applied sciences

MDPI

Article

Hybrid Optimization and Explainable Deep Learning for Breast Cancer Detection

Maral A. Mustafa [1,2,*], Osman Ayhan Erdem [3] and Esra Söğüt [3]

[1] Graduate School of Natural and Applied Sciences, Department of Computer Engineering, Gazi University, Ankara 06560, Türkiye
[2] Production Mechanics Techniques, Kirkuk Technical Institute, Northern Technical University, Kirkuk 36001, Iraq
[3] Department of Computer Engineering, Faculty of Technology, Gazi University, Ankara 06560, Türkiye; ayerdem@gazi.edu.tr (O.A.E.); esrasogut@gazi.edu.tr (E.S.)
* Correspondence: 24833304401@gazi.edu.tr or maralanwer@ntu.edu.iq

Abstract

Breast cancer continues to be one of the leading causes of women's deaths around the world, and this has emphasized the necessity to have novel and interpretable diagnostic models. This work offers a clear learning deep learning model that integrates the mobility of MobileNet and two bio-driven optimization operators, the Firefly Algorithm (FLA) and Dingo Optimization Algorithm (DOA), in an effort to boost classification appreciation and the convergence of the model. The suggested model demonstrated excellent findings as the DOA-optimized MobileNet acquired the highest performance of 98.96 percent accuracy on the fusion test, and the FLA-optimized MobileNet scaled up to 98.06 percent and 95.44 percent accuracies on mammographic and ultrasound tests, respectively. Further to good quantitative results, Grad-CAM visualizations indeed showed clinically consistent localization of the lesions, which strengthened the interpretability and model diagnostic reliability of Grad-CAM. These results show that lightweight, compact CNNs can be used to do high-performance, multimodal breast cancer diagnosis.

Keywords: deep learning; transfer learning; breast cancer detection; optimization algorithms; explainable AI (XAI); Grad-CAM

check for updates

Academic Editor: Gang Wei

Received: 23 June 2025
Revised: 25 July 2025
Accepted: 26 July 2025
Published: 30 July 2025

Citation: Mustafa, M.A.; Erdem, O.A.; Söğüt, E. Hybrid Optimization and Explainable Deep Learning for Breast Cancer Detection. *Appl. Sci.* **2025**, *15*, 8448. https://doi.org/10.3390/app15158448

Copyright: © 2025 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Breast cancer (BC) is the interruption of the normal growth of cells and the development of abnormal and uncontrolled cellular growth in breast tissue, which causes a tumor [1]. Despite this, it still accounts for the main cause of cancer-related deaths in women in 9 out of 10 countries, and there is an occurrence of more than 2.3 million new cases every year [2]; about 0.4 million female deaths every year. In particular, clinical outcomes are increasingly influenced by systemic immune-inflammation indices, especially in HER2-positive metastatic cases, where early stratification may guide treatment selection [3]. The disease arises from abnormal and uncontrolled growth of breast tissue cells, most commonly manifesting as ductal carcinoma among over 20 histological subtypes [4]. Key diagnostic features include architectural distortion, microcalcifications, and mass formation [5]. Early diagnosis significantly reduces mortality [6], with treatment options spanning surgery, radiation, and chemotherapy [7]. Despite advances in screening technologies [8], including spectroscopy-based methods such as surface-enhanced Raman spectroscopy using porous silicon Bragg reflectors for enhanced signal sensitivity in breast

cancer detection [9], challenges persist due to diagnostic delays, noise in medical images, and interpretability limitations in AI-based systems [10–12]. Traditional image processing techniques, such as shape features, local binary patterns, and gray-level co-occurrence matrices, have improved discrimination in mammography [13]. However, deep learning (DL), particularly convolutional neural networks (CNNs), has demonstrated superior performance in medical image analysis with a 6% performance gain over classical methods [14–17]. DL models demand extensive annotated datasets, computational resources, and expert knowledge constraints often unmet in clinical settings [18]. Transfer learning (TL), which repurposes pretrained CNNs for medical imaging, has thus emerged as a robust alternative [19,20]. As imaging data grow in complexity, dimensionality reduction and feature selection (FS) techniques have become critical for improving model generalization and reducing noise [21–25]. Such biological and molecular complexity, including stress-induced molecular mechanisms affecting cellular systems, has implications for the interpretability of cancer models [26]. Metaheuristic FS algorithms such as FLA and DOA have shown effectiveness in boosting classification by selecting high-quality features [27,28]. However, DL models often behave as black boxes, raising concerns regarding clinical trust and decision transparency [29–32]. To address this, explainable AI (XAI) has introduced interpretability tools like Grad-CAM, which visually highlights image regions influencing model predictions [33,34]. This is particularly valuable in healthcare, where understanding decision boundaries is crucial for clinical acceptance. Numerous studies have explored similar directions.

- Development of a hybrid diagnostic framework using mammography and ultrasound images, constructing a diverse fusion dataset for binary breast cancer classification;
- Fine-tuning MobileNet using the Firefly Algorithm (FLA) and Dingo Optimization Algorithm (DOA) to optimize feature selection and improve classification performance;
- Comparative evaluation of multiple transfer learning models (CNN, MobileNet, Xception, DenseNet121, EfficientNet, ResNet50, and VGG16) to identify the most effective architecture for breast cancer diagnosis;
- Performance validation on three datasets: MIAS, Breast Ultrasound, and their fusion;
- Integration of Grad-CAM to generate interpretable visual explanations for model predictions;
- Demonstration of the framework's superiority over existing methods in terms of accuracy, precision, recall, and F1-score.

For example, ensemble CNNs combining AlexNet, GoogLeNet, VGG16, and ResNet-50 enhanced microcalcification detection in mammograms [35]. DenseNet with attention mechanisms and multi-level TL surpassed 84% accuracy [36], while Few-Shot Learning integrated with TL achieved 95.6% accuracy and 0.97 AUC [37]. In infrared thermography, ensemble models with attention-guided Grad-CAM achieved 98.04% accuracy [38], while CNN-IoT frameworks supported real-time diagnosis with 95% accuracy [39]. In FS research, LASSO-enhanced Random Forests reached 90.68% accuracy [40], and grid-searched RF and AdaBoost models hit 99% and 98%, respectively [41]. For metastasis prediction, XGBoost and RF yielded 93.6% accuracy and 91.3% AUC [42]. Hybrid techniques such as TLBO-SSA [43] and BinJOA-S [44] achieved 98.46% accuracy and advanced multi-objective FS. Outside of BC, flower pollination-based segmentation improved lung cancer detection [45], while bacterial foraging and emperor penguin optimization hybrids excelled in BC classification with >98% accuracy and AUC > 0.998 [46]. Inspired by tumor growth dynamics, the Liver Cancer Algorithm (LCA) reached 98.704% accuracy on the MAO dataset [47]. FS wrappers using RF, SVM, and DT achieved AUCs up to 0.98 in early detection [48]. In ultrasound-based CAD systems, Autoencoders and DeepLabV3+ achieved 97.4% accuracy with strong Grad-CAM interpretability [49]. Recent segmentation strategies using

cluster-enhanced Transformer architectures, such as CenterFormer, have demonstrated robust performance in delineating complex structures in unconstrained imaging scenarios [50], which may inform future exploration of transformer-based breast cancer models. SHAP-based XAI integrated with RF outperformed clinical nomograms in NSLN metastasis prediction [51], while CNNs combined with XAI methods (LIME, SHAP, Saliency Maps) showed high interpretability [52]. Patho-Net (GRU + U-Net) achieved 98.9% accuracy in histopathology via GLCM and resizing-based preprocessing [53]. Other studies have used DenseNet201 with LIME and Grad-CAM++ for up to 99% accuracy in mammograms [54], shape-based models with t-SNE and UMAP for 98.2% AUC in 3D tomosynthesis [55], and condensed CNNs for 96.67% accuracy on MIAS + DDSM [56]. Lastly, ResNet50, optimized by the Improved Marine Predators Algorithm, attained 98.88% accuracy on MIAS and CBIS-DDSM [57]. Table 1 summarizes the methods and performance metrics of these studies. Despite these advances, most prior studies rely on single-modality imaging, lack hybrid optimization, or offer limited interpretability. To address these gaps, our work introduces a hybrid framework that integrates mammography and ultrasound data, uses MobileNet fine-tuned with DOA and FLA, and incorporates Grad-CAM for clinically reliable visual explanations.

Table 1. Comparative table of related work.

Ref.	Methodology	Dataset or Application Context	Major Findings	Noted Limitations	Distinctive Contribution Relative to the Present Work
[35]	Ensemble CNNs (AlexNet, GoogLeNet, VGG16, ResNet-50) + Transfer Learning	Mammograms	Confidence: 0.9305 (microcalcifications), 0.8859 (normal)	Focus on microcalcifications; no multimodal input	Improved microcalcification detection using ensemble learning
[36]	DenseNet with attention + Multi-level Transfer Learning	Augmented mammogram dataset	Accuracy > 84%	Performance dependent on attention layers	Robust classification via attention-enhanced DenseNet
[37]	Few-Shot Learning (FSL) + Transfer Learning	Limited annotated mammograms	Accuracy: 95.6%, AUC: 0.970	Specialized FSL models may not generalize broadly	Combined FSL with TL for better generalization in low-data settings
[38]	Ensemble CNNs + Attention-Guided Grad-CAM	Infrared thermograms (DMR)	Accuracy: 98.04%, AUC: 0.97	Domain-specific (infrared imaging only)	Enhanced interpretability using channel and spatial attention
[39]	CNN + IoT Integration	Real-time breast cancer detection	Accuracy: 95%	Not evaluated on public benchmarks	Integrated IoT with CNN for scalable diagnostics
[40]	LASSO + RF, KNN, MLP	Breast cancer classification	RF: Accuracy 90.68%, F1: 94.60%; KNN Recall: 98.80%	No DL used	Demonstrated impact of LASSO FS on ML model accuracy
[41]	RF, AdaBoost, GNB, Logistic Regression + Grid Search	BC classification	RF: 99%, GNB: 91%	No XAI or interpretability included	Performance benchmarking across traditional ML algorithms
[42]	XGBoost, RF, SVM, GBDT, Logistic Regression	Distant metastasis prediction	Accuracy: 93.6%, F1: 88.9%, AUC: 91.3%	Limited to metastasis focus	Ensemble model for metastasis risk prediction
[43]	SSA + TLBO + ANN	651 mammograms, UCI dataset	Accuracy: 98.46%, Sensitivity: 98.81%, AUC: 0.997	Dependent on heuristic FS quality	Hybrid FS using SSA + TLBO for enhanced selection
[44]	BinJOA-S (Binary Jaya Optimization with scalarization)	Breast cancer datasets	High classification accuracy and fitness score	Algorithm complexity not discussed	Multi-objective FS via scalarization with optimization

Table 1. *Cont.*

Ref.	Methodology	Dataset or Application Context	Major Findings	Noted Limitations	Distinctive Contribution Relative to the Present Work
[45]	Snake-spline segmentation + Flower Pollination Optimization	CT lung cancer dataset	Accuracy: 84% using 33 features	Focused on lung CT, not BC	Hybrid segmentation and FS using FPO for SVM
[46]	Bacterial Foraging + Emperor Penguin Optimization (EPO), hGSAEPO	WDBC BC dataset	Accuracy: 98.31%, AUC > 0.998	Computational complexity of hybrids	High-performing hybrid FS + XAI integration
[47]	Liver Cancer Algorithm (LCA) + SVM optimization	MonoAmine Oxidase (MAO) dataset	Accuracy: 98.704%	Not breast cancer-specific	Novel evolutionary optimization inspired by tumor dynamics
[48]	Wrapper-based FS + RF, SVM, DT	Early BC biomarker detection	AUC: 0.89–0.98, Sensitivity: 0.94, Specificity: 0.90	Requires feature engineering	Sequential backward FS enhancing early biomarker identification
[49]	DeepLabV3+ + Autoencoder + Grad-CAM + GLCM	Breast ultrasound datasets	Accuracy: 97.4%, Dice: 0.981	Specific to ultrasound; not multimodal	Strong segmentation and interpretability using XAI
[51]	Random Forest + SHAP (XAI) + Clinical features	NSLN metastases prediction	Accuracy: 72.2%, AUC: 0.77	Limited dataset (183 cases)	Interpretable model using SHAP for clinical guidance
[52]	CNN (VGG16, VGG19, ResNet) + XAI (LIME, SHAP, Saliency Maps)	Histopathology images	Accuracy: 92.59% (VGG19)	Downsampled input, lacks real-time validation	Evaluation of XAI methods for clinical interpretability
[53]	Patho-Net (GRU + U-Net + GLCM + XAI)	BreakHis 100X histopathology	Accuracy: 98.90%	Focused on histopathology only	End-to-end DL with built-in interpretability and preprocessing
[54]	CNNs + LIME, Grad-CAM, Grad-CAM++	Mammography images	DenseNet201: Accuracy 99%	Limited multiclass analysis	Comprehensive use of multiple XAI methods
[55]	3D tomosynthesis + Grad-CAM + LIME + t-SNE + UMAP	Lesion-region images	AUC: 98.2%	High-dimensional visualization needed	Structural + visual interpretability in multiclass BC diagnosis
[56]	Condensed CNN	MIAS, DDSM, Combined	Accuracy: 96.67% (combined dataset)	Limited XAI integration	Efficient CNN optimized for performance and speed
[57]	ResNet50 + Improved Marine Predators Algorithm (IMPA)	MIAS, CBIS-DDSM	Accuracy: 98.88%	Dataset-specific performance	Metaheuristic TL optimization with opposition learning

2. Proposed Methodology

As shown in Figure 1, a novel and comprehensive deep learning-based framework for the early detection of breast cancer is proposed in this study. It makes use of the complementary strengths of two different medical imaging datasets: The Breast Ultrasound Images dataset and the MIAS (Mammographic Image Analysis Society) dataset. The approach begins with independent preprocessing of both datasets to ensure consistency in format, resolution, and class labeling.

To increase model generalization and data diversity, a fusion of the datasets is then carried out. The fused dataset is then subjected to an additional preprocessing phase to unify input dimensions and normalize image values, ensuring compatibility with DL architectures. The unified dataset is divided into training and testing sets after preprocessing, preparing the way for a dual-path modeling approach. Several cutting-edge transfer learning techniques are used in the first path, such as CNN, MobileNet, Xception, DenseNet121, EfficientNet, ResNet50, and VGG16, which are fine-tuned on the training data to learn high-level image representations. In parallel, the second path introduces

optimized transfer learning, where feature extraction using MobileNet is enhanced through two nature-inspired metaheuristic algorithms: the DOA and the FLA. By choosing the most pertinent characteristics, these algorithms cut down on redundancy and enhance classification performance.

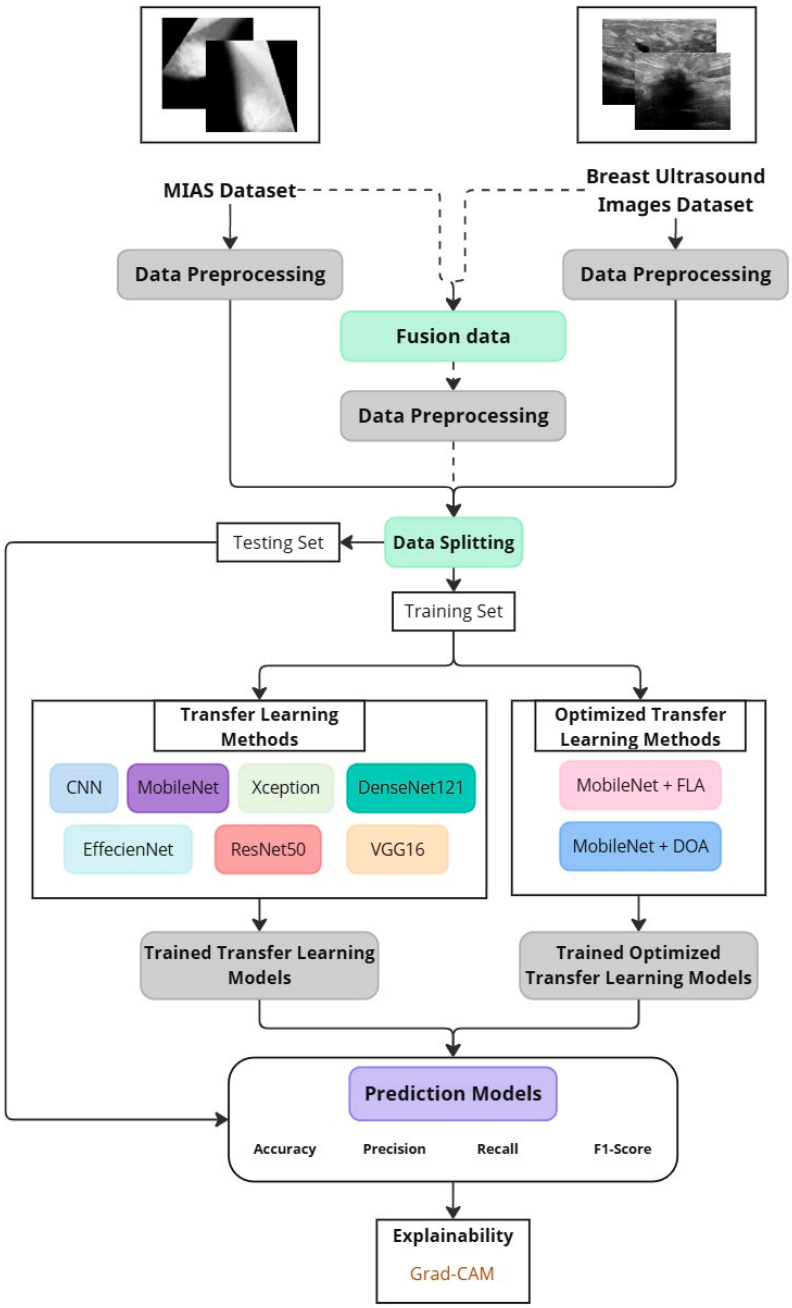


Figure 1. A Hybrid DL Framework for Multi-Modal Data Fusion and Optimization in the Detection of Breast Cancer.

Both sets of models, optimized versions and traditional transfer learning, are assessed using a strong prediction framework that incorporates performance criteria, including F1-score, accuracy, precision, and recall. To improve interpretability and trust in the predictions, the framework highlights important areas in mammograms and ultrasound images that affected the predictions using Grad-CAM, which provides a visual explanation of model decisions.

2.1. Data Description

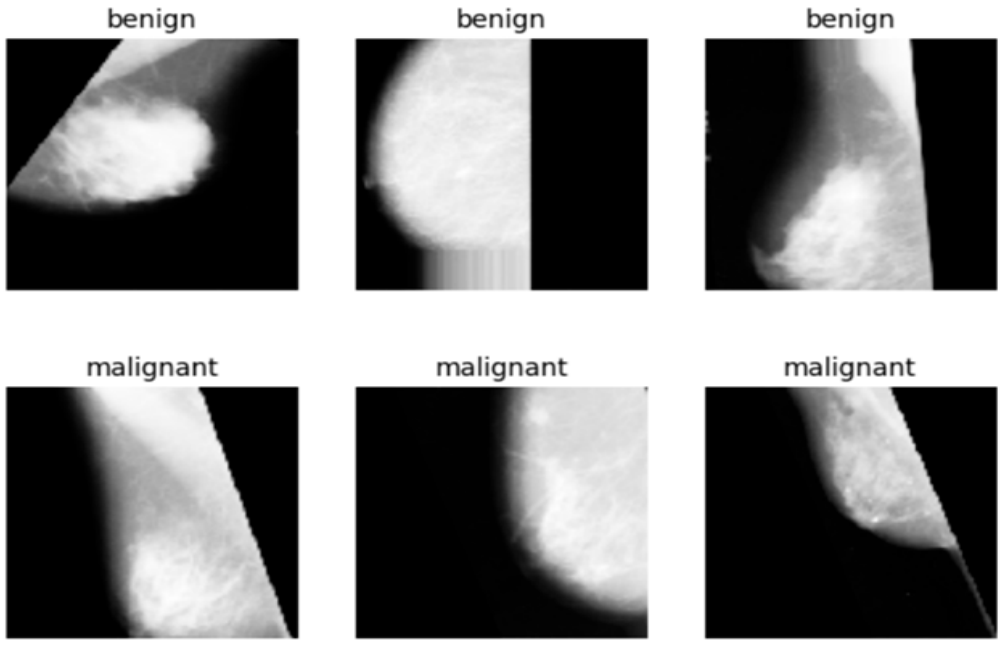
2.1.1. MIAS Dataset Description

The most common benchmark dataset used in the validation of computer-aided diagnostic (CAD) systems of breast cancer detection is the Mammographic Image Analysis Society (MIAS) dataset [57], which is by far one of the most well-known benchmark datasets in medical image processing. There are 322 digitized mammograms with a resolution of 1024×1024 pixels and annotated by accurate ground truth labels that mention whether the tumor is benign or malignant. The MIAS dataset is especially valuable due to its diversity: it contains samples of different densities of the breast tissue, a wide range of varying shapes and localizations of the lesions, and of different pathologies. This explains why this dataset makes ML models capable of generalizing well across diverse patient profiles as well as clinical situations, and therefore, the dataset is very suitable in the real-world deployment of ML models in clinical decision support systems.

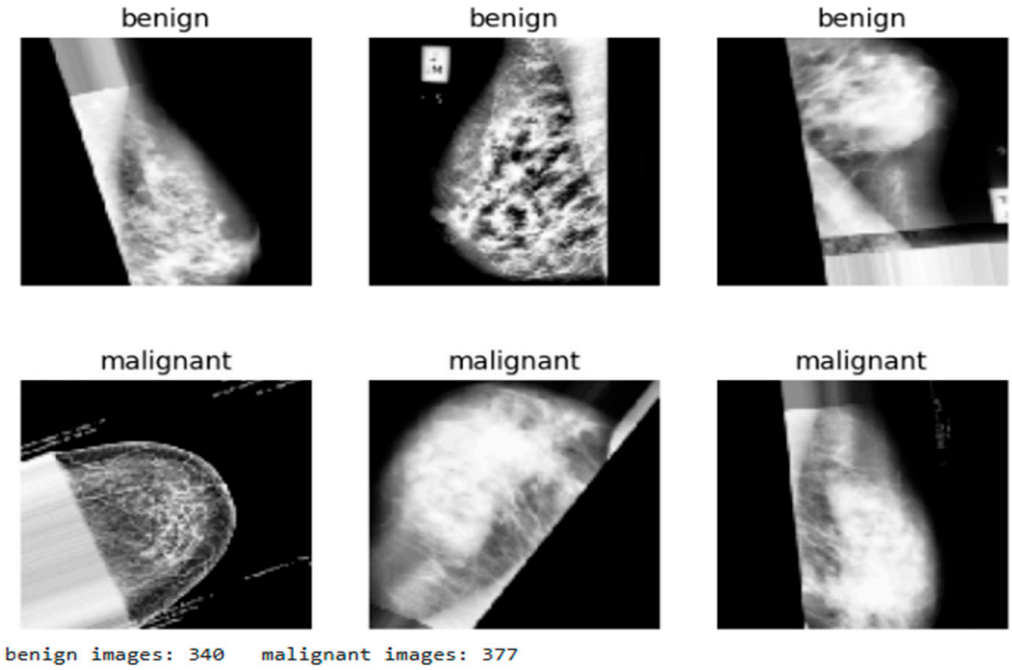
In order to demonstrate the quality of the image and pathological diversity of the MIAS dataset, in Figure 2a, a list of six typical mammograms that portray three benign cases and three malignant ones is organized in two rows. The upper row has benign samples, whereas the lower row has malignant ones. The visual comparison enables the reader to obtain an insight into the distinguishing features between the two classes. Benign lesions are likely to have smoother and more homogenous grains and sharp edges, whereas malignant lesions are likely to have heterogeneous intensities and irregular shapes as well as architectural distortions. Such images give key diagnostic information that should be used by the radiologists in their diagnosis, and this is what DL models seek to grasp and label. The fact that both types of lesions are included with the figure sets forth the necessity of navigation in the elaboration of sensitive textural distinctions that cannot always be seen using the human eye but can be learned by complex computational mechanisms.

To continue the visual investigation, the corresponding Figure 2b shows a second group of mammograms, on which, once more, benign and malignant ones were randomly assigned (i.e., in equal numbers). These present a more general picture of intra-class variation and underline the inter-patient heterogeneity within each class. The benign ones in this figure illustrate continuous tissue patterns, though not quite recognizable as abnormal, whereas the malignant forms include high-contrast masses, jagged edges, and show signs of tissue disruption, indicating that the diagnosis of these conditions is not straightforward and relies on the visual basis alone, which is extremely difficult. This number also contributes to the balance in representation of difficult cases to achieve the fact that the DL models are trained with various real-life manifestations of abnormalities in breasts.

To provide an extended review of the image dataset, Figure 3 presents a 5×5 array of randomly chosen images in the MIAS; a wide coverage of the data better describes benign and malignant cases. Every picture has its labels, which help readers to learn about morphological disparity in the dataset. The observations described by this figure can be examined with the help of side-by-side visual comparison and can offer an idea of the possible repeatability or variance of some aspects, e.g., tumor density, margins of the mass, or patterns of calcification. They are useful not only as a qualitative reference to the complexity of the dataset but also as evidence of why an automated diagnostic tool with the ability to handle visual noise, anatomical variability, and ambiguous patterns, which even an experienced radiologist might find difficult to work with, is needed.



(**a**)



(**b**)

Figure 2. (**a**) MIAS Dataset Description. (**b**) MIAS Dataset Description.

In short, Figures 2a,b and 3 are presented to provide a visual location of the reader as to the nature and complexity of the MIAS dataset from the basis of much of the modeling development and model evaluation in this paper. These values together show why the process of breast cancer diagnosis with the help of mammography needs highly discriminative computational models and why one should resort to explainable DL methods that could reliably classify more or less distinctive pathological aspects in a vast variety of patients.

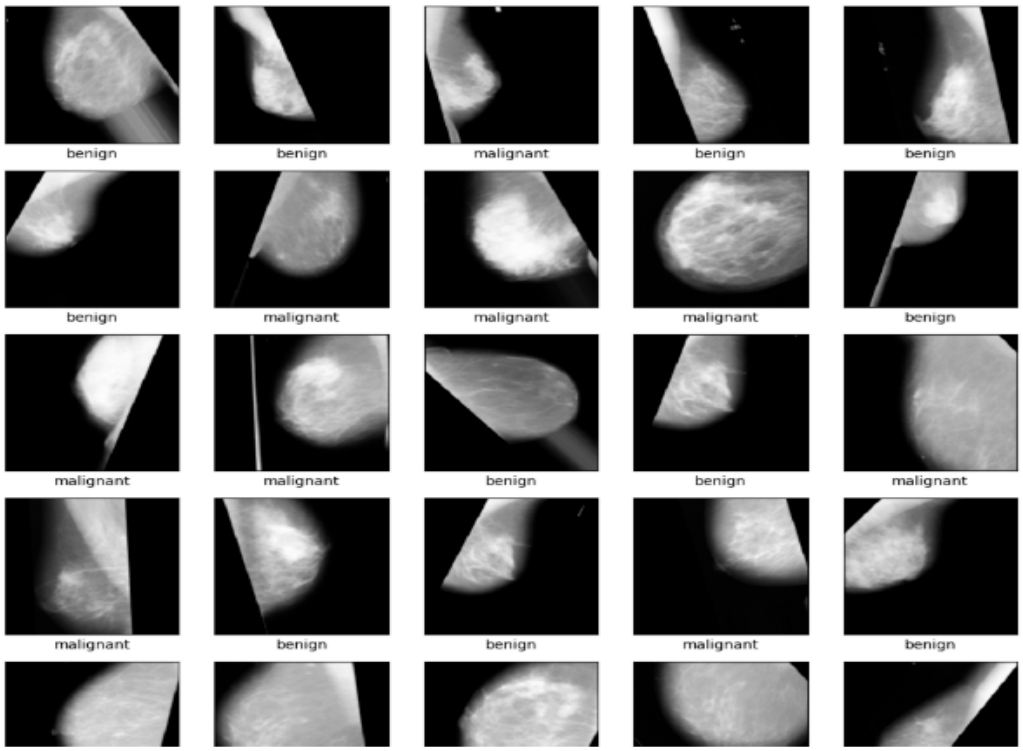


Figure 3. MIAS Dataset Description.

2.1.2. Breast Ultrasound Images Dataset Description

One of the biggest causes of death for women worldwide is still breast cancer, and patient outcomes are greatly enhanced by early detection. A useful resource for developing computer-aided diagnosis techniques is the Breast Ultrasound Images Dataset, which is accessible on Kaggle [58]. Three different kinds of ultrasound scans, normal, benign, and malignant, are included in this collection. This kind of classification makes it easier to create and assess ML models for tasks including segmentation, detection, and classification in the diagnosis of breast cancer. In this study, only two classes, benign and malignant, are considered, excluding the normal category to focus on differentiating pathological cases. The dataset, which was gathered in 2018, consists of 780 high-resolution photos that are saved in PNG format and have an average size of 500 $\times$ 500 pixels. These pictures were taken from 600 female patients who ranged in age from 25 to 75 years. For each case, corresponding ground truth notations are provided, allowing for supervised learning and precise model evaluation. The inclusion of both pathological and non-pathological samples makes this dataset particularly suitable for training robust DL models aimed at enhancing diagnostic accuracy in breast ultrasound analysis.

Figure 4 displays sample images from the mammography dataset, divided into benign and malignant categories.

- Top Row (Benign): Three benign images show uniform textures with minimal irregularities, indicating non-cancerous conditions;
- Bottom Row (Malignant): Three malignant images exhibit distinct anomalies, suggesting potential malignancies.

The 'normal' class in the BUSI dataset was excluded from training due to its limited representation and lack of pathological markers, which rendered it insufficient for robust learning and generalization. Moreover, preliminary experiments indicated that inclusion of the normal class introduced class imbalance without contributing significantly to model

discriminability, justifying its omission to preserve diagnostic focus on benign versus malignant classification.

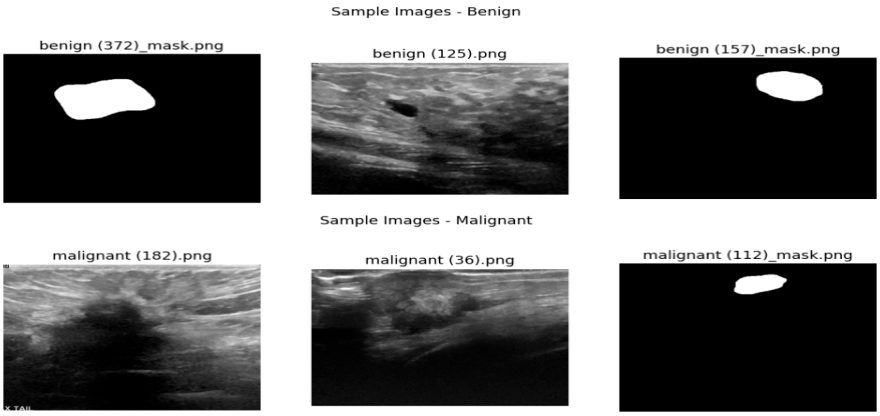


Figure 4. Breast Ultrasound Images Dataset Description.

2.1.3. Fusion Dataset Description

A fusion dataset was constructed by combining the MIAS mammographic dataset and the Breast Ultrasound Images Dataset. This fusion approach integrates two distinct imaging modalities, mammography and ultrasound, each offering complementary diagnostic information. While mammographic images provide high-resolution anatomical details and are widely used in breast cancer screening, ultrasound images are particularly effective in differentiating between cystic and solid lesions, especially in dense breast tissues.

Combining these datasets improves the model's capacity to generalize across various clinical circumstances by allowing it to learn from a wider variety of visual aspects and lesion characteristics. Prior to fusion, both datasets were independently preprocessed to ensure consistency in image resolution, format, and labeling schemes. Only the benign and malignant classes were retained from both datasets to maintain a uniform binary classification framework. The fused dataset contains samples from both mammograms and ultrasound scans, ensuring a balanced representation of each imaging type. In addition to expanding the amount and diversity of training data, this integration makes use of multimodal learning's advantages, which have been demonstrated to greatly improve performance in medical picture analysis. The resultant dataset provides a strong basis for creating a computer-aided diagnostic system that is more dependable and flexible.

Figure 5 displays sample images from the mammography dataset, divided into benign and malignant categories.

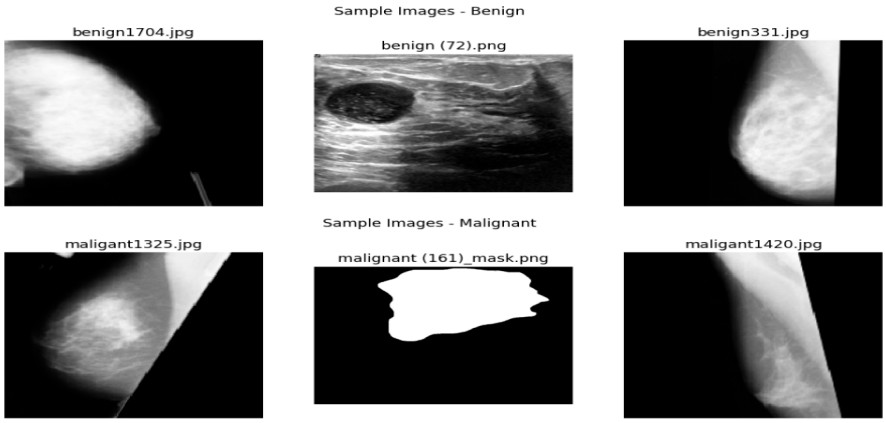


Figure 5. Fusion Dataset Description.

- Top Row (Benign): Three benign images show uniform textures with minimal irregularities, indicating non-cancerous conditions;
- Bottom Row (Malignant): Three malignant images exhibit distinct anomalies, suggesting potential malignancies.

In order to have a coherent fusion dataset, MIAS (mammography) and BUSI (ultrasound) datasets consist of images matched according to identical class labels. Since there was such a marked imbalance in the classes, especially the overrepresentation of benign cases, balancing techniques were used. These were random undersampling of majority classes and oversampling of minority classes by augmenting the data by flipping, rotation, and contrast enhancement. This method enhanced the generalizability of the model and reduced bias in training. Combining the process of mammography and ultrasound images presented an issue of differences in the modalities in terms of texture, resolution, and contrast. In combination, all image sizes were uniformed to a standardized input (224 224) and standardized on an equalized intensity range. To ensure the clinical relevancy of the fused dataset, modality-specific preprocessing methods were employed to particularly retain diagnostic qualities specific to imaging modality, e.g., sharp boundary delineation in the case of mammograms and echotextural transforms of ultrasound.

2.2. Data Preprocessing

In order to guarantee the consistency and quality of the input data used for DL model training and evaluation, the preprocessing stage is essential. In this work, we employed the ImageDataGenerator class from TensorFlow Keras to preprocess and augment the input images before training. This approach enabled us to rescale the pixel values and apply various real-time transformations such as rotation, zooming, and horizontal flipping. These augmentation techniques increased the diversity of the training data and contributed to improving the model's generalization performance by reducing the risk of overfitting. Additionally, the tool facilitated the efficient loading and batching of images directly from organized directories, which streamlined the data pipeline during training. Three datasets were preprocessed for this study: the merged Fusion dataset, the Breast Ultrasound Images dataset, and the MIAS mammography dataset. The photos in the MIAS dataset were initially categorized into test, validation, and training sets before being divided into two classes: benign and malignant. Prior to scaling all photos to a consistent resolution of 224×224 pixels and converting them to RGB format for convolutional neural network compatibility, the first step in preprocessing was to confirm the image format and content. Normalization (scaling pixel values between 0 and 1) and grayscale conversion were used to provide numerical stability throughout training. The matching class labels were then encoded once the photographs were put into memory as NumPy arrays. After dividing the dataset in an 80/20 ratio into training and test sets, one-hot vectors were used for label encoding. To ensure balanced learning and mitigate bias, class distributions were visualized, and data were shuffled before model input, as illustrated in Figure 6.

The Breast Ultrasound Images dataset was subject to a unique preprocessing pipeline for ultrasound images. Images were extracted from two named subfolders, namely, benign and malignant, and non-image files were discarded. All the images were rescaled to the size 224×224, converted into grayscale, and normalized in the range [0, 1]. Grayscale images were converted to RGB by replicating the channel because pretrained CNN models require three-channel input. The dataset was divided by means of stratified sampling into the training and the testing sets, with the preservation of the class proportions, and label encoding was performed (0 benign; 1 malignant). As Figure 7 demonstrates, there is a severe imbalance in the classes in the dataset, as occurring benign cases are more frequent than malignant ones. To overcome this, on the minority (malignant) class, data

augmentation was applied using transformations, which included rotation, horizontal flip, and contrast. Not only did this increase class balance in the training data but it also better enabled the model to learn more of the underrepresented patterns.

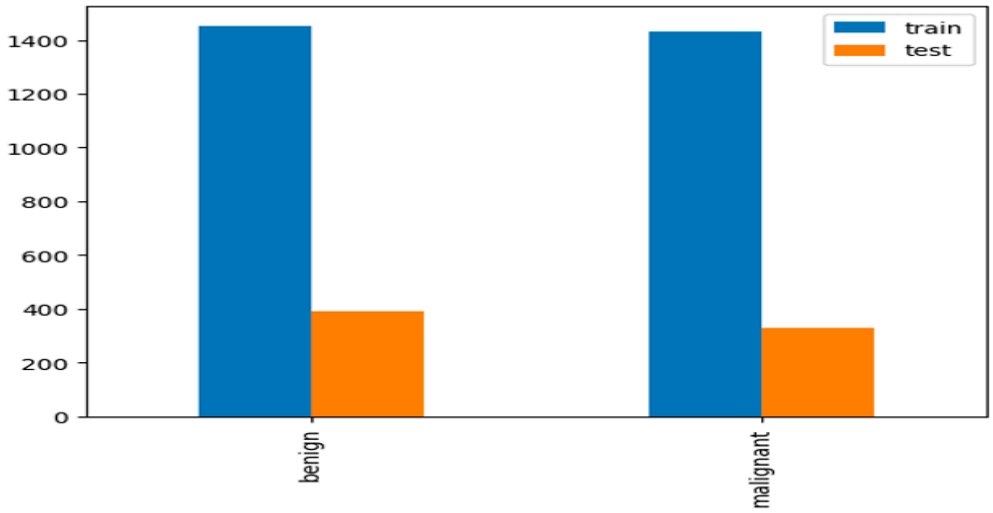


Figure 6. Class Distribution in the MIAS Dataset.

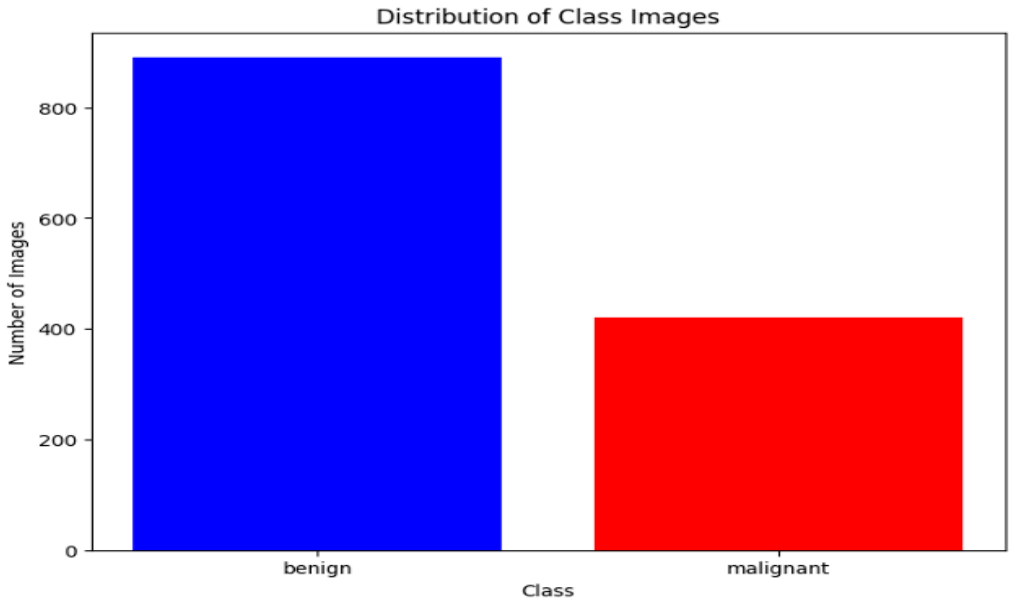


Figure 7. Class Distribution in the Breast Ultrasound Images Dataset.

Finally, the Fusion dataset, constructed by merging the MIAS and Breast Ultrasound datasets, required unified pre-processing to align differences in imaging modalities and formats. All images were standardized to 224 × 224 RGB format and normalized in the same way as the previous datasets. The labels were harmonized into a binary classification scheme. This combined dataset aimed to improve model generalizability by exposing the model to greater variability in breast cancer presentations across imaging types. Following pre-processing, an 80/20 split with label stratification was also employed to separate the fused dataset into training and test sets. The class distribution after fusion, shown in Figure 8, reveals a more balanced and expanded dataset, which serves as a robust foundation for the subsequent training of transfer learning and optimized classification models.

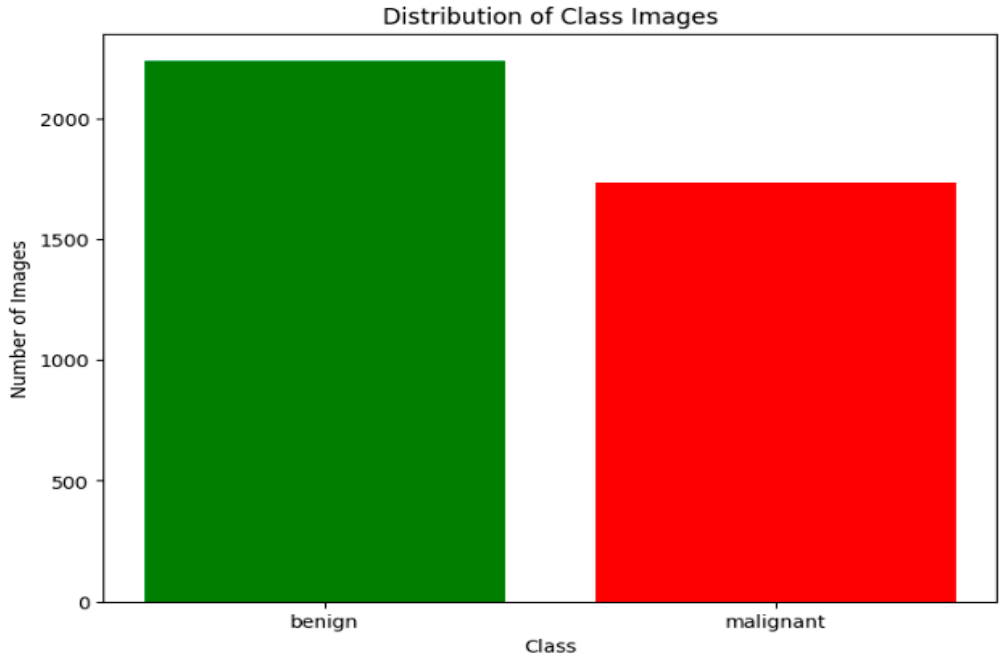


Figure 8. Class Distribution in the Fusion Dataset.

2.3. Transfer Learning Models

A wide range of transfer learning techniques was used to improve the suggested breast cancer classification system's performance and capacity for generalization. Since transfer learning allows information obtained from large-scale datasets like ImageNet to be reused, it has emerged as a key strategy in medical image analysis, especially when working with sparse datasets. Several trained models, including MobileNet, Xception, DenseNet201, VGG16, ResNet50, and EfficientNetB0, as well as specially created CNNs, are among the tactics used. Each model was fine-tuned with custom dense layers for binary classification (benign vs. malignant), offering unique structural advantages in feature extraction and computational efficiency.

- MobileNet inspired a bespoke CNN that uses depth-wise separable convolutions to lower the number of parameters and computational cost without sacrificing speed. The four separate convolution blocks in this lightweight design are followed by batch normalizing and max pooling, which results in dense layers that are completely linked and include dropout regularization. It serves as a foundational architecture for comparison against more complex pre-trained models;
- MobileNet, a highly efficient network designed for mobile and embedded vision applications, further leverages depth-wise separable convolutions and a streamlined structure. In this work, the base MobileNet was frozen and extended with dense layers (1024, 512) and dropout to prevent overfitting. Its main advantage lies in its low latency and memory footprint while achieving strong performance on small medical datasets;
- The Xception model, built upon depth-wise separable convolutions and residual connections, expands on Inception modules by replacing them with extreme Inception blocks. In this approach, a stack of thick layers (2024, 1024, 512) with dropout was added for binary classification, while the ImageNet-trained layers were frozen. The strength of Xception is its ability to preserve parameter efficiency while capturing fine-grained spatial hierarchies in medical pictures;
- In order to encourage feature reuse and mitigate the vanishing gradient issue, DenseNet201, which is renowned for its dense connection across layers, was also used. DenseNet promotes gradient flow and increases representational power by

feed-forwardly linking each layer to every other layer. Here, it was augmented with dense layers (2024 × 2, 1024) and dropout layers to tailor the model for binary decision-making in cancer detection;

- VGG16 was employed for its simplicity and consistency. VGG16's consistent architecture of 3 × 3 convolution layers and 2 × 2 pooling layers makes it resilient and interpretable even if it lacks the sophisticated advances of more current models. In this study, VGG16's convolutional base was frozen, and new dense layers (2048, 1024, 512) were stacked to adapt it for the classification task. The key strength of VGG16 lies in its deep hierarchical structure, which is particularly beneficial for capturing visual patterns in high-resolution mammographic and ultrasound images;
- ResNet50, another widely used model, addresses the degradation problem in very deep networks through the introduction of residual learning. The residual blocks al low gradients to flow directly through skip connections, enabling the training of deeper architectures. In this implementation, the ResNet50 model was extended with multiple dense layers (2024, 1024, 512, 256, 128) and dropout, providing a deep and expressive architecture that balances depth and training stability;
- The method was modified to use EfficientNetB0, a more modern architecture that scales depth, breadth, and resolution in a compound way. It is renowned for using fewer FLOPs and parameters to achieve excellent precision. EfficientNet's efficient scaling strategy allows it to outperform traditional CNNs on various benchmarks. In this work, the EfficientNetB0 backbone was frozen, and dense layers (1024, 512, 256, 128) were added with dropout to optimize performance for the binary classification task. Its compact design makes it particularly suitable for real-time and resource-constrained deployment scenarios.

Each of these transfer learning methods contributes uniquely to the proposed framework. While models like MobileNet and EfficientNetB0 are optimized for speed and efficiency, deeper architectures such as DenseNet201 and ResNet50 provide richer feature representations. The hybrid inclusion of lightweight and heavyweight models enables robust comparative analysis and the potential for ensemble learning. Ultimately, these transfer learning strategies not only expedite model development by leveraging prior knowledge but also significantly improve classification accuracy and generalization when applied to the fused dataset of breast ultrasound and mammographic images.

2.4. Optimization Algorithms

2.4.1. Algorithm for Dingo Optimization

DOA [59] is a new bio-inspired metaheuristic based on the more advanced predation and survival techniques of Australian dingoes to solve complex, nonlinear, high-dimensional optimization issues. Several behaviors, including cooperative group attacks, individual over pursuit and opportunistic scavenging as well as behaviors that are flexible to environmental limitations, are integrated in DOA to have a successful balance between exploration and exploitation phases. DOA is applied in the long-established setting of constrained optimization that can be described as

$$\min_{x \in R^n} f(x) \text{ subject to } g_i(x) \leq 0,\ h_j(x) = 0 \tag{1}$$

where $\vec{x} \in$ Rn represents the decision variables, $f(\vec{x})$ is the objective function, and g_i and h_j denote inequality and equality constraints, respectively. This formulation follows the canonical representation of constrained optimization problems as outlined in foundational texts such as Boyd and Vandenberghe [60]. Four key dingo-inspired behaviors are mathematically modeled to guide the evolution of candidate solutions through the search space,

forming the core operational logic of DOA. The update rules defined in Equations (2)–(5) are core components of the DOA, adapted from gravitational interaction concepts introduced by [61] in the Gravitational Search Algorithm (GSA) framework. These equations do not represent novel contributions but are included here to illustrate the internal mechanics of DOA. All symbols and parameters are defined at first mention to ensure clarity. The Group Attack strategy, which simulates coordinated hunting, updates the position of a dingo (candidate solution) as follows:

$$\overline{x_i}(t+1) = \beta_1 \left(\frac{1}{n_a} \sum_{k=1}^{n_a} \overline{x_{atk}}(t) - \overline{x_i}(t) \right) - \overline{x^*}(t) \tag{2}$$

where

- $\overline{x_i}(t)$: current position of the i-th candidate;
- $\overline{x_{atk}}(t)$: positions of n_a randomly selected attackers;
- $\overline{x^*}(t)$: best-known solution so far;
- $\beta_1 \sim \mathrm{U}(-2, 2)$: stochastic scaling factor for update magnitude.

The Persecution strategy mimics an individual dingo pursuing prey, as follows:

$$\overline{x_i}(t+1) = \overline{x^*}(t) + \beta_1 e^{\beta^2} (\overline{x_{r1}}(t) - \overline{x_i}(t)) \tag{3}$$

where

- $\overline{x_{r1}}(t)$: randomly selected peer position;
- $\beta^2 \sim \mathrm{U}(-1, 1)$: additional stochastic factor.

The Scavenging strategy, simulating opportunistic feeding behavior, updates as

$$\overline{x_i}(t+1) = \frac{1}{2} \left[e^{\beta^2} \overline{x_{r1}}(t) - (-1)\sigma \overline{x_i}(t) \right] \tag{4}$$

where $\sigma \in \{0, 1\}$ is a binary variable that introduces randomness in direction. Lastly, the Survival Rate Adaptation strategy determines whether a dingo (solution) should be replaced based on its survival rate, as follows:

$$\mathrm{Survival}(i) = \frac{f_{max} - f_i}{f_{max} - f_{min}} \tag{5}$$

where f_i represents the current solution's fitness and f_{max} and f_{min} represent the population's worst and greatest fitness values, respectively. Individuals with low survival probability are replaced using a hybridized movement.

$$\overline{x_i}(t+1) = \overline{x^*}(t) + \frac{1}{2} [\overline{x_{r1}}(t) - (-1)\sigma \overline{x_2}(t)] \tag{6}$$

The algorithm employs a probabilistic controller with two switching parameters, P and Q, to decide whether the population enters hunting, scavenging, or survival mode in each iteration. This adaptive mechanism allows DOA to navigate rugged, multi-modal search spaces effectively, avoiding premature convergence while accelerating progress toward global optima. DOA is especially well suited for challenging real-world applications, such as engineering design, biomedical image analysis, and feature selection tasks, since it can combine several behavioral heuristics into a single optimization technique. In this study, DOA is applied to optimize deep feature sets extracted from medical images, demonstrating its capacity to reduce feature dimensionality and enhance classification performance in breast cancer detection.

2.4.2. Firefly Algorithm

The FLA is a recently developed metaheuristic optimization algorithm, with its precursor developed by [62] along with mechanism elaborated by [63]. This algorithm is based upon the social and photosensory nature of fireflies, namely the bioluminescent flickering code in which they detect their mates or predators. The fireflies in the optimization context are considered as the candidate solutions whose attractiveness is defined with the quality of the objective value, usually metaphorically described as the brightness or intensity of the firefly lighting. This behavior is exploited by the algorithm, which guides less optimal fireflies toward the direction of brighter and more optimal fireflies. This leads to the build-up of a dynamic search pattern that searches for promising areas of the solution space and thus, FLA is effective to solve complex, nonconvex, and multi-mode optimization problems. The strength of the FLA is that it compensates between the search of new areas with the use of known high quality solutions by three main operations: action based on attractiveness, adaptation of the light intensity changes, and random movements to diversify the method [61,64].

During the initialization stage, the FLA randomly populates a multidimensional search space with a population of fireflies. This is vital to have the diversity of starting points that is essential to preventing premature convergence. The position of each of the fireflies in the space is normally generated by drawing uniformly in the specified ranges of each dimension, which is governed by the following formula:

$$X_{i,n} = L_n + \text{rand} \times (U_n - L_n) \tag{7}$$

where L_n and U_n are the lower and upper bounds for the nth dimension, and $X_{i,n}$ is the position of the ith firefly in that dimension. The term 'rand' represents a uniformly distributed random variable in the range [0, 1].

The light intensity, representing the fitness value, influences the attractiveness of a firefly. A key component of FLA is the attractiveness function, which decays with distance, as follows:

$$\beta(r) = \beta_0 \cdot e^{-\gamma r^2} \tag{8}$$

Here, β_0 is the maximum attractiveness at distance $r = 0$, γ is the light absorption coefficient controlling the decay, and r is the Euclidean distance between fireflies, as follows:

$$r_{ij} = |X_i - X_j| = \sqrt{\sum_{k=1}^{d} \left(X_{i,k} - X_{j,k} \right)^2} \tag{9}$$

A firefly i moves toward a brighter (more attractive) firefly j according to

$$X_i^{t+1} = X_i^t + \beta\left(r_{ij}\right) \cdot \left(X_j^t - X_i^t \right) + \alpha \cdot (\text{rand} - 0.5) \tag{10}$$

where α is a randomization parameter controlling the stochastic behavior of the algorithm. This movement integrates both exploitation (movement toward better solutions) and exploration (random perturbation).

The light intensity I of each firefly is related to the objective function $f(X)$, as follows:

$$I_i \propto f(X_i) \tag{11}$$

Better solutions emit stronger "light", making them more attractive to others. Over iterations, fireflies converge toward brighter peers, allowing the population to intensively search promising regions of the space.

To avoid premature convergence and local optima, the algorithm introduces random walks and can be enhanced using dynamic parameters such as adaptive ∝, chaotic maps, or Lévy flights to increase diversity. The FLA's simplicity and parallelism make it suitable for high-dimensional, nonlinear, and multi-modal optimization problems.

Despite its strengths, the standard FLA may suffer in convergence speed or stagnation in complex landscapes. Enhanced variants like the Adaptive Firefly Algorithm (AFA) or Hybrid FLA with Sine-Cosine or Genetic Operators have been developed to improve robustness and balance exploration–exploitation. Nonetheless, the foundational FLA remains widely applicable in engineering design, feature selection, scheduling, and medical image segmentation tasks.

To ensure fair comparison and reproducibility, both the FLA and the DOA were configured with consistent parameters throughout all experiments. FLA used a population size of 30, a maximum of 100 iterations, and control parameters $\alpha = 0.2$, $\beta_0 = 1$, and $\gamma = 1$. For DOA, the population size was also set to 30 with 100 iterations. Probability-based parameters were P = 0.7 and Q = 0.3, with dynamic adjustment factors $\beta_1 \sim U(-2, 2)$ and $\beta_2 \sim U(-1, 1)$. These values were selected based on empirical trials and consistency with prior studies.

In order to facilitate reproducibility and transparency, we consider the sensitivity of the hyper parameters in the metaheuristic optimization. The chosen population size (30) and maximum iterations (100) were based on the previous benchmark works on the medical image analysis and feature selection problems. When a subset of the training data was used to perform some preliminary experiments, it was found that growth in population size beyond 30 and iterations beyond 100 did not deliver improvements in significance (<0.5% increase in accuracy) but increased too much in the amount of computation being performed. On the other hand, it yielded premature convergence when the values were low. Therefore, the selected optimal setup is an economical balance between the performance effectiveness, which appears in the biomedical optimization articles of similar applications. Further research can implement a complete sensitivity analysis: ablation or grid-based search.

2.5. Explainable AI

In high-stakes areas like breast cancer detection, where open ness and trust are crucial, XAI has become an essential feature of medical picture analysis. Despite their impressive accuracy, traditional DL models frequently operate as "black boxes," providing no insight into the decision-making process. In therapeutic settings, where comprehension of the reasoning behind a prediction is just as crucial as the forecast itself, this opacity presents a substantial obstacle to adoption. Grad-CAM is one gradient-based visualization approach that has been created to improve model interpretability in order to meet this difficulty. By emphasizing the prominent areas of an image that have the most impact on the model's prediction, Grad-CAM offers visual explanations that promote diagnostic validation and foster clinician trust [65]. Calculating the gradient of the class score y^c (i.e., the output of a target class c) in relation to the feature mappings Ak of a convolutional layer is how Grad CAM operates mathematically. For every feature map k, the important weight α_k^c is determined as follows:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{12}$$

The gradient of the class score with regard to the activation at spatial position (i, j) is ∂y^c, ∂A_{ij}^k, where Z is the total number of pixels in the feature map A^k. After a weighted combination of the feature maps, the final Grad-CAM heatmap $L_{Grad-CAM}^c$ is produced.

Only the features that have a positive impact on the class of interest are kept after a ReLU activation.

$$L^{c}_{Grad-CAM} = \text{ReLU}\left(\sum_{k} \alpha^{c}_{k} A^{k}\right) \tag{13}$$

The regions that most influence the model's choice are graphically shown by superimposing this heatmap, which has been up sampled to the original picture quality. For this investigation, the outputs of the improved MobileNet models (both FLA-enhanced and DOA-enhanced) were subjected to Grad-CAM on three datasets: the fusion dataset, Breast Ultrasound, and MIAS. This allowed us to visualize how the model interprets various pathological features such as masses and calcifications, thereby validating the clinical relevance of predictions. The integration of Grad-CAM not only augments diagnostic performance but also aligns with the growing demand for ethically responsible and explainable AI in healthcare applications [66–68].

3. Experimental Setup

All the experiments were performed on the Google Colab Pro platform that provided the required computational means of the deep learning problem a Tesla T4 GPU (16 GB VRAM), 24 GB of RAM, and dual-core Intel Xeon processor. The time taken to train a fold was between 12 min and 18 min depending with the size of the databases and complexity of the model. The ease of accessibility, including to those arriving at the project without high-end local systems, as well as scalability was guaranteed by this cloud-based infrastructure. We used the following four common measures of performance to assess the performance provided by the proposed framework of breast cancer classification: accuracy, precision, recall, and F1-score. Accuracy shows the percentage of correctly predicted cases (true positive and true negative) out of all cases, whereas precision shows the percentage of true positively predicted cases out of all the predicted cases or cases that were given the status of positive. Recall (or sensitivity) measures how well a model can find the reality with respect to the positives, and F1-score, which acts as the harmonic mean of precision and recall, is a good metric to be employed to an imbalanced dataset [69]. Although the current research widely uses these measurements, their soundness is based on the pioneering works on information retrieval such as the definitions and analysis framework worked out by [70] and the exhaustive principles of evaluations presented in [71], as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

$$\text{F1}-\text{score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

where

- TP (True Positives): Number of correctly predicted positive cases;
- TN (True Negatives): Number of correctly predicted negative cases;
- FP (False Positives): Number of negative cases incorrectly predicted as positive;
- FN (False Negatives): Number of positive cases incorrectly predicted as negative.

These well-established formulas provide a comprehensive view of model performance, particularly in imbalanced medical datasets where relying on accuracy alone may be

misleading. They also ensure consistency in evaluating model robustness across various benchmark datasets used in this study.

4. Results and Discussion

This section evaluates the effectiveness of the proposed explainable deep learning framework in achieving its primary objective: enhancing breast cancer classification performance while preserving interpretability across different imaging modalities.

4.1. Performance of Hybrid Models in the MIAS Dataset

The MIAS dataset has been applied as an initial benchmark to compare the classification performance of different types of deep learning models within the scope of the mammographic images due to its diversity in terms of imaging conditions as well as pathological cases. Table 2 indicates that MobileNet showed the best results with the highest accuracy (93.76%) and equal precision, recall, and F1-score (0.9376), which mainly proves its high efficiency when it comes to feature extraction despite having few data. Xception and DenseNet were the next similar architectures with robust but slightly less accuracy (88.9% and 88.4%, respectively), though, such models as VGG16, ResNet50, EfficientNet, and the custom CNN fared so badly that the overall accuracy of VGG16 and ResNet50 made only 54.09% each. These results state the security and applicability of MobileNet on mammographic image classification and place it as an encouraging selection to be progressed and used in practice further.

Table 2. Assessment of Various Models' Performance on the MIAS Dataset.

Model	Accuracy	Precision	Recall	F1-Score
MobileNet	0.937587	0.937604	0.937587	0.937594
Xception	0.889043	0.889281	0.889043	0.888823
DenseNet	0.884882	0.885869	0.884882	0.884459
CNN	0.610264	0.610927	0.610264	0.610538
EfficientNet	0.540915	0.292589	0.540915	0.379761
VGG16	0.540915	0.292589	0.540915	0.379761
ResNet50	0.540915	0.292589	0.540915	0.379761

To further improve the classification performance on the MIAS dataset, two hybrid models were developed by integrating optimization algorithms with the MobileNet architecture: one using the DOA and the other employing the FLA. These bio-inspired optimizers were used to enhance the model's parameter tuning and convergence behavior. The evaluation of both optimized models was conducted using standard classification metrics, supported by confusion matrices and classification reports.

The confusion matrix of the MobileNet model optimized with DOA is shown in Figure 9. It reveals that 370 benign and 309 malignant cases were properly categorized, whereas 20 benign and 22 malignant instances were misclassified. Table 3 displays the relevant classification report, which indicates a test accuracy of 94.17% with precision, recall, and F1-score values all centered around 0.94. These outcomes demonstrate how well DOA directs the optimization process, enabling the model to continue delivering accurate and balanced performance across both classes.

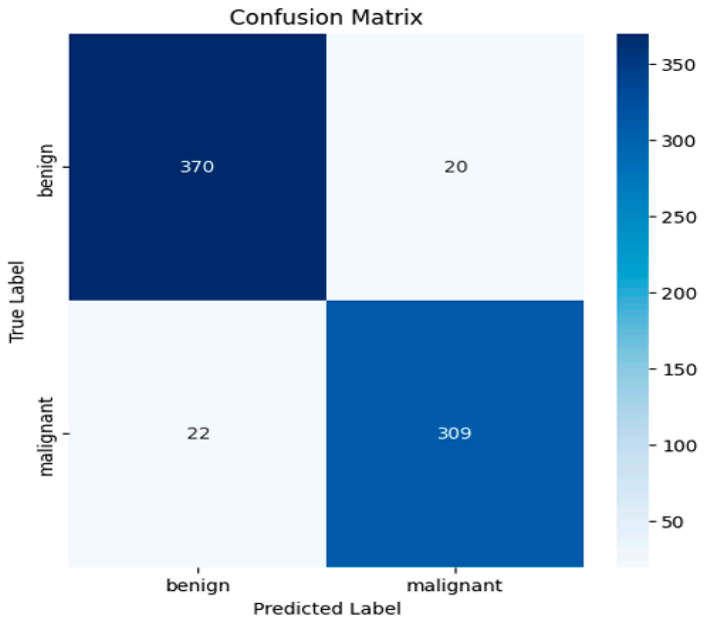


Figure 9. Confusion Matrix of Optimized MobileNet with DOA on MIAS Dataset.

Table 3. Assessment of Various Models' Performance on the MIAS Dataset.

Class	Precision	Recall	F1-Score
Benign	0.94	0.95	0.95
Malignant	0.94	0.93	0.94
Overall	0.94	0.94	0.94
Accuracy		**0.9417**	

In comparison, the model optimized with FLA achieved even better results. As shown in Figure 10, it correctly classified 382 benign and 325 malignant images, while only misclassifying 8 benign and 6 malignant cases. The comprehensive classification metrics are shown in Table 4, which reports a test accuracy of 98.06% with precision, recall, and F1-score values for both classes exceeding 0.98. These findings demonstrate how effectively FLA improves the model's capacity to generalize and accurately collect pertinent characteristics.

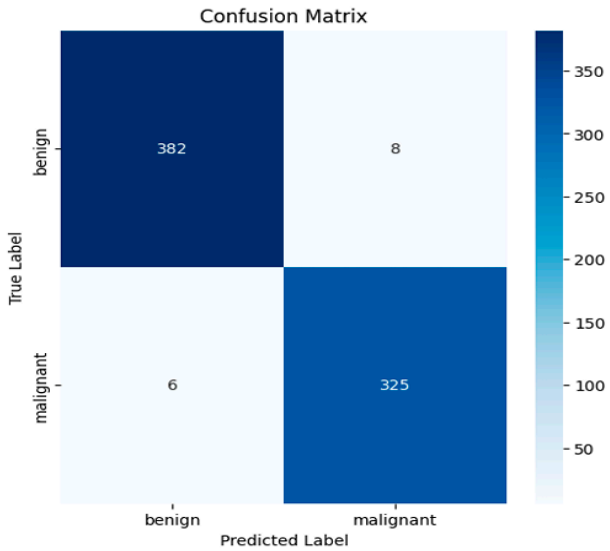


Figure 10. Optimized MobileNet with the FLA Confusion Matrix on the MIAS Dataset.

Table 4. Assessment of Various Models' Performance on the MIAS Dataset.

Class	Precision	Recall	F1-Score
Benign	0.98	0.98	0.98
Malignant	0.98	0.98	0.98
Overall	0.98	0.98	0.98
Accuracy		**0.9806**	

The FLA-optimized model outperformed the DOA-optimized version in all metrics, confirming its superior ability to fine-tune the learning process. The fusion of MobileNet with FLA provided the highest level of diagnostic accuracy, making it a highly effective solution for breast cancer detection based on mammographic images.

Figure 11 illustrates the AUC-ROC curve for the FLA-optimized MobileNet model applied to the MIAS dataset. The curve demonstrates excellent separability between the benign and malignant classes, with the model achieving an area under the curve (AUC) of 0.9982. This near-perfect AUC indicates a highly effective classifier with outstanding discriminative ability, where the true positive rate remains consistently high across a wide range of false positive rates. The curve's steep ascent and close proximity to the top-left corner of the plot reflect minimal trade-off between sensitivity and specificity, further validating the model's robustness and reliability in clinical breast cancer detection tasks.

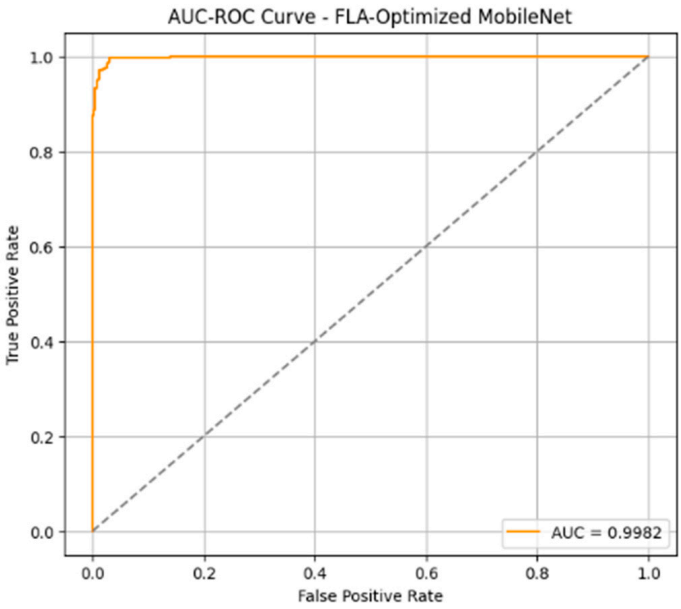


Figure 11. AUC-ROC of the FLA-Optimized MobileNet.

4.2. Performance of Hybrid Models in the Breast Ultrasound Images Dataset

The second community assessment was conducted on the Breast Ultrasound Images Dataset, which is known to be complex to diagnose due to noise, low contrast, and variation in the images. Table 5 shows that MobileNet performed the best of all models, indicating the highest level of accuracy (93.53%), with all other measures, F1-score (0.8982) and recall (0.8929), also performing well, signifying that it was the most suitable model in processing ultrasound information. DenseNet and Xception demonstrated competitive results as well, with the former displaying an excellent precision (0.9221) but poor recall and the latter performing similarly in accuracy (90.49%) but different in precision–recall trade-offs. Although ResNet50 has a high accuracy (94.12%), it's very low recall (0.3809), however, is of significant concern since this leads to a high percentage of false negatives. Custom CNN and EfficientNet resulted in the lowest performance, especially on recall and F1-score. In

general, these results indicate the versatile and stable behavior of MobileNet in the realm of ultrasound-based breast cancer detection because of its moderately high results in the most important indexes of effectiveness.

Table 5. Assessment of Various Models' Performance on the Breast Ultrasound Image Dataset.

Model	Accuracy	Precision	Recall	F1-Score
MobileNet	0.935361	0.903614	0.892857	0.898204
DenseNet	0.927757	0.922078	0.845238	0.881988
Xception	0.904943	0.904110	0.785714	0.840764
VGG16	0.904943	0.855422	0.845238	0.850299
ResNet50	0.794677	0.941176	0.380952	0.542373
CNN	0.711027	0.633333	0.226190	0.333333
EfficientNet	0.680608	0.340000	0.500000	0.680000

In order to assess how well the hybrid optimization techniques performed on the Breast Ultrasound Images Dataset, two improved models were created by combining the MobileNet architecture with FLA and DOA, respectively. These optimizers were employed to improve the network's parameter configuration and convergence behavior, aiming for greater diagnostic accuracy and generalizability.

Figure 12 shows the confusion matrix of the Optimized MobileNet with DOA, where the model correctly classified 170 benign and 77 malignant cases, while misclassifying 9 benign and 7 malignant instances. The corresponding classification report in Table 6 indicates a test accuracy of 94.22%. All three metrics—precision, recall, and F1-score—were 0.95 for the benign class and 0.91 for the malignant class. The accuracy, recall, and F1-score averages were all 0.93, indicating a good trade-off between sensitivity and specificity. This demonstrates that the DOA-assisted model is effective at distinguishing between cancerous and non-cancerous ultrasound images, though with slightly more errors in identifying malignant cases.

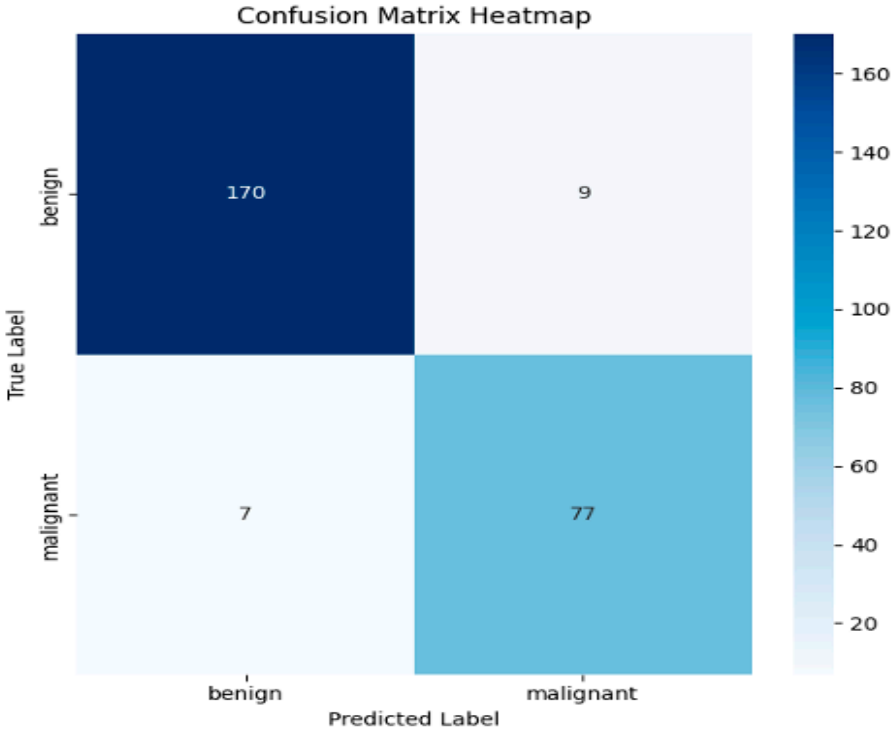


Figure 12. Confusion Matrix of Optimized MobileNet with DOA on the Breast Ultrasound Images Dataset.

Table 6. Classification Report of Optimized MobileNet with DOA on the Breast Ultrasound Images Dataset.

Class	Precision	Recall	F1-Score
Benign	0.95	0.95	0.95
Malignant	0.91	0.91	0.91
Overall	0.93	0.93	0.93
Accuracy		**0.9422**	

In comparison, Figure 13 presents the confusion matrix of the Optimized MobileNet with FLA, which further improves classification accuracy. Only 5 benign and 7 malignant cases were incorrectly classified by the model, whereas 174 benign and 77 malignant cases were accurately predicted. With precision, recall, and F1-score values of 0.96, 0.97, and 0.97 for benign cases and 0.94, 0.92, and 0.93 for malignant instances, the test accuracy was 95.44%, as shown in Table 7. A well-balanced and extremely dependable model that could reduce both false positives and false negatives was demonstrated by the overall average scores of 0.95 for all criteria.

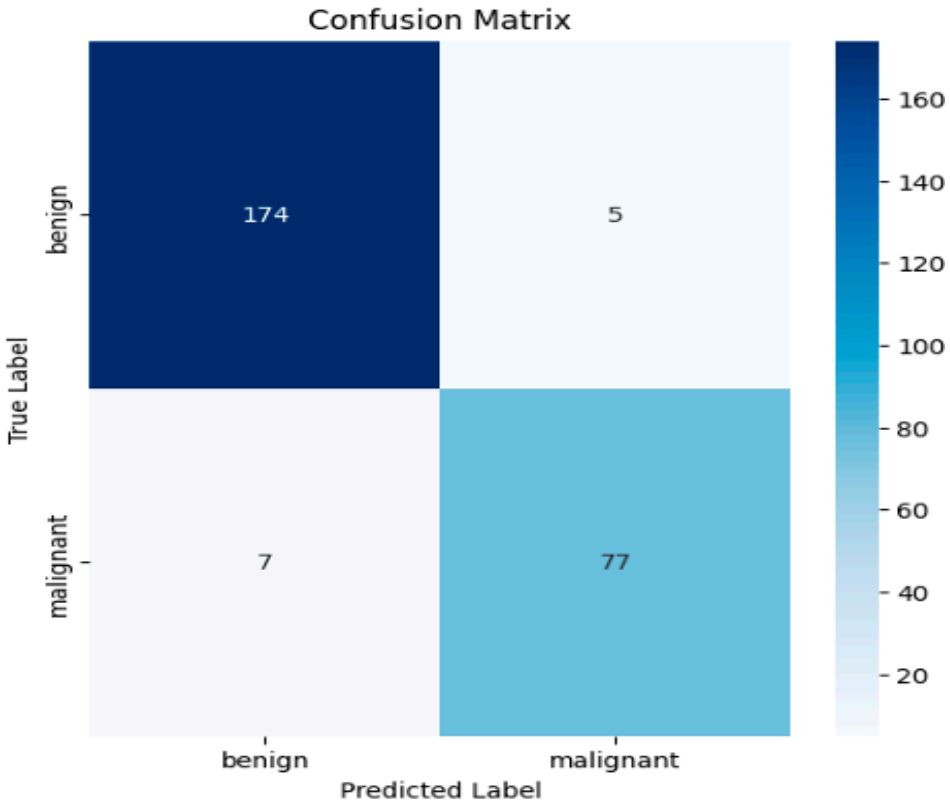


Figure 13. Confusion Matrix of Optimized MobileNet with FLA on the Breast Ultrasound Images Dataset.

Table 7. Classification Report of Optimized MobileNet with FLA on the Breast Ultrasound Images Dataset.

Class	Precision	Recall	F1-Score
Benign	0.96	0.97	0.97
Malignant	0.94	0.92	0.93
Overall	0.95	0.94	0.95
Accuracy		**0.9544**	

The algorithm correctly predicted 174 benign and 77 malignant cases, whereas only five benign and seven malignant cases were misclassified. As indicated in Table 6, the test accuracy was 95.44% with precision, recall, and F1-score values of 0.96, 0.97, and 0.97 for benign cases and 0.94, 0.92, and 0.93 for malignant examples. The aggregate average scores of 0.95 for all criteria showed that the model was very reliable and well-balanced, capable of reducing both false positives and false negatives. These results confirm that the FLA-based optimization yields superior classification performance on the Breast Ultrasound Images Dataset, making it a more robust choice for breast cancer detection using ultrasound imaging. It is especially well-suited for clinical applications where diagnostic accuracy is essential due to its capacity to efficiently adjust the model's decision bounds and enhance generalization.

Figure 14 presents the AUC-ROC curve for the FLA-Optimized MobileNet model evaluated on the MIAS dataset. The model achieved an AUC score of 0.9678, indicating excellent discriminative capability in distinguishing between benign and malignant breast tissue. The curve demonstrates a strong trade-off between true positive and false positive rates, with the trajectory remaining well above the diagonal baseline, reflecting high sensitivity and specificity. This result confirms the model's robustness and diagnostic reliability, underscoring the effectiveness of Firefly Algorithm-based optimization in enhancing MobileNet's classification performance in mammographic image analysis.

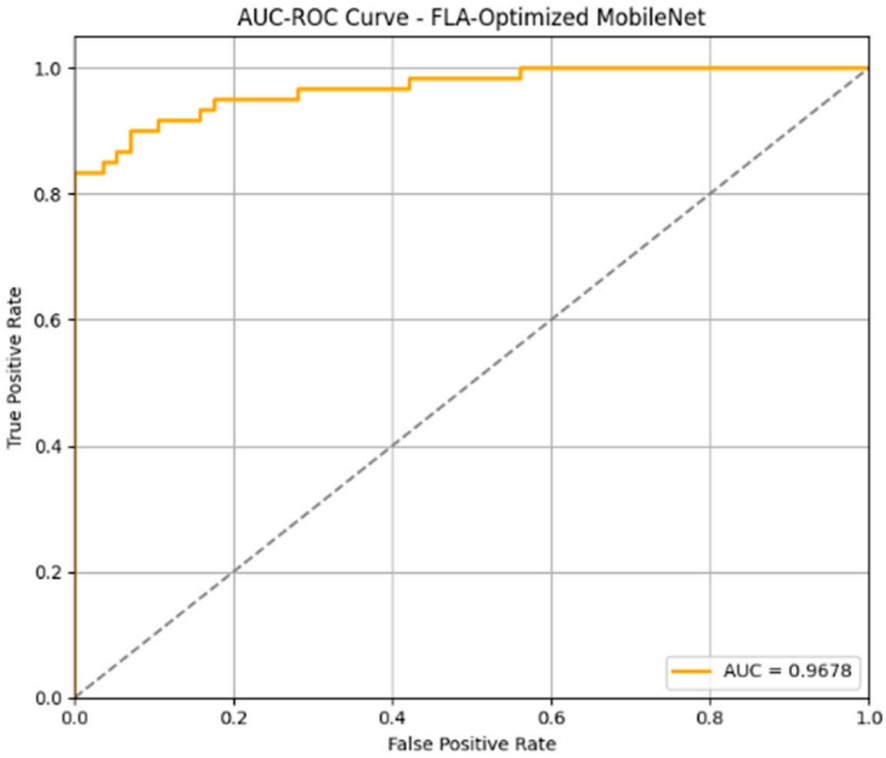


Figure 14. AUC-ROC Curve of the FLA-Optimized MobileNet.

4.3. Performance of Hybrid Models in the Fusion Dataset

The last stage of evaluation utilizes the Fusion Dataset, which unifies the MIAS mammographic set and Breast Ultrasound Images to bring more heterogeneity and resemble a more closely real-life diagnostic setting. Multimodal data make models hard to generalize across imaging variability. MobileNet also proved to be the strongest in all the metrics of accuracy (91.95%), precision (0.92), recall (0.92), and F1-score (0.92), as it is also lightweight in structure and the feature extraction is consequently effective due to depthwise separable convolutions (see Table 8). DenseNet and Xception also turned out

to be competitive, having scaled 85.91 and 85.41 percentages of accuracy, respectively, but they were not as balanced or consistent as MobileNet. VGG16 performed better (75.22% accuracy), whereas ResNet50, CNN, and EfficientNet were way behind, with EfficientNet performing particularly poorly in terms of precision (0.32), indicating a high ratio of false positives. The results confirm that MobileNet is effective and reliable in multimodal breast cancer classification, providing support for its potential application in real life concerning various diagnostic facilities.

Table 8. Performance Evaluation of Different Models on the Fusion Dataset.

Model	Accuracy	Precision	Recall	F1-Score
MobileNet	0.9195	0.92	0.92	0.92
DenseNet	0.8591	0.86	0.85	0.86
Xception	0.8541	0.85	0.85	0.85
VGG16	0.7522	0.75	0.75	0.75
ResNet50	0.6138	0.62	0.58	0.56
CNN	0.5849	0.59	0.58	0.59
EfficientNet	0.5635	0.32	0.56	0.41

To enhance classification performance, the MobileNet architecture was optimized using two bio-inspired metaheuristic algorithms: the DOA and FLA.

The confusion matrix for the Optimized MobileNet with DOA is displayed in Figure 15. Only 5 benign and 10 malignant samples were misclassified, while 443 benign and 337 malignant samples were properly predicted by the model. This model obtained a test accuracy of 98.96%, with precision, recall, and F1-score values of 0.98 or higher for both classes, as shown in Table 9. The DOA-enhanced model is well-suited for real-world diagnostic applications where input variability is widespread, since these results show that it performs robust and highly accurate classification on heterogeneous data.

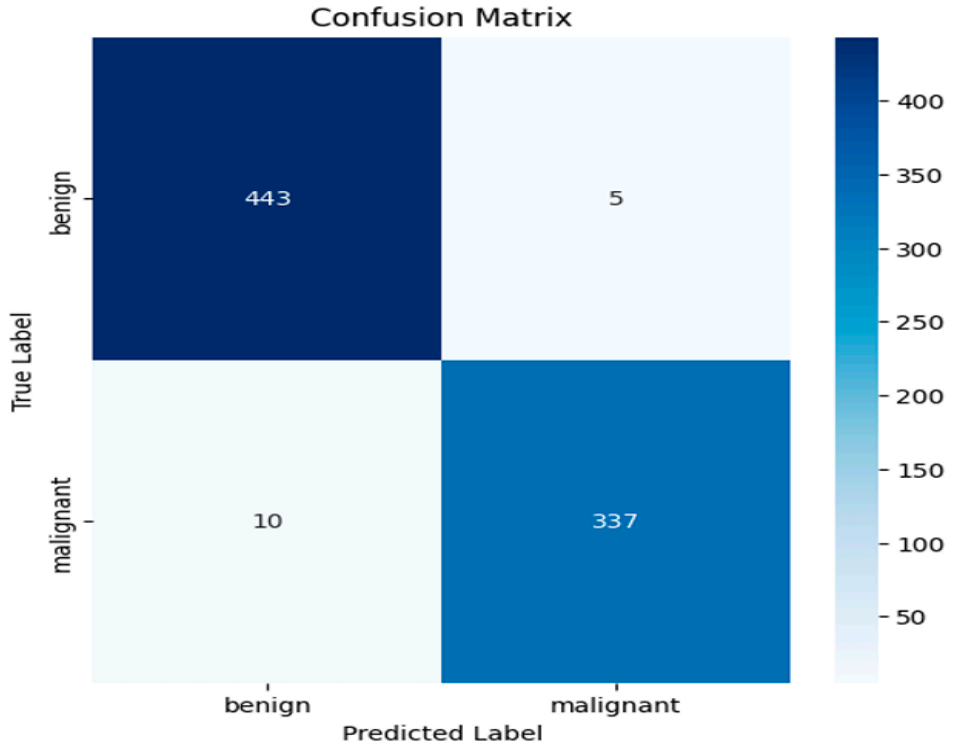


Figure 15. Confusion Matrix of Optimized MobileNet with DOA on the Fusion Dataset.

Table 9. Confusion Matrix of Optimized MobileNet with DOA on the Fusion Dataset.

Class	Precision	Recall	F1-Score
Benign	0.98	0.99	0.99
Malignant	0.98	0.97	0.97
Overall	0.98	0.98	0.98
Accuracy		**0.9896**	

In comparison, the performance of the Optimized MobileNet with FLA is shown in Figure 16. This model also exhibited strong classification ability, correctly identifying 443 benign and 336 malignant samples, while misclassifying 5 benign and 11 malignant cases. The accuracy is 98.70%, and the precision, recall, and F1-score values for both categories are consistently near 0.98, according to the relevant classification report in Table 10. While this performance is outstanding, it is slightly lower than the DOA-optimized counterpart in terms of overall accuracy and true positive predictions for malignant cases.

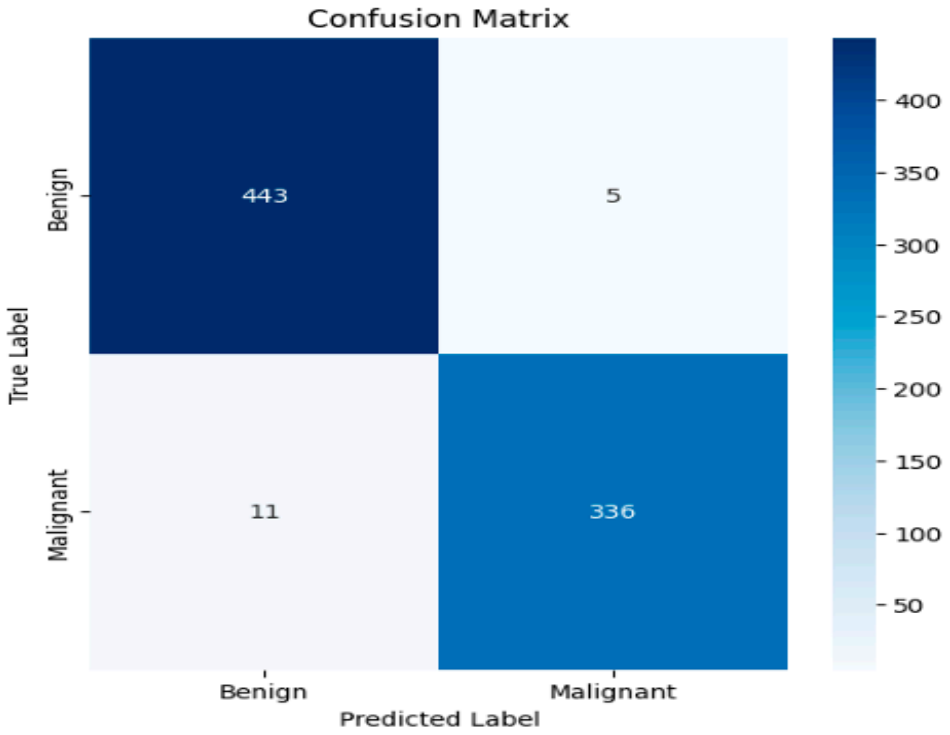


Figure 16. Confusion Matrix of the Optimized MobileNet with FLA on the Fusion Dataset.

Table 10. Classification Report for the Fusion Dataset Using the Optimized MobileNet with FLA.

Class	Precision	Recall	F1-Score
Benign	0.99	0.99	0.98
Malignant	0.97	0.97	0.97
Overall	0.98	0.98	0.98
Accuracy		**0.9870**	

Based on the comparative summary, the DOA-optimized MobileNet slightly outperformed the FLA-optimized version in all performance indicators. The final comparison study revealed that, in contrast to the FLA model's accuracy of 98.70%, the DOA model had the greatest overall accuracy (98.96%), precision (0.98), recall (0.98), and F1-score (0.98). These findings confirm that the use of the DOA provided a marginal yet valuable advantage in optimizing the MobileNet architecture for fusion-based breast cancer classification. This

demonstrates the potential of DOA as a superior optimization strategy when applied to complex multi-source medical image datasets.

Figure 17 illustrates the AUC-ROC curve for the DOA-Optimized MobileNet model applied to the fusion dataset. The curve demonstrates a strong classification performance, with the model achieving an Area Under the Curve (AUC) of 0.9940. This near-perfect score indicates the model's exceptional ability to distinguish between benign and malignant breast tumors. The curve rises sharply toward the top-left corner of the plot, reflecting a high true positive rate (sensitivity) and a low false positive rate across various threshold levels. Such performance underscores the model's robustness, making it highly suitable for real-world clinical applications where diagnostic precision is critical. The AUC value further supports the findings from the confusion matrix and classification report, affirming the DOA-optimized MobileNet's reliability and discriminative power in multimodal breast cancer detection.

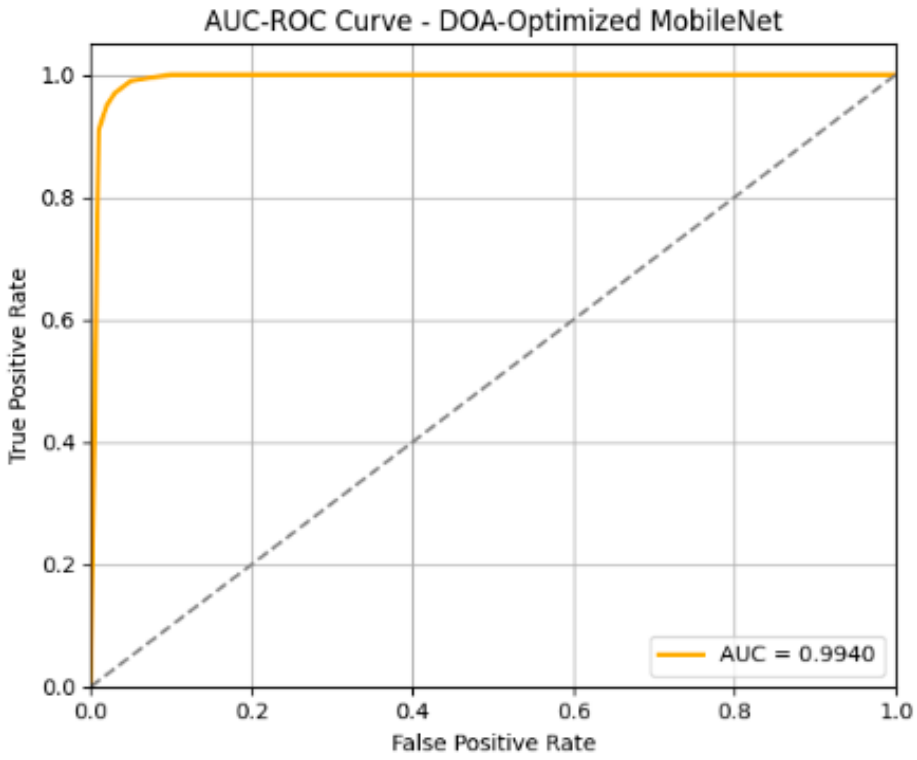


Figure 17. AUC-ROC Curve DOA-Optimized MobileNet.

A higher number of factors can judge the high speed of MobileNet than that of deep networks like ResNet50 and VGG16. First, the relatively small size and little variation in the datasets involved (especially MIAS and BUSI) make it favorable to use lightweight models that can be effectively generalized, and they do not overfit. The depthwise separable convolutions that MobileNet employs greatly limit the complexity of the model and, to the same extent, preserve the representational capacity, due to which it is more effective to use on small and high-noise medical imaging assignments. Finally, optimization algorithms (FLA and DOA) were used to test the influence of optimization on the performance of MobileNet; the results were further improved due to the optimization of parameters and enhanced convergence. Alternatively, more complicated models could have needed large amounts of data or greater tuning in order to perform optimally. These results are all similar regardless of the three datasets involved, giving importance to the potential of MobileNet in data-scant, diagnostic applications.

4.4. GRAD-CAM Explainability Analysis for Breast Cancer Identification

Gradient-weighted Class Activation Mapping (Grad-CAM) was applied in this research to visualize the part of input images that contributed the most to model classification decisions. These visualizations can improve the interpretability of models and provide insight into the inner states of CNNs and are of high importance in high-stakes applications like breast cancer diagnosis. Grad-CAM is helpful in giving an account of whether the model is attending to clinically relevant regions, which in a way promotes model transparency and confidence in AI-aided diagnosis systems.

Grad-CAM was used to estimate the spatial focus of the optimized MobileNets models wrapped in three imaging modalities, i.e., in mammography, in ultrasound, and in a combined fusion of both. In the mammographic case of MIAS (Figure 18), the FLA-optimized MobileNet model was quite accurate in the localization of a dense and irregular area that is often related to malignancy. Likewise, the presence of a heatmap in the ultrasound image (Figure 19) indicated a significant focus in proximity around the lesion boundary and echotexture within the internal area. In the fusion dataset (Figure 20), the DOA-optimized model exhibited the sharp activation on the tumor area with an apparent inhibition on the background noise, which suggests that the model can exploit the multimodal information in a very effective manner.

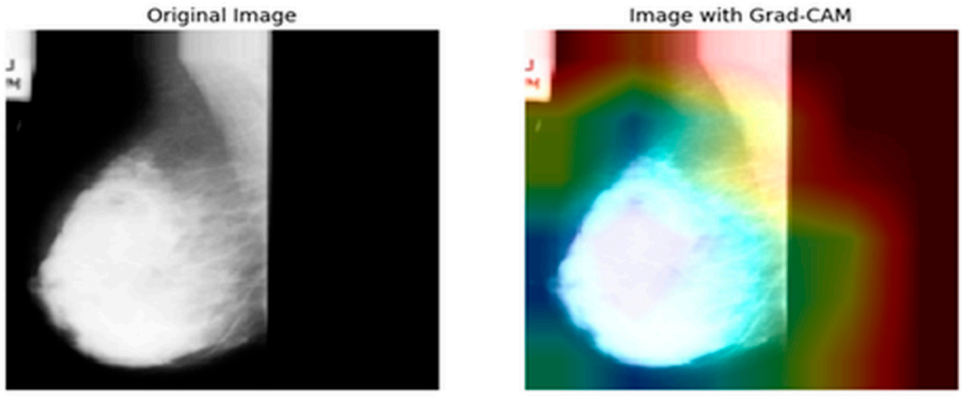


Figure 18. Grad-CAM visualization using Optimized MobileNet-FLA highlights clinically relevant lesion regions in a MIAS mammogram, confirming malignancy-focused model attention.

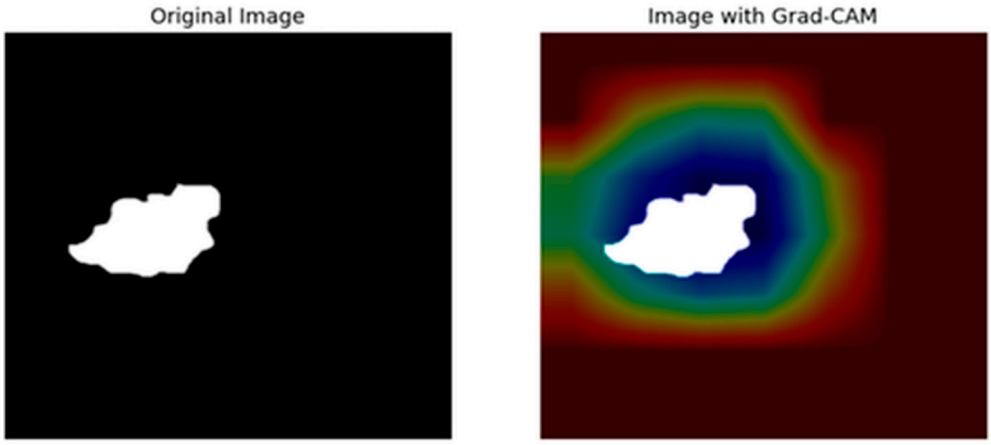


Figure 19. High-resolution Grad-CAM visualization on a Breast Ultrasound image using the Optimized MobileNet-FLA model. The left panel displays the original ultrasound image, and the right panel shows the Grad-CAM heatmap, emphasizing areas with the highest contribution to malignancy prediction.

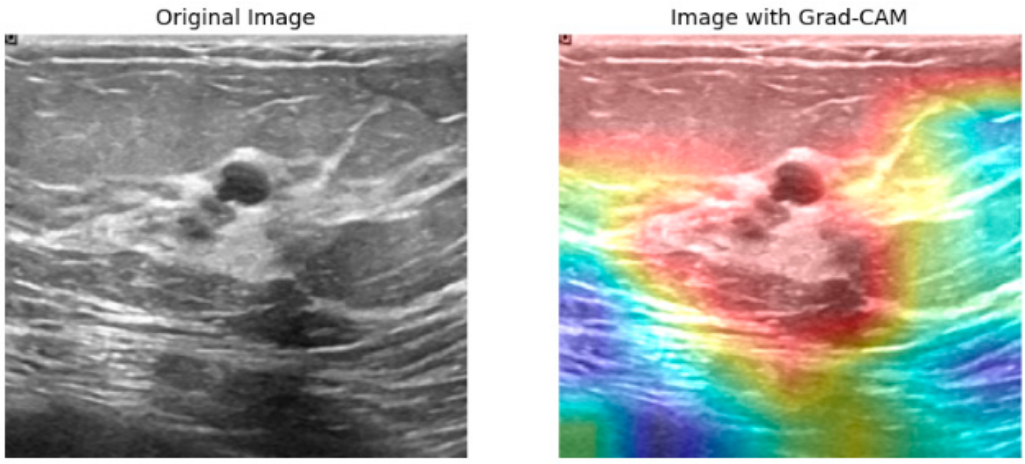


Figure 20. Grad-CAM visualization on a fused mammographic–ultrasound image using the Optimized MobileNet-DOA model. The heatmap demonstrates focused activation in the lesion area, illustrating the model's ability to leverage multimodal features for accurate malignancy prediction.

Although such visualizations indicate that the model focuses on clinically relevant areas, we admit that consideration of expert radiologist validation was beyond the scope of the given research, which weakens the overall findings concerning the matching of diagnoses. In subsequent studies, we intend to use qualitative scores and markings in certified radiologists to prove the relevance of the areas of interest in terms of clinical significance. The integration will enhance the claims of interpretability and put the model decision-making process in line with determined diagnostic standards.

4.5. Comparison Results

To assess the overall performance and generalizability of the proposed hybrid optimization framework, we conducted a comparative analysis across three datasets: the MIAS dataset, the Breast Ultrasound Images dataset, and the Fusion dataset that combines both sources. This comprehensive evaluation enables a deeper understanding of how different models, particularly the baseline MobileNet and its optimized variants using Firefly Law Algorithm FLA and DOA, perform across diverse imaging conditions and modalities.

Figure 21 demonstrates an overall comparative evaluation of several deep learning models on three benchmark datasets: MIAS (mammographic images), the Image dataset of breast ultrasound, and the combined dataset that includes a combination of both modalities. The figure uses a grouped bar chart with color coding to identify an area of confusion because visual clarity is enhanced by using green to represent the MIAS dataset, light blue to represent ultrasound, and steel blue to represent the fusion dataset.

Optimization of the mobileNet designs, with the use of metaheuristic techniques (FLA and DOA), were the best-performing designs in terms of achieving better accuracy for all datasets. The FLA-optimized MobileNet achieved a 98.06% accuracy on the MIAS dataset compared with its DOA version (94.17%) and the original MobileNet (92.93%). Regarding the ultrasound dataset, the same FLA version performed better with 95.44%, followed by DOA (94.22%), which is significantly better than other conventional architectures such as VGG16, ResNet50, or CNN. The best performance was noted on the fusion dataset, where the DOA-optimized MobileNet outperformed the FLA-enhanced one with an MAE of 98.96 percent compared to its 98.70 percent, respectively. These findings show that hybrid techniques of optimization have high levels of resilience and applicability in heterogeneous medical data. The stable high results on all the datasets confirm the effectiveness of the mobile-based frameworks, especially after adding bio-inspired optimization strategies, to improve the accuracy of breast cancer classification. These findings confirm that opti-

mization techniques significantly enhance model performance, with the Fusion dataset delivering the highest accuracies due to the rich, diverse features it contains, as presented in Table 11. The DOA-optimized MobileNet model, in particular, proved most effective under these complex conditions.

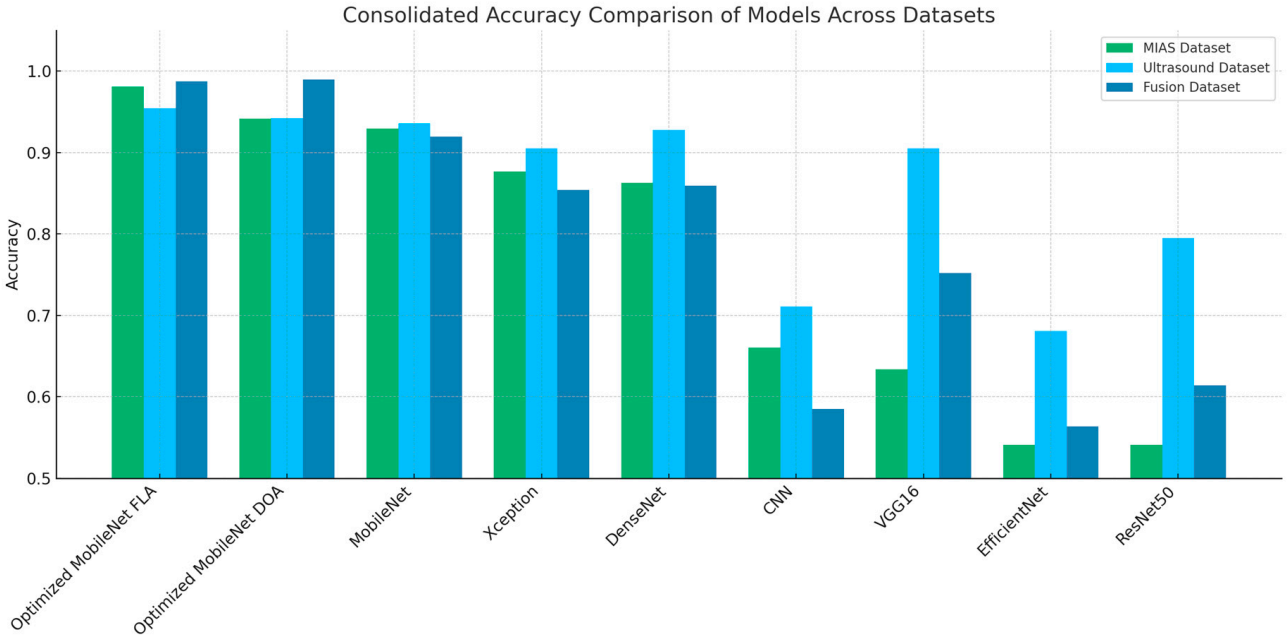


Figure 21. Consolidated Accuracy Comparison of Deep Learning Models on MIAS, Breast Ultrasound, and Fusion Datasets.

Table 11. Performance Comparison of Best Models Across All Datasets.

Model	Dataset	Accuracy	Precision	Recall	F1-Score
Optimized MobileNet FLA	MIAS	0.9806	0.9800	0.9800	0.9800
Optimized MobileNet DOA	MIAS	0.9417	0.9400	0.9400	0.9400
MobileNet	MIAS	0.9376	0.9376	0.9376	0.9376
Optimized MobileNet FLA	Breast Ultrasound	0.9544	0.9500	0.9500	0.9500
Optimized MobileNet DOA	Breast Ultrasound	0.9422	0.9400	0.9400	0.9400
MobileNet	Breast Ultrasound	0.9354	0.9036	0.8929	0.8982
Optimized MobileNet DOA	Fusion	0.9896	0.9800	0.9800	0.9800
Optimized MobileNet FLA	Fusion	0.9870	0.9800	0.9800	0.9800
MobileNet	Fusion	0.9195	0.9800	0.9200	0.9200

According to Table 12, a thorough comparative study was performed to compare the suggested hybrid optimization framework with the breast cancer detection methodologies involving ensemble learning, transfer learning, explainable AI, and metaheuristic optimization. As one of the notable works, Teoh et al. [34] exploit an ensemble of pretrained CNNs to detect micro calcification with 93.05% accuracy; Sarwar et al. [36] utilize an ensemble of pretrained CNNs with the idea of few-shot learning combined with transfer learning, where they report 95.6% and 0.970 AUC; Raghavan et al. [37] make use of attention-guided Grad-CAM and an ensemble of CNNs to detect microcalcification, with the results of 9 A simplified CNN provided 94.23% and 96.67% on MIAS and fusion dataset, respectively [42]; and ResNet50, with the Improved Marine Predators Algorithm, performed skillfully at 98.88% [49]. Naas et al. [51] incorporated DeepLabV3+, autoencoders, and Grad-CAM and achieved 97.4% on ultrasound images. Our optimized MobileNet models performed competitively or better in comparison when tested with MIAS with FLA (98.06% accuracy,

surpassing the results of Alhsnony et al. [42]), on ultrasound with a strong F1-score and high interpretability (95.44%, beating the results of Naas et al. [51]), and on the fusion dataset with DOA (98.96% vs. all the previously referenced methods). The results discussed indicate the strength, applicability, and broader usefulness of the suggested framework applicable to various imaging systems.

Table 12. Comparative Summary Table of Related Works and Proposed Method.

References	Model/Method	Dataset	Results
Teoh et al., 2024 [34]	Ensemble (AlexNet, GoogLeNet, VGG16, ResNet-50), Transfer Learning	Mammograms	93.05% (microcalcifications), 88.59% (normal)
Sarwar et al., 2025 [36]	Few-Shot Learning + Transfer Learning (Relation Network)	Mammograms (Few-shot)	Accuracy: 95.6%, AUC: 0.970
Raghavan et al., 2024 [37]	Attention-Guided Grad-CAM+Ensemble CNNs	Infrared Thermograms (DMR)	Accuracy: 98.04%, AUC: 0.97
Alhsnony et al., 2024 [42]	Custom Simplified CNN	MIAS, DDSM, Fusion	Accuracy: 94.23%(MIAS), 95.53% (DDSM), 96.67% (Fusion)
Houssein et al., 2022 [49]	ResNet50 + Improved Marine Predators Algorithm (IMPA)	MIAS, CBIS-DDSM	Accuracy: up to 98.88%
Naas et al., 2025 [51]	DeepLabV3++ Autoencoder + Grad-CAM (XAI)	Breast Ultrasound	Accuracy: 97.4%, Dice: 0.981
Proposed Method	**Optimized MobileNet + FLA**	**MIAS Dataset**	**Precision/Recall/F1: 0.98. Accuracy: 98.06%**
Proposed Method	**Optimized MobileNet + FLA**	**Breast Ultrasound Images**	**Accuracy:95.44%, Precision/Recall/F1: 0.95**
Proposed Method	**Optimized MobileNet + DOA**	**Fusion Dataset**	**Accuracy: 98.96% Precision/Recall/F1: 0.98**

5. Conclusions

Breast cancer remains one of the most critical public health challenges globally, necessitating the advancement of accurate and interpretable diagnostic methodologies. This study introduced an explainable deep learning framework for breast cancer classification, integrating the computational efficiency of the MobileNet architecture with the optimization capabilities of two bio-inspired algorithms: the DOA and the FLA. The proposed approach follows a systematic pipeline—initial model selection, integration of metaheuristic optimization, and multimodal evaluation—aimed at enhancing diagnostic accuracy while preserving model transparency.

Extensive experiments were conducted on three benchmark datasets: the MIAS mammography dataset, the Breast Ultrasound Images dataset, and a combined Fusion dataset. The FLA-optimized MobileNet achieved superior results on the MIAS dataset, with an accuracy of 98.06%, precision, recall, and F1-score all reaching 0.98, and an AUC of 0.9982, indicating excellent class separability. On the ultrasound dataset, the same model attained an accuracy of 95.44%, with F1-scores of 0.97 and 0.93 for benign and malignant classes, respectively. On the multimodal Fusion dataset, the DOA-optimized MobileNet outperformed all models, achieving an accuracy of 98.96%, and precision, recall, and F1-score values of 0.98, demonstrating robust generalization across diverse imaging modalities.

Future research directions include expanding the framework to incorporate additional imaging modalities such as MRI and thermographic scans, integrating attention-based mechanisms to further enhance interpretability, and implementing and validating the system in real-time clinical settings to assess its practical diagnostic utility.

Author Contributions: The first author, M.A.M., conducted experiments, wrote the manuscript, and executed the software process. The second author, O.A.E., was responsible for supervision and correcting the direction of the work. The third author, E.S., was responsible for project administration and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in this study are openly available in the Apollo-University of Cambridge Repository at https://doi.org/10.17863/CAM.105113 (accessed on 25 February 2025), and in the Data in Brief journal repository at https://doi.org/10.1016/j.dib.2019.104863 (accessed on 20 February 2025).

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

References

1. Tan, Y.; Sim, K.-S.; Ting, F.F. Breast cancer detection using convolutional neural networks for mammogram imaging system. In Proceedings of the 2017 International Conference on Robotics, Automation and Sciences (ICORAS), Melaka, Malaysia, 27–29 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–5.
2. López-Cabrera, J.D.; Rodríguez, L.A.L.; Pérez-Díaz, M. Classification of breast cancer from digital mammography using deep learning. *Intel. Artif.* **2020**, *23*, 56–66. [CrossRef]
3. Pang, J.; Ding, N.; Liu, X.; He, X.; Zhou, W.; Xie, H.; Feng, J.; Li, Y.; He, Y.; Wang, S.; et al. Prognostic Value of the Baseline Systemic Immune-Inflammation Index in HER2-Positive Metastatic Breast Cancer: Exploratory Analysis of Two Prospective Trials. *Ann. Surg. Oncol.* **2025**, *32*, 750–759. [CrossRef]
4. World Health Organization. Breast Cancer. 2023. Available online: https://www.who.int/news-room/fact-sheets/detail/breast-cancer (accessed on 13 March 2024).
5. Han, L.; Yin, Z. A hybrid breast cancer classification algorithm based on meta-learning and artificial neural networks. *Front. Oncol.* **2022**, *12*, 1042964. [CrossRef]
6. Halim, A.; Andrew, A.M.; Yasin, M.N.M.; Rahman, M.A.A.; Jusoh, M.; Veeraperumal, V.; Rahim, H.A.; Illahi, U.; Karim, M.K.A.; Scavino, E. Existing and emerging breast cancer detection technologies and its challenges: A review. *Appl. Sci.* **2021**, *11*, 10753. [CrossRef]
7. Vijayarajeswari, R.; Parthasarathy, P.; Vivekanandan, S.; Basha, A.A. Classification of mammogram for early detection of breast cancer using svm classifier and hough transform. *Measurement* **2019**, *146*, 800–805. [CrossRef]
8. Sarosa, S.J.A.; Utaminingrum, F.; Bachtiar, F.A. Mammogram breast cancer classification using gray-level co-occurrence matrix and support vector machine. In Proceedings of the 2018 International Conference on Sustainable Information Engineering and Technology (SIET), Malang, Indonesia, 10–12 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 54–59.
9. Ma, X.; Cheng, H.; Hou, J.; Jia, Z.; Wu, G.; Lü, X.; Li, H.; Zheng, X.; Chen, C. Detection of breast cancer based on novel porous silicon Bragg reflector surface-enhanced Raman spectroscopy-active structure. *Chin. Opt. Lett.* **2020**, *18*, 051701. [CrossRef]
10. Torres, R.E.O.; Gutiérrez, J.R.; Jacome, A.G.L. Neutrosophic-based machine learning techniques for analysis and diag nosis the breast cancer. *Int. J. Neutrosophic Sci. (IJNS)* **2023**, *21*, 1.
11. Yadav, A.R.; Vaegae, N.K. Development of an early prediction system for breast cancer using machine learning techniques. In Proceedings of the 2023 International Conference on Next Generation Electronics (NEleX), Vellore, India, 14–16 December 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
12. Resch, D.; Gullo, R.L.; Teuwen, J.; Semturs, F.; Hummel, J.; Resch, A.; Pinker, K. Ai-enhanced mammography with digital breast tomosynthesis for breast cancer detection: Clinical value and comparison with human performance. *Radiol. Imaging Cancer* **2024**, *6*, e230149. [CrossRef]
13. Shaaban, S.M.; Nawaz, M.; Said, Y.; Barr, M. Anefficient breast cancer segmentation system based on deep learning techniques. *Eng. Technol. Appl. Sci. Res.* **2023**, *13*, 12415–12422. [CrossRef]
14. Gurumoorthy, R.; Kamarasan, M. Breast cancer classification from histopathological images using future search optimization algorithm and deep learning. *Eng. Technol. Appl. Sci. Res.* **2024**, *14*, 12831–12836. [CrossRef]
15. Shen, D.; Wu, G.; Suk, H.-I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [CrossRef]
16. Le Cun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *7553*, 436–444. [CrossRef]

17. Islam, U.; Al-Atawi, A.A.; Alwageed, H.S.; Mehmood, G.; Khan, F.; Innab, N. Detection of renal cell hydronephrosis in ultrasound kidney images: A study on the efficacy of deep convolutional neural networks. *PeerJ Comput. Sci.* **2024**, *10*, e1797. [CrossRef]
18. Razavian, A.S.; Azizpour, H.; Sullivan, J.; Carlsson, S. Cnn features off-the-shelf: An astounding baseline for recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 806–813.
19. Penatti, O.A.; Nogueira, K.; Santos, J.A.D. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 44–51.
20. Tajbakhsh, N.; Shin, J.Y.; Gurudu, S.R.; Hurst, R.T.; Kendall, C.B.; Gotway, M.B.; Liang, J. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans. Med. Imaging* **2016**, *35*, 1299–1312. [CrossRef]
21. Nasir, I.M.; Raza, M.; Shah, J.H.; Khan, M.A.; Nam, Y.-C.; Nam, Y. Improved shark smell optimization algorithm for human action recognition. *Comput. Mater. Contin* **2023**, *76*, 2667–2684.
22. Nasir, I.M.; Raza, M.; Ulyah, S.M.; Shah, J.H.; Fitriyani, N.L.; Syafrudin, M. Enga: Elastic net-based genetic algorithm for human action recognition. *Expert Syst. Appl.* **2023**, *227*, 120311. [CrossRef]
23. Nasir, I.M.; Raza, M.; Shah, J.H.; Wang, S.-H.; Tariq, U.; Khan, M.A. Harednet: A deep learning based architecture for autonomous video surveillance by recognizing human actions. *Comput. Electr. Eng.* **2022**, *99*, 107805. [CrossRef]
24. Wang, L.; Jiang, S.; Jiang, S. A feature selection method via analysis of relevance, redundancy, and interaction. *Expert Syst. Appl.* **2021**, *183*, 115365. [CrossRef]
25. Tariq, J.; Alfalou, A.; Ijaz, A.; Ali, H.; Ashraf, I.; Rahman, H.; Armghan, A.; Mashood, I.; Rehman, S. Fast intra mode selection in hevc using statistical model. *Comput. Mater. Contin* **2022**, *70*, 3903–3918. [CrossRef]
26. Gao, Y.; Wang, C.; Wang, K.; He, C.; Hu, K.; Liang, M. The effects and molecular mechanism of heat stress on spermatogenesis and the mitigation measures. *Syst. Biol. Reprod. Med.* **2022**, *68*, 331–347. [CrossRef] [PubMed]
27. Shafipour, M.; Fadaei, S. Particle distance rank feature se lection by particle swarm optimization. *Expert Syst. Appl.* **2021**, *185*, 115620. [CrossRef]
28. Nasir, I.M.; Rashid, M.; Shah, J.H.; Sharif, M.; Awan, M.Y.; Alkinani, M.H. An optimized approach for breast cancer classification for histopathological images based on hybrid feature set. *Curr. Med. Imaging Rev.* **2021**, *17*, 136–147. [CrossRef]
29. Samieinasab, M.; Torabzadeh, S.A.; Behnam, A.; Aghsami, A.; Jolai, F. Meta-health stack: A new approach for breast cancer prediction. *Health Care Anal.* **2022**, *2*, 100010. [CrossRef]
30. Mushtaq, I.; Umer, M.; Imran, M.; Nasir, I.M.; Muhammad, G.; Shorfuzzaman, M. Customer prioritization for medical supply chain dur ing COVID-19 pandemic. *Comput. Mater. Contin.* **2021**, *70*, 59–72.
31. Nardin, S.; Mora, E.; Varughese, F.M.; D'Avanzo, F.; Vachanaram, A.R.; Rossi, V.; Saggia, C.; Rubinelli, S.; Gennari, A. Breast cancer survivor ship, quality of life, and late toxicities. *Front. Oncol.* **2020**, *10*, 864. [CrossRef]
32. Nasir, I.M.; Raza, M.; Shah, J.H.; Khan, M.A.; Rehman, A. Human action recognition using machine learning in uncontrolled environment. In Proceedings of the 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), Riyadh, Saudi Arabia, 6–7 April 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 182–187.
33. Alkhanbouli, R.; Almadhaani, H.M.A.; Alhosani, F.; Simsekler, M.C.E. The role of explainable artificial intelligence in disease prediction: A systematic literature review and future research directions. *BMC Med. Inform. Decis. Mak.* **2025**, *25*, 110. [CrossRef] [PubMed]
34. Teoh, J.R.; Hasikin, K.; Lai, K.W.; Wu, X.; Li, C. Enhancing early breast cancer diagnosis through automated microcalcification detection using an optimized ensemble deep learning framework. *PeerJ Comput. Sci.* **2024**, *10*, e2082. [CrossRef] [PubMed]
35. Wang, W.; Li, Y.; Yan, X.; Xiao, M.; Gao, M. Breast cancer image classification method based on deep transfer learning. In Proceedings of the International Conference on Image Processing, Machine Learning and Pattern Recognition, Guangzhou, China, 13–15 September 2024; pp. 190–197.
36. Sarwar, N.; Al-Otaibi, S.; Irshad, A. Optimizing breast cancer detection: Integrating few-shot and transfer learning for enhanced accuracy and efficiency. *Int. J. Imaging Syst. Technol.* **2025**, *35*, e70033. [CrossRef]
37. Raghavan, K.; B, S.; v, K. Attention guided grad-cam: An improved explain able artificial intelligence model for infrared breast cancer detection. *Multimed. Tools Appl.* **2024**, *83*, 57551–57578. [CrossRef]
38. Naz, A.; Khan, H.; Din, I.U.; Ali, A.; Husain, M. An efficient optimization system for early breast cancer diagnosis based on internet of medical things and deep learning. *Eng. Technol. Appl. Sci. Res.* **2024**, *14*, 15957–15962. [CrossRef]
39. Hassan, M.M.; Hassan, M.M.; Yasmin, F.; Khan, M.A.R.; Zaman, S.; Islam, K.K.; Bairagi, A.K. A comparative assessment of machine learning algorithms with the least absolute shrinkage and selection operator for breast cancer detection and prediction. *Decis. Anal. J.* **2023**, *7*, 100245. [CrossRef]
40. Malakouti, S.M.; Menhaj, M.B.; Suratgar, A.A. Ml: Early breast cancer diagnosis. *Curr. Probl. Cancer Case Rep.* **2024**, *13*, 100278. [CrossRef]

41. Duan, H.; Zhang, Y.; Qiu, H.; Fu, X.; Liu, C.; Zang, X.; Xu, A.; Wu, Z.; Li, X.; Zhang, Q.; et al. Machine learning-based prediction model for distant metastasis of breast cancer. *Comput. Biol. Med.* **2024**, *169*, 107943. [CrossRef]
42. Alhsnony, F.H.; Sellami, L. Advancing breast cancer detection with convolutional neural networks: A comparative analysis of mias and ddsm datasets. In Proceedings of the 2024 IEEE 7th International Conference on Advanced Technologies, Signal and Image Processing (ATSIP) 1, Sousse, Tunisia, 11–13 July 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 194–199.
43. Thawkar, S. A hybrid model using teaching–learning-based optimization and salp swarm algorithm for feature selection and classification in digital mammography. *J. Ambient Intell. Humaniz. Comput.* **2021**, *12*, 8793–8808. [CrossRef]
44. Sheth, P.; Patil, S. Improved jaya optimization algorithm for feature selection on cancer diagnosis data using evolutionary binary coded approach. *Solid State Technol.* **2020**, *29*, 992–1006.
45. Johnson, D.S.; Johnson, D.L.L.; Elavarasan, P.; Karunanithi, A. Feature selection using flower pollination optimization to diagnose lung cancer from ct images. In *Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC), San Francisco, CA, USA, 5–6 March 2020*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 2, pp. 604–620.
46. Singh, L.K.; Khanna, M.; Singh, R. Feature subset selection through nature inspired computing for efficient glaucoma classification from fundus images. *Multimed. Tools Appl.* **2024**, *83*, 77873–77944. [CrossRef]
47. Houssein, E.H.; Oliva, D.; Samee, N.A.; Mahmoud, N.F.; Emam, M.M. Liver cancer algorithm: A novel bio-inspired optimizer. *Comput. Biol. Med.* **2023**, *165*, 107389. [CrossRef] [PubMed]
48. Alnowami, M.R.; Abolaban, F.A.; Taha, E. A wrapper-based feature selection approach to investigate potential biomarkers for early detection of breast cancer. *J. Radiat. Res. Appl. Sci.* **2022**, *15*, 104–110. [CrossRef]
49. Houssein, E.H.; Emam, M.M.; Ali, A.A. An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators' algorithm. *Neural Comput. Appl.* **2022**, *34*, 18015–18033. [CrossRef]
50. Song, W.; Wang, X.; Guo, Y.; Li, S.; Xia, B.; Hao, A. CenterFormer: A Novel Cluster Center Enhanced Transformer for Unconstrained Dental Plaque Segmentation. *IEEE Trans. Multimed.* **2024**, *26*, 10965–10978. [CrossRef]
51. Naas, M.; Mzoughi, H.; Njeh, I.; Slima, M.B. Deep learning-based computer aided diagnosis (cad) tool supported by explainable artificial intelligence for breast cancer exploration. *Appl. Intell.* **2025**, *55*, 679. [CrossRef]
52. Fattahi, A.S.; Hoseini, M.; Dehghani, T.; Nia, R.G.N.N.; Naseri, Z.; Ebrahimzadeh, A.; Mahri, A.; Eslami, S. Explainable machine learning versus known nomogram for predicting non-sentinel lymph node metastases in breast cancer patients: A comparative study. *Comput. Biol. Med.* **2025**, *184*, 109412. [CrossRef]
53. Murugan, T.K.; Karthikeyan, P.; Sekar, P. Efficient breast cancer detection using neural networks and explainable artificial intelligence. *Neural Comput. Appl.* **2025**, *37*, 3759–3776. [CrossRef]
54. Manojee, K.S.; Kannan, A.R. Patho-net: Enhancing breast cancer classification using deep learning and explainable artificial intelligence. *Am. J. Cancer Res.* **2025**, *15*, 754. [CrossRef] [PubMed]
55. Ariyametkul, A.; Tamang, S.; Paing, M.P. Explainable ai (xai) for breast cancer diagnosis. In Proceedings of the 2024 16th Biomedical Engineering International Conference (BMEiCON), Pattaya, Thailand, 21–24 November 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–5.
56. Hussain, S.M.; Buongiorno, D.; Altini, N.; Berloco, F.; Prencipe, B.; Moschetta, M.; Bevilacqua, V.; Brunetti, A. Shape-based breast lesion classification using digital tomosynthesis images: The role of explainable artificial intelligence. *Appl. Sci.* **2022**, *12*, 6230. [CrossRef]
57. Suckling, J.; Parker, J.; Dance, D.; Astley, S.; Hutt, I.; Boggis, C.; Ricketts, I.; Stamatakis, E.; Cerneaz, N.; Kok, S.; et al. *Mammographic Image Analysis Society (MIAS) Database v1.21*; Apollo-University of Cambridge Repository: Online, 2015. [CrossRef]
58. Al-Dhabyani, W.; Gomaa, M.; Khaled, H.; Fahmy, A. Dataset of breast ultrasound images. *Data Brief* **2020**, *28*, 104863. [CrossRef]
59. Peraza-Vázquez, H.; Peña-Delgado, A.F.; Echavarría-Castillo, G.; Morales-Cepeda, A.B.; Velasco-Álvarez, J.; Ruiz-Perez, F. A bio-inspired method for engineering design optimization inspired by dingoes hunting strategies. *Math. Probl. Eng.* **2021**, *2021*, 9107547. [CrossRef]
60. Boyd, S.P.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
61. Rashedi, E.; Nezamabadi-Pour, H.; Saryazdi, S. GSA: A gravitational search algorithm. *Inf. Sci.* **2009**, *179*, 2232–2248. [CrossRef]
62. Yang, X.-S. Firefly algorithms for multimodal optimization. In *Stochastic Algorithms: Foundations and Applications*; SAGA: London, UK; Springer: Berlin/Heidelberg, Germany, 2009; LNCS 5792; pp. 169–178.
63. Yang, X.-S.; Deb, S. Engineering optimisation by cuckoo search. *Int. J. Math. Model. Numer. Optim.* **2010**, *1*, 330–343. [CrossRef]
64. Hashim, F.A.; Mostafa, R.R.; Hussien, A.G.; Mirjalili, S.; Sallam, K.M. Firefly algorithm: A physical law-based algorithm for numerical optimization. *Knowl.-Based Syst.* **2023**, *260*, 110146. [CrossRef]
65. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.

66. Holzinger, A.; Biemann, C.; Pattichis, C.S.; Kell, D.B. What do we need to build explainable ai systems for the medical domain? *arXiv* **2017**, arXiv:1712.09923.
67. Gunning, D.; Aha, D. Darpa's explainable artificial intelligence (xai) program. *AI Mag.* **2019**, *40*, 44–58.
68. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Inf. Fusion* **2020**, *58*, 115. [CrossRef]
69. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1.
70. van Rijsbergen, C.J.; Lalmas, M. Information calculus for information retrieval. *J. Am. Soc. Inf. Sci.* **1996**, *47*, 385–398. [CrossRef]
71. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv* **2020**, arXiv:2010.16061.

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.