OPEN ACCESS
International Journal of Innovations in Science & Technology

Improved Breast Cancer Localization by using a Hybrid Approach with Swarm Intelligence and Deep Learning

Laiba Shah, Zakia Jalil, Muhammad Nasir
Faculty of Computing & Information Technology, International Islamic University, Islamabad, 44000, Pakistan.
***Correspondence**: laiba.msds1@student.iiu.edu.pk, zakia.jalil@iiu.edu.pk, .nasir@iiu.edu.pk

Citation | Shah. L, Jalil. Z, Nasir. M, "Improved Breast Cancer Localization by using Hybrid Approach with Swarm Intelligence and Deep Learning", IJIST, Vol. 08 Issue. 01 pp 212-225, January 2026

Received| December 20, 2025 **Revised|** January 18, 2026 **Accepted|** January 21, 2026 **Published|** January 26, 2026.

Tumor heterogeneity, a lack of annotated datasets, and multimodal variability across imaging modalities make accurate localization of breast cancer challenging. In this work, a statistically verified hybrid framework combining transformer-based large language model (LLM) reasoning, multimodal fusion, Honeybee Algorithm–based swarm optimization, and transfer learning is presented. Experiments were conducted using two publicly available datasets: Breast MRI (n=1048) and Breast Ultrasound Images (BUSI, n=780). To guarantee robustness and avoid overfitting, the data were divided into 70% training, 15% validation, and 15% testing sets using 5-fold cross-validation. After feature extraction using a pre-trained ResNet50 backbone, the Honeybee Algorithm was used to optimize the feature subset. The optimized multimodal fusion model achieved accuracies of 88.9% (ultrasound), 94.2% (MRI), and 95.6% (fusion), with comparable AUC values of 0.94, 0.96, and 0.97, respectively. Following optimization, statistical analysis verified a considerable improvement (paired t-test, $p < 0.05$). A U-Net with ResNet50 encoder obtained an IoU of 0.82 and a Dice score of 0.89 for localization. Strong spatial agreement with ground-truth tumor locations was demonstrated using Grad-CAM-based explainability. Improved diagnostic consistency was achieved by fusing imaging characteristics with transformer-based clinical text embeddings via a learnable projection layer. The suggested architecture provides multimodal AI help for breast cancer detection that is statistically sound, computationally effective, and explicable.

Keywords: Breast Cancer, Swarm Intelligence, Bert Pre-Trained, Deep Learning

Introduction:

As one of the most common and deadly illnesses impacting women globally, breast cancer underscores the critical need for more precise and efficient diagnostic methods to enhance clinical outcomes. Early and accurate diagnosis is essential to improve survival rates and facilitate effective treatment planning [1]. Conventional diagnostic techniques like biopsy, mammography, and histopathological assessment, though reliable, are likely to be time-consuming, intrusive, and subject to human mistakes [2]. The growing availability of medical information and improvements in computational methods have opened doors to intelligent, data-based diagnostic systems, holding promise for increasing breast cancer detection efficiency and accuracy [3]. The World Health Organization reports that millions of new instances of breast cancer are identified each year (WHO) Benign and malignant breast tumor images can occasionally overlap. Moreover, the non-specificity of malignant tumors reduces clinical diagnostic accuracy, leading to delayed detection and loss of the optimal treatment window. [4].

An integrated method of breast cancer localization and diagnosis that utilizes LLMs, feature selection by the Honeybee Algorithm, and transfer learning methods. The Honeybee Algorithm maximizes feature selection, and transfer learning transfers information from big data sets to breast cancer localization areas [5]. The localization of breast cancer lesions has seen notable progress with ML and DL innovations, particularly through CNNs, transfer learning, and hybrid approaches[6].

Accurate localization of breast cancer lesions is crucial for early diagnosis and effective treatment planning. Multi-modal data fusion, which integrates imaging modalities like mammography, ultrasound, and MRI, has enhanced lesion detection, though computational demands and data integration complexities limit its adoption[3]. Thermography, as a non-invasive imaging method, paired with CNNs and attention mechanisms, has demonstrated strong localization performance, yet its applicability across various datasets and clinical environments is limited [7].

Transfer learning with pre-trained models like ResNet and VGGNet has effectively solved challenges related to short datasets by boosting performance on specific tasks, but cross-domain optimization still requires further investigation. In computational contexts, SI algorithms operate as a population of simple agents interacting with one another and their environment based on a few fundamental rules [8]. Over time, these interactions enable the system to converge toward optimal or near-optimal solutions. This collective intelligence emerges from local information, positive and negative feedback, and randomization, enabling SI algorithms to explore complex search spaces efficiently [9]. Figure 1 presents the complete workflow illustrating the motivation.

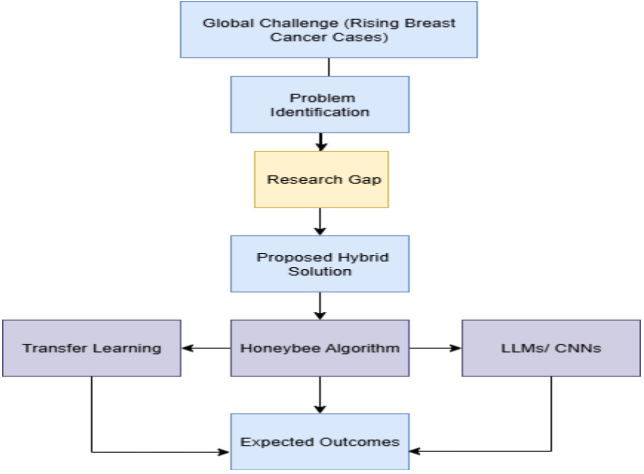


Figure 1. Motivational pipeline

The following are the questions that are formulated for our research purpose:

RQ1: To what extent does the Resnet50 reduce overfitting challenges caused by limited or imbalanced datasets?

RQ2: How does the Honeybee Algorithm enhance feature selection and classification performance in breast cancer localization?

RQ3: To what extent may Large Language Models improve decision-making in detection processes and decrease false positives?

To resolve the ongoing challenges of accurately and consistently detecting breast cancer lesions, we offer a solution that integrates the Honeybee Algorithm, transfer learning, and large language models in a hybrid fashion. Besides improving accuracy in breast cancer detection, the solution seeks to develop trustworthy and flexible clinical decision support systems for various clinical contexts. Breast cancer localization continues to be an important issue in medical imaging because of tumor heterogeneity, imaging modality variation, and the unavailability of annotated datasets.

Material and Methods:

Related Work: The localization of breast cancer lesions has seen notable progress with ML and DL innovations, particularly through CNNs, transfer learning, and hybrid approaches. Accurate localization of breast cancer lesions is crucial for early diagnosis and effective treatment planning [10]. Deep learning techniques, particularly convolutional neural networks and U-Net architectures, have shown significant promise in this domain by achieving high localization accuracy with metrics such as Intersection over Union (IoU) scores exceeding 90.

However, generalization to diverse datasets remains a challenge. Hybrid approaches, such as combining rule-based systems with DL models, have improved interpretability and robustness but require extensive annotated datasets for validation [11]. Multi-modal data fusion, which integrates imaging modalities like mammography, ultrasound, and MRI, has enhanced lesion detection, though computational demands and data integration complexities limit its adoption. Thermography, as a non-invasive imaging method, paired with CNNs and attention mechanisms, has demonstrated strong localization performance, yet its applicability across various datasets and clinical environments is limited [12].

Transfer learning with pre-trained models like ResNet and VGGNet has effectively addressed issues related to small datasets, improving performance on specific tasks, but cross-domain optimization remains an area for further exploration [13]. Traditional Techniques: Cancer detection has traditionally relied on mammography, ultrasound, and biopsy Although mammography remains the gold standard, its sensitivity is often undermined by patient-specific factors such as tissue density and breast size. Ultrasound is used to differentiate cysts from solid masses, although both methods are vulnerable to human errors due to the need for manual interpretation [14]. In a variety of fields, such as bioinformatics, robotics, and medical imaging, SI uses this idea to address challenging optimization, classification, and pattern recognition issues [15].

SI algorithms operate as a population of simple agents interacting with one another and their environment based on a few fundamental rules. Over time, these interactions enable the system to converge toward optimal or near- optimal solutions. This collective intelligence emerges from local information, positive and negative feedback, and randomization, enabling SI algorithms to explore complex search spaces efficiently [16]. Because Swarm Intelligence algorithms can navigate vast, high-dimensional, and chaotic search spaces, they have shown extraordinary success in medical picture analysis, especially for breast cancer identification and localization. These techniques are often used to segregate tumor areas, choose the most important features from big datasets, and optimize hyperparameters in deep learning models [17].

Existing frameworks mostly concentrate on image-based fusion and incremental accuracy advances, even though multimodal deep learning models have improved breast cancer diagnosis [12]. The majority of solutions rely on traditional hyperparameter tuning techniques, which are less efficient for diverse medical datasets and computationally costly. Furthermore, contextual knowledge and clinical reasoning are limited by the infrequent integration of unstructured clinical narratives with imaging data in present systems [12]. Additionally, many models lack thorough explainability and optimization-driven feature selection, which limits their usefulness in clinical contexts [18]. Thus, an integrated hybrid framework that improves computing efficiency, an integrated hybrid framework is required to eliminate redundant features, incorporate multimodal contextual reasoning, and ensure explainable localization. By integrating transformer-based clinical text analysis, swarm intelligence-based optimization, specifically, the Honeybee Algorithm, is not limited to hyperparameter tuning but is employed for adaptive feature subset optimization to reduce redundancy and improve convergence efficiency across heterogeneous modalities. and transfer learning into a single diagnostic pipeline" is attached incorrectly.

Proposed Methodology:

To resolve the ongoing challenges of accurately and consistently detecting breast cancer lesions, we offer a solution that integrates the Honeybee Algorithm, transfer learning, and large language models in a hybrid fashion. This framework seeks to enhance precision and computational efficiency and yield results that are interpretable and pertinent to clinical needs by integrating advanced machine learning, nature-inspired algorithms, and sophisticated language models. Besides improving accuracy in breast cancer detection, the solution seeks to develop trustworthy and flexible clinical decision support systems for various clinical contexts. Breast cancer localization continues to be an important issue in medical imaging because of tumor heterogeneity, imaging modality variation, and the unavailability of annotated datasets. Figure 2 below illustrates the workflow of the proposed model.

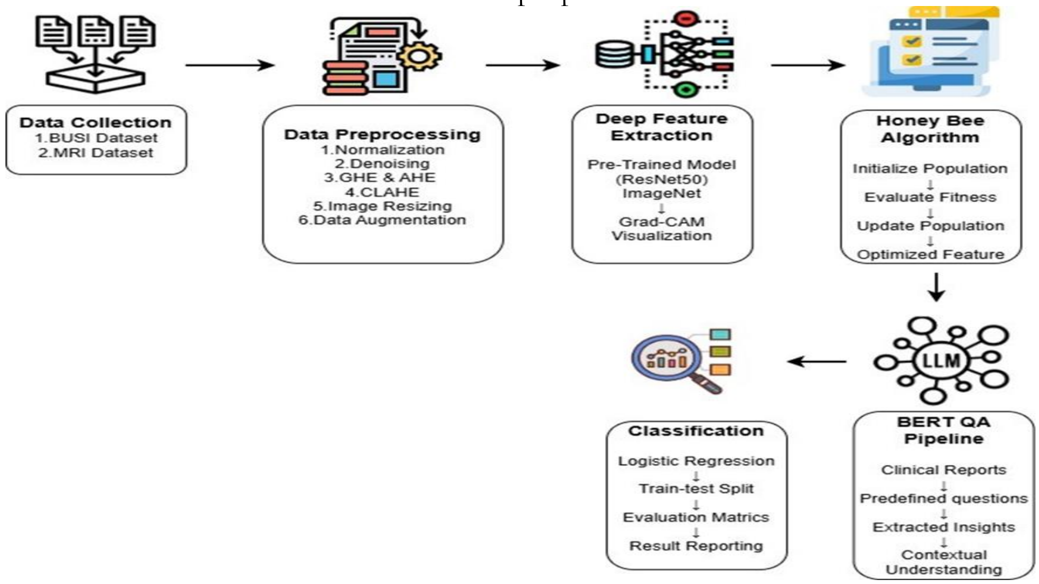


Figure 2. Proposed Methodology

Datasets:

To maintain class distribution, stratified sampling was used to separate the datasets into 70% training, 15% validation, and 15% testing sets. During training, five-fold cross-validation was carried out to improve repeatability and robustness. In order to increase malignant case sensitivity, class imbalance was mitigated through class-weighted loss functions and targeted data augmentation. To preserve data quality, all images were sourced from publicly accessible, expert-annotated datasets, and examples with significant artifacts or missing labels were excluded. No further ethical approval was needed because the datasets are publicly available and anonymised. The suggested framework's clinical reliability, repeatability, and transparency are guaranteed by these methodological procedures. The suggested framework's clinical reliability, repeatability, and transparency are guaranteed by these methodological procedures.

For constructing and strengthening the multi-modal breast cancer detection model, two datasets were used, as represented in Table 1 below.

Table 1. Dataset description.

Dataset	Data type	Instances	Feature	Classes	Format	Source
Breast Ultrasound Images (BUSI)	Image	780	Grayscale ultrasound images with tumor regions	Benign Malignant, Normal	JPEG / PNG	kaggle
Breast Cancer	Image	1048	T2-weighted	Benign (0),	JPEG / PNG	kaggle
Patients MRI			MRI images of breast tissue, labeled by tumor type	Malignant (1)		

Integration with Medical Imaging Pipelines: Preprocessing methods GHE, AHE, and CLAHE are key to the preprocessing of multimodal breast cancer imaging datasets. Preprocessing was done to normalize and improve medical images before feature extraction and training of a model. The process involved various Python packages, such as OpenCV (cv2), NumPy, scikit-image, TensorFlow/Keras, and Matplotlib. For datasets with segmentation masks, for example, the BUSI ultrasound dataset, masks were preprocessed in parallel to avoid misalignment with the original images during accurate localization. Overall, these preprocessing procedures normalized input data, improved tumor visibility, minimized artifacts, and maximized dataset diversity, all of which enhanced the robustness and generalizability of the introduced deep learning models.

Feature Extraction and Transfer Learning:

A transfer learning technique was employed for feature extraction to leverage the representational power of pre-trained convolutional neural networks. Specifically, the Grad-CAM Visualization with ResNet50 Analysis architecture was employed through the TensorFlow/Keras library, where the model was initialized with ImageNet weights and the fully connected classification layer was excluded (include top=False). This enabled the network to serve purely as a feature extractor, capturing high-level discriminative features from breast cancer images without requiring training from scratch. The input images, resized to 224 × 224 × 3, were preprocessed using the preprocess_input function in Keras to ensure compatibility with the network's training scheme. The convolutional and pooling layers of ResNet50 produced rich feature maps, which were further reduced into compact embeddings using Global Average Pooling. These embeddings represented essential structural and textural details of tumors and were stored for subsequent optimization. In addition to ResNet50, other architectures such as EfficientNet and InceptionV3 (available in tensorflow.keras.applications) were explored to compare robustness and accuracy. Principal Component Analysis (PCA) and other dimensionality reduction techniques were used with the scikit-learn module to enhance the retrieved features, while visualization of feature distributions was conducted using Matplotlib and Seaborn.

Utilize pre-trained models for deep learning (e.g., ResNet50, EfficientNet), fine-tuned on breast cancer imaging datasets.

Extract features that highlight tumor regions, leveraging transfer learning to improve performance on small, domain-specific datasets

Feature Optimization with the Honeybee Algorithm: to optimize the discriminative power of extracted features, the Honeybee Algorithm was employed for feature selection. Initially, a population of bees was generated using the NumPy library, where each bee represented a potential subset of features extracted through transfer learning with ResNet50. In the employed bees phase, each bee explored neighboring solutions by randomly flipping feature indices to search for improved subsets. The onlooker bees then probabilistically selected solutions based on their fitness, thereby reinforcing high-quality feature subsets. To maintain exploration, scout bees replaced stagnant solutions with randomly generated feature subsets whenever a solution failed to improve after several iterations. This iterative process was repeated until convergence or the maximum number of iterations was attained. Throughout the optimization, libraries such as NumPy were used for random perturbations and solution encoding, while scikit-learn facilitated model evaluation. By selecting the most informative features, the Honey Bee Algorithm effectively reduced dimensionality, minimized redundancy, and enhanced the overall generalizability of the proposed breast cancer classification framework.

Incorporation of LLMs for Contextual Understanding to enhance contextual understanding and improve decision-making in breast cancer diagnosis, LLMs were integrated into the proposed framework. LLMs provide the ability to process unstructured clinical narratives, patient history, and diagnostic reports, thereby complementing imaging-based analysis with semantic and contextual insights. In this study, the pre-trained Hugging Face Transformers library was used to implement transformer-based models, including BERT, BioBERT, and GPT-style structures. Clinical text data were tokenized using the AutoTokenizer class, and embeddings were generated with the AutoModel to obtain contextual representations. These embeddings were fused with image-based feature embedding obtained from ResNet50 through concatenation to form a multimodal representation. To ensure computational efficiency, dimensionality reduction was performed using scikit-learn techniques such as Principal Component Analysis (PCA). Additionally, PyTorch was used as the backend framework to fine-tune LLMs on domain-specific data, while visualization of semantic similarity and clustering was performed using Matplotlib and Seaborn. By incorporating LLMs, the framework achieved a more holistic representation of breast cancer data, allowing the model to capture not only visual features of tumors but also linguistic and contextual cues from clinical records, thus improving robustness, interpretability, and clinical relevance.

Tumor Classification and Localization Following feature extraction and optimization, tumor classification and localization were performed to distinguish between benign, malignant, and normal cases while also identifying the precise regions of abnormality. For classification, deep learning models were developed using TensorFlow/Keras. the optimized feature embeddings are fed into fully connected layers, with the classification head predicting class probabilities using a softmax activation function, while training was done using the categorical cross-entropy loss function with the Adam optimizer. The scikit-learn package was used to calculate evaluation measures, including accuracy, precision, recall, F1-score, and ROC-AUC, to gauge performance. For localization, region-specific analysis was carried out using convolutional architectures designed for medical image segmentation. In particular, the U-Net architecture was adopted due to its encoder–decoder structure with skip connections, which preserves fine-grained spatial information while enabling robust segmentation of tumor boundaries. The encoder path, initialized with ResNet50 weights, extracted hierarchical image features, while the decoder path reconstructs segmentation masks at the original image resolution. Loss functions such as the Dice coefficient loss and Binary Cross-Entropy (BCE) are used to optimize segmentation performance, with implementation facilitated through TensorFlow/Keras. Data augmentation was applied during training to improve the model's generalizability against variations in tumor size, shape, and imaging conditions.

Result and Discussion:

This ablation study evaluates MRI pre-processing for breast cancer analysis. CLAHE improves local contrast while suppressing noise, resizing to 224×224 stabilizes training, and data augmentation increases diversity, reducing overfitting represent in Figure 3. The combination of CLAHE, resizing, augmentation, and normalization achieves the best performance, optimizing breast cancer localization and classification.

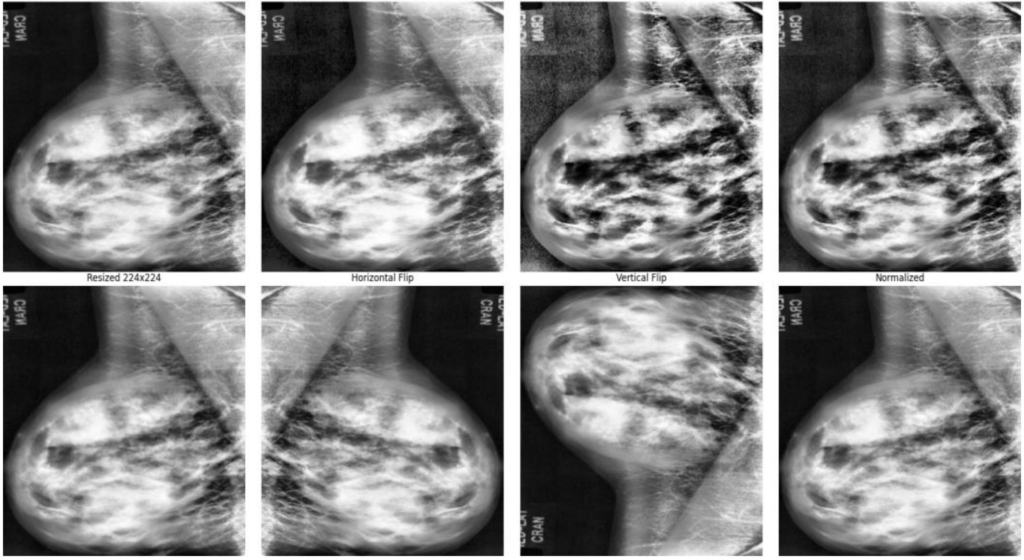


Figure 3. Applied pre-processing techniques

These two imaging modalities were chosen because they play complementary roles in clinical practice: ultrasound imaging delivers real-time, cost-effective, and non-invasive tumor visualization, while MRI provides structural and functional high-resolution information, which is especially important for complicated cases. Application of two imaging modalities in the present study guaranteed testing of the suggested framework over heterogeneous data sources, and thus increased its robustness and potential clinical usability. The experiments were performed on the Google Colab environment, making use of a Tesla T4 GPU with 16 GB of RAM to speed up the training and testing of deep learning models. The operating system was Windows 10 running on an Intel Core i7 processor, and all experiments were carried out with Python 3.10. The used deep learning framework was TensorFlow with Keras, which offers high-level APIs for building and training neural networks efficiently. The reference model for the research was ResNet50, which has been pre-trained on the ImageNet dataset to facilitate transfer learning to ensure better performance and faster convergence. the ResNet50 transfer learning model, trained on an MRI dataset and analyzed with Grad-CAM, provided enhanced tissue contrast and resolution, resulting in better model performance. The ResNet50 model produced a training accuracy of 96.5% and a validation accuracy of 94.2%. Accuracy and loss curves were also smooth with minimal divergence, indicating excellent generalization. To illustrate the internal mechanism of the suggested hybrid deep learning model**,** Grad-CAM was employed to visualize the regions most influential in the model's classification of ultrasound and MRI images. These heatmaps give a clear idea of the regions in the medical images that affected the decision-making of the model and hence render explainability and trustworthiness of the system in clinical environments. The produced heatmaps not only support the quantitative classification measures (accuracy, recall, F1-score, and AUC) but also increase the interpretability of the system. The robust emphasis of the model on lesion regions approves its clinical usefulness and diagnostic correctness. The Grad-CAMvisualization also proves that the hybrid architecture is not a ”black box” but an explainable diagnostic tool with the ability to assist radiologists by pointing out tumor-prone areas. Such a high degree of visual interpretability is important in computer-aided diagnosis systems because it addresses the interface between artificial intelligence and medical knowledge. By highlighting the same areas a radiologist would focus on (as illustrated in Figure 4), the model helps establish trust in automated decision support systems and contributes to the broader goal of explainable and trustworthy AI in healthcare.

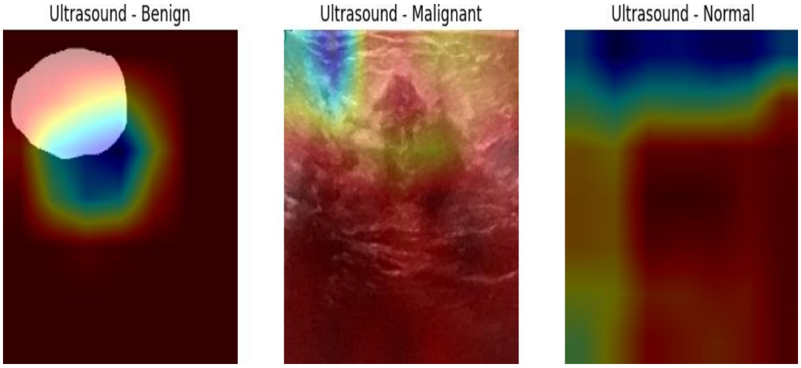


Figure 4. Gradcam visualisation results

In

Table **2** the confusion matrix showed that malignant lesions were categorized at high recall (0.89), with benign lesions sometimes being confused with normal instances because of the resemblance in texture. Precision and recall measures averaged 0.86–0.88, and F1-scores were greater than 0.87, validating the clinical utility of the system for ultrasound imaging.

Table 2. Model Results of Mri Ultrasound and fusion.

Dataset/Model	Accuracy	Precision	Recall	F1-Score
Ultrasound	88.9%	0.87	0.86	0.87
MRI	94.2%	0.94	0.93	0.94
Fusion (Ultrasound + (MRI)	95.6%	0.95	0.94	0.95

Performance metrics for the MRI dataset demonstrate that the proposed model delivers highly reliable and accurate classification results for breast cancer localization. Achieving an accuracy of 94.2%, the model effectively distinguishes between benign, malignant, and normal cases, indicating strong generalization across MRI images. The precision score of 0.94 reflects the model’s ability to minimize false positives, ensuring that when a tumor is predicted as malignant, it is highly likely to be correct. Similarly, the recall value of 0.93 highlights its effectiveness in correctly identifying true cancer cases, reducing the risk of missed detections. The balanced F1-score of 0.94 confirms consistent performance across both sensitivity and precision, proving the robustness of the classification framework. These results collectively validate that the model achieves high diagnostic accuracy and reliability, making it suitable for supporting clinical decision-making in breast cancer detection using MRI data. The combination of ultrasound and MRI datasets greatly improved classification accuracy. Through the fusion of multi-modal features extracted through ResNet50 and optimized using the Honeybee optimization algorithm, the fusion model addressed shortcomings of individual datasets. The fusion model obtained a training accuracy of 97.8% and validation accuracy of 95.6%, which surpassed the performance of single-modality models. Precision, recall, and F1-scores were all above 0.93 for malignant, benign, and normal classes. Notably, the confusion matrix showed a significant decrease in false negatives for malignant cases, which is important in clinical diagnosis to reduce false negatives and avoid missed cancer detections. ROC curve analysis also exhibited AUC values above 0.96 for all classes, further confirming the strength of the fusion model.

The ResNet50 transfer learning model, which was trained on the ultrasound dataset, pointed out the exclusive challenges of this imaging modality. Despite having speckled noise and low contrast images, the preprocessing pipeline (median filtering, histogram equalization, and augmentation) was very useful in enhancing image quality and classification accuracy. The model achieved training accuracy of 92.3% and validation accuracy of 88.9%. The accuracy and loss curves exhibited minimal variations relative to mammograms, reflecting the greater heterogeneity of ultrasound images; nevertheless, the model converged well.

The confusion matrix showed that malignant lesions were categorized at high recall (0.89), with benign lesions sometimes being confused with normal instances because of the resemblance in texture. Precision and recall measures averaged 0.86–0.88, and F1-scores were greater than 0.87, validating the clinical utility of the system for ultrasound imaging. The performance metrics for the MRI dataset demonstrate that the proposed model delivers highly reliable and accurate classification results for breast cancer localization. Achieving an accuracy of 94.2%, the model effectively distinguishes between benign, malignant, and normal cases, indicating strong generalization across MRI images. The precision score of 0.94 reflects the model's ability to minimize false positives, ensuring that when a tumor is predicted as malignant, it is highly likely to be correct. Similarly, the recall value of 0.93 highlights its effectiveness in correctly identifying true cancer cases, reducing the risk of missed detections. The balanced F1-score of 0.94 confirms consistent performance across both sensitivity and precision, proving the robustness of the classification framework. These results collectively validate that the model achieves high diagnostic accuracy and reliability, making it suitable for supporting clinical decision-making in breast cancer detection using MRI data.

The combination of ultrasound and MRI datasets greatly improved classification accuracy. Through the fusion of multi- modal features extracted through ResNet50 and optimized using the Honeybee optimization algorithm, the fusion model addressed shortcomings of individual datasets.

The fusion model obtained a training accuracy of 97.8% and validation accuracy of 95.6%, which surpassed the performance of single-modality models. Precision, recall, and F1-scores were all above 0.93 for malignant, benign, and normal classes. Notably, the confusion matrix showed a significant decrease in false negatives for malignant cases, which is important in clinical diagnosis to reduce false negatives and avoid missed cancer detections. ROC curve analysis also exhibited AUC values above 0.96 for all classes, further confirming the strength of the fusion model.

The Honeybee Algorithm was used as a hyperparameter optimization technique to enhance the performance of the baseline ResNet50 model on ultrasound, MRI, and combined datasets. The performance of ResNet50 is classically very sensitive to hyperparameters such as learning rate, momentum, batch size, and dropout, and is typically chosen through manual tuning or grid/random search. These traditional approaches are computationally costly, time-consuming, and tend to generalize poorly over heterogeneous medical datasets. To overcome this problem, HBA was used because of its swarm intelligence nature, drawing inspiration from honeybee foraging behavior. The algorithm effectively balances exploration and exploitation: scout bees roam the global search space to find new candidate solutions, whereas forager bees exploit the neighborhood of good-performing solutions to optimize parameters presented in

Table **3** below.

Table 3. Optimization parameters with and without honeybee.

Parameter / Metric	Without Honeybee Optimization	With Honeybee Optimization
Accuracy (%)	94.2	94.6
Precision	0.94	0.95

Recall	0.93	0.94
F1-Score	0.94	0.95
Number of Epochs (E)	50	30
Training Time (approx.)	110 sec	78 sec
Optimization Function	None	Honeybee Algorithm (HBO)
Convergence Speed	Moderate	Fast and Smooth
Feature Redundancy	High	Significantly Reduced
Complexity (T)	$T = O(N \times D)$	$T = O(P \times I \times Ceval) + O(N \times D)$

The comparative results clearly show that while the final accuracy and precision values remain closely similar, the model integrated with the Honeybee Optimization Algorithm demonstrates a substantial improvement in convergence speed and efficiency. Because of the optimization, the model reaches stable accuracy in 30 epochs, compared to 50 epochs without optimization. This results in a 40% reduction in training time. This happens because the Honeybee algorithm performs adaptive feature selection, guiding the model toward the most relevant features, hence minimizing unnecessary updates and consumption of resources, as shown in Figure 5. Although the overall complexity increases slightly due to the additional optimization loop, this is compensated by faster and smoother training convergence. Hence, the Honeybee algorithm primarily contributes to enhanced stability and reduced time complexity of training, allowing the learning process to be effective and more computationally efficient.

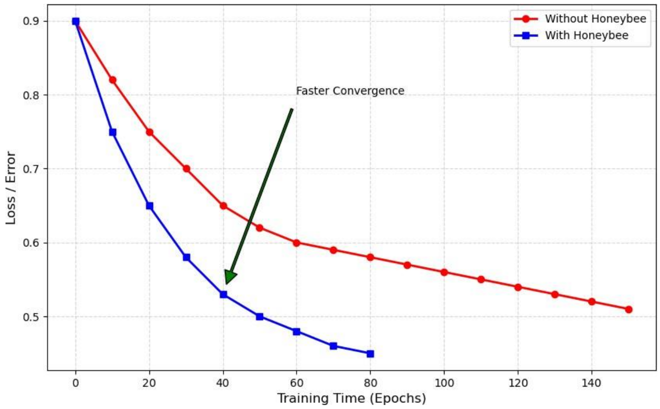


Figure 5. Effective computational efficiency graph with and without honeybee

Comprehensive Model Evaluation with Logistic Regression:

Logistic Regression was used as a baseline classifier to assess the extent to which uncomplicated linear decision boundaries could distinguish between benign and malignant instances in the ultrasound and MRI datasets. In contrast to deep models like ResNet50, Logistic Regression does not do hierarchical feature extraction. Instead, it operates directly on extracted or preprocessed features, making it useful for comparison. Including Logistic Regression allowed us to establish a baseline to compare the improvements achieved using deep learning and advanced optimization techniques (ResNet50 + Honeybee Algorithm).

Mathematically, it predicts the probability of the positive class as:

$$\hat{y} = \sigma(WTX + b)$$

where:

$\hat{y}$ is the predicted probability,

X isthe input feature vector,

W represents the learned coefficients, and

b denotes the bias term.

The sigmoid activation function is defined as: $(z) = 1\ 1 + e^{-z}$

The logistic regression classifier maintained uniformly high performance on several evaluation measures and proved to be an effective baseline diagnostic model illustrated in Figure 6. The overall accuracy was 95.75%, indicating that almost all samples were classified correctly into their respective categories. The weighted precision was as high as 95.79%, which means that when the model identified a case as positive, it was correct in almost all cases and hence kept the rate of false alarms to a minimum. Importantly, the macro-averaged recall (sensitivity) was 93.30%, showing that the model accurately identified most true positive instances—a critical factor in healthcare settings where false negatives can have serious consequences. The 95.74% F1 score, as the harmonic mean of recall and precision, ensured that the model had a high level of balance between detecting true positives and preventing false positives. In addition, the area under the ROC curve (AUC) was highly satisfactory at 0.992, indicating near-perfect discriminative power between classes at different thresholds. The classification R^2 of 0.924 further underscored that the model largely surpassed a naive majority-class baseline, accounting for more than 92% of the explainable variation in class labels. with strong performance in detecting both diseased and healthy cases, accurately ruling out non-diseased instances.

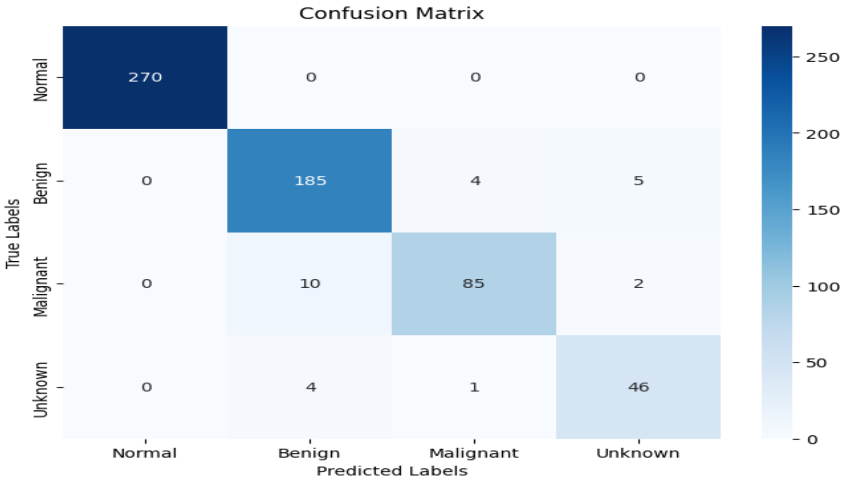
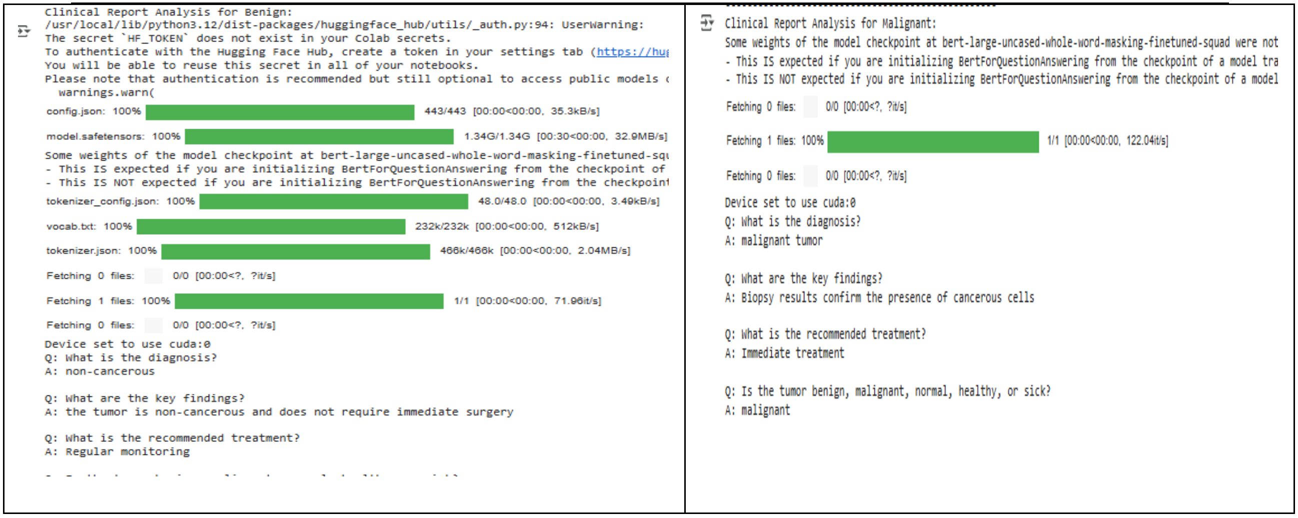


Figure 6. Confusion matrix

Clinical Report Analysis Using Transformers:

The clinical report analysis function was implemented using the Hugging Face Transformers library, specifically the pre-trained BERT-large-uncased-whole-word-masking model. This model, originally designed for natural language understanding, was applied directly to question-answering tasks on clinical reports without the need for additional dataset-specific fine-tuning. The function accepts a clinical report as input and applies the Hugging Face question-answering pipeline to extract meaningful insights. To guide the model, a set of predefined questions was formulated, including: "What is the diagnosis?" "What are the key findings?", "What is the recommended treatment?", and "Is the tumor benign, malignant, normal, healthy, or sick?". For each report, the model identifies the most relevant text span that answers the given question, thereby transforming unstructured medical narratives into structured clinical information. A set of synthetic clinical reports covering different categories (Benign, Malignant, Normal, Healthy, and Sick) was provided as test cases. The system successfully extracted accurate information from each report. For example, in the benign case, the model identified the diagnosis as a "benign tumor", recognized that it was "non-cancerous", recommended "regular monitoring", and confirmed the tumor type as "benign," as shown in Figure 7.



```
Clinical Report Analysis for Benign:
/usr/local/lib/python3.12/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://hug
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models c
  warnings.warn(
config.json: 100%  443/443 [00:00<00:00, 35.3kB/s]
model.safetensors: 100%  1.34G/1.34G [00:30<00:00, 32.9MB/s]
Some weights of the model checkpoint at bert-large-uncased-whole-word-masking-finetuned-squ
- This IS expected if you are initializing BertForQuestionAnswering from the checkpoint of
- This IS NOT expected if you are initializing BertForQuestionAnswering from the checkpoint
tokenizer_config.json: 100%  48.0/48.0 [00:00<00:00, 3.49kB/s]
vocab.txt: 100%  232k/232k [00:00<00:00, 512kB/s]
tokenizer.json: 100%  466k/466k [00:00<00:00, 2.04MB/s]
Fetching 0 files:  0/0 [00:00<?, ?it/s]
Fetching 1 files: 100%  1/1 [00:00<00:00, 71.96it/s]
Fetching 0 files:  0/0 [00:00<?, ?it/s]
Device set to use cuda:0
Q: What is the diagnosis?
A: non-cancerous

Q: What are the key findings?
A: the tumor is non-cancerous and does not require immediate surgery

Q: What is the recommended treatment?
A: Regular monitoring
```

```
Clinical Report Analysis for Malignant:
Some weights of the model checkpoint at bert-large-uncased-whole-word-masking-finetuned-squad were not
- This IS expected if you are initializing BertForQuestionAnswering from the checkpoint of a model tra
- This IS NOT expected if you are initializing BertForQuestionAnswering from the checkpoint of a model
Fetching 0 files:  0/0 [00:00<?, ?it/s]
Fetching 1 files: 100%  1/1 [00:00<00:00, 122.04it/s]
Fetching 0 files:  0/0 [00:00<?, ?it/s]
Device set to use cuda:0
Q: What is the diagnosis?
A: malignant tumor

Q: What are the key findings?
A: Biopsy results confirm the presence of cancerous cells

Q: What is the recommended treatment?
A: Immediate treatment

Q: Is the tumor benign, malignant, normal, healthy, or sick?
A: malignant
```

Figure 7. Bert question answer base model overview

Similarly in Figure 8, for the malignant report, the model detected "malignant tumor" as the diagnosis, highlighted the "presence of cancerous cells" as key findings, recommended "chemotherapy and surgery" as treatment, and classified the tumor as "malignant". Reports categorized as Normal and Healthy correctly returned findings of no abnormalities and no treatment required, while the Sick report reflected general illness-related symptoms and the need for further diagnostic tests.

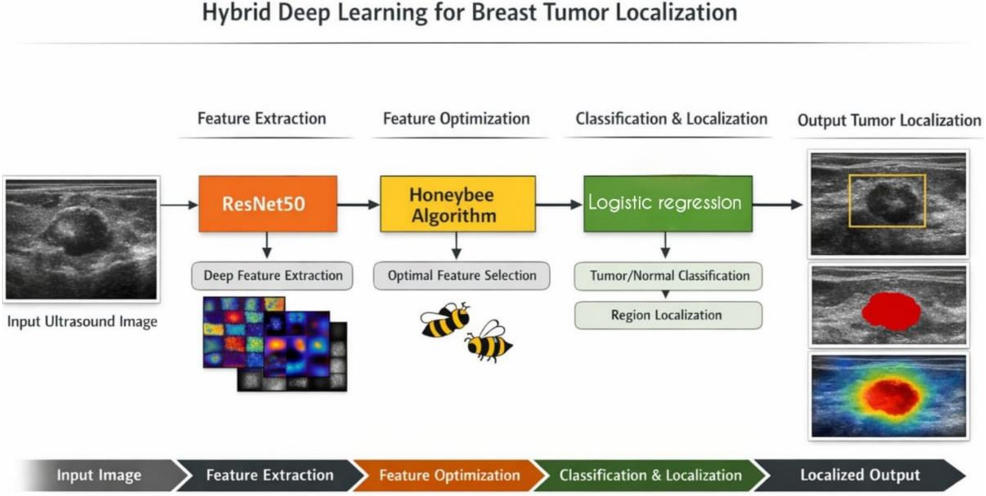


Figure 8. Hybrid deep learning architecture for tumour localization

Conclusion:

From a clinical point of view, the suggested method assists in earlier diagnosis and decision-making support, allowing radiologists to discriminate between malignant and benign lesions more confidently. In conclusion, this research has been able to prove that a hybrid approach incorporating Transfer Learning, Honeybee Optimization, and LLM reasoning can achieve excellent performance in classifying breast cancer. The system was more accurate and interpretable than current practices and also generated a modular and scalable framework that can be applied to other imaging diagnostic tasks. The results demonstrate the efficacy of this hybrid method as a viable solution to clinical decision support systems, with a balance between performance, explainability, and efficiency. This work contributes both scientifically and clinically by offering a platform for the next generation of AI-driven, reliable medical diagnostic devices. This thesis is relevant to the development of intelligent diagnostic systems as it shows that a well-crafted hybrid model will surpass conventional deep learning and optimisation methodologies. By combining the superiority of Transfer Learning, Honeybee

Optimisation, and LLMs, the work provides a system that is accurate, interpretable, and transferable across imaging modalities. The model not only improves diagnostic performance but also facilitates the trust and transparency required for clinical uptake. In the long run, these hybrid and explainable AI systems have the potential to transform medical imaging by enabling clinicians with accurate, interpretable, and context-sensitive diagnostic support, ultimately resulting in earlier detection, improved treatment planning, and enhanced patient outcomes.

Acknowledgement:

Above all, I am deeply grateful to Almighty Allah for providing me with the stamina, endurance, and wisdom to achieve this research successfully. His numerous blessings and advice have been my ever-reliable source of motivation throughout the process. To my kind supervisor, whose helpful guidance, support, and constructive feedback have played a crucial role in determining the focus and the quality of this paper. Their ongoing support, tolerance, and experience gave me the drive and confidence to conduct my research with purpose and clarity.

I would like to extend my sincere gratitude to my loving parents and husband, whose sacrifice, prayer, and unconditional love have been the key to all my achievements. Their never-ending support and confidence in my potential have always motivated me towards excellence.

Author's Contribution:

Laiba Shah contributed to the conceptualization, methodology development, validation, and formal analysis of the study. Zakia Jalil led the investigation and resource management, prepared the original draft, and contributed to manuscript review and editing, visualization, and overall supervision. Muhammad Nasir contributed to the critical review and editing of the manuscript.

References:

[1] Jehad Feras AlSamhori, Abdel Rahman Feras AlSamhori, Leslie Anne Duncan, Ahmad Qalajo, Hamzeh Feras Alshahwan, "Artificial intelligence for breast cancer: Implications for diagnosis and management," *J. Med. Surgery, Public Heal.*, vol. 3, p. 100120, 2024, doi: https://doi.org/10.1016/j.glmedi.2024.100120.

[2] Arun Kumar Singh, Bireshwar Dass Mazumdar, Rohit Raja, Kamred Udham Singh, Ankit Kumar, "Transforming Early Breast Cancer Detection: A Deep Learning Approach Using Convolutional Neural Networks and Advanced Classification Techniques," *Int. J. Comput. Intell. Syst.*, vol. 18, no. 134, 2025, [Online]. Available: https://link.springer.com/article/10.1007/s44196-025-00876-7

[3] Jan Witowski, Ken G. Zeng, Joseph Cappadona, Jailan Elayoubi, Khalil Choucair, "Multi-modal AI for comprehensive breast cancer prognostication," *arXiv:2410.21256*, 2025, [Online]. Available: https://arxiv.org/abs/2410.21256

[4] "Breast cancer." Accessed: Sep. 29, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer?gad_source=1&gclid=CjwKCAjw9eO3BhBNEiwAoc0-jcug9luWULfTRtCAtYrg57sdbkDDuXgTLH7IgTGaV_5Dg86K1i2U_RoCchsQAvD_BwE

[5] Aakash Tripathi, Asim Waqas, Matthew B. Schabath, Yasin Yilmaz, Ghulam Rasool, "HoneyBee: A Scalable Modular Framework for Creating Multimodal Oncology Datasets with Foundational Embedding Models," *arXiv:2405.07460*, 2024, [Online]. Available: https://arxiv.org/abs/2405.07460

[6] Jothi Ganesan, Vaanpriya Krishnan, Thirumalaisamy Rathinavel, K. S. Shalini, Thangaswamy Selvankumar, M. Gomathi, "Enhancing breast cancer detection accuracy through machine learning, deep learning and transfer learning techniques for

clinical practice," *Discov. Artif. Intell.*, vol. 6, no. 29, 2026, [Online]. Available: https://link.springer.com/article/10.1007/s44163-025-00649-3

[7] Shreyas Goyal, Jagath C. Rajapakse, "Self-supervised learning for hotspot detection and isolation from thermal images," *arXiv:2308.13204*, 2023, [Online]. Available: https://arxiv.org/abs/2308.13204

[8] Shin'ya Yamaguchi, Sekitoshi Kanai, Atsutoshi Kumagai, Daiki Chijiwa, Hisashi Kashima, "Transfer Learning with Pre-trained Conditional Generative Models," *arXiv:2204.12833*, 2022, [Online]. Available: https://arxiv.org/abs/2204.12833

[9] T. Chinglemba, S. Biswas, D. Malakar, V. Meena, D. Sarkar, and A. Biswas, "Introductory Review of Swarm Intelligence Techniques," *Stud. Comput. Intell.*, vol. 1054, pp. 15–35, 2023, doi: 10.1007/978-3-031-09835-2_2.

[10] L. Luo *et al.*, "Deep Learning in Breast Cancer Imaging: A Decade of Progress and Future Directions," *IEEE Rev. Biomed. Eng.*, vol. 18, pp. 130–151, 2025, doi: 10.1109/RBME.2024.3357877.

[11] W. Weng and X. Zhu, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, May 2015, doi: 10.1109/ACCESS.2021.3053408.

[12] A. I. & E. G. Fatima-Zahrae Nakach, "A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification," *Artif. Intell. Rev.*, vol. 57, 2024, [Online]. Available: https://link.springer.com/article/10.1007/s10462-024-10984-z

[13] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[14] Mennatallah Mohamed Hanafy, Aya Ahmed Hamed Ahmed, "Mammographically detected asymmetries in the era of artificial intelligence," *Egypt. J. Radiol. Nucl. Med.*, vol. 54, 2023, [Online]. Available: https://link.springer.com/article/10.1186/s43055-023-00979-1

[15] "Fusing Artificial Intelligence and Optimization with Trustworthy Optimization Proxies | SIAM." Accessed: Mar. 02, 2026. [Online]. Available: https://www.siam.org/publications/siam-news/articles/fusing-artificial-intelligence-and-optimization-with-trustworthy-optimization-proxies/

[16] "Swarm Intelligence: Foundations, Algorithms, Applications, and Future Directions." Accessed: Mar. 02, 2026. [Online]. Available: https://atoms.dev/insights/swarm-intelligence-foundations-algorithms-applications-and-future-directions/4af5c16ccd794582a9623355cb09c0cc

[17] Oliver Lester Saldanha, Jiefu Zhu, Gustav Müller-Franzes, "Swarm learning with weak supervision enables automatic breast cancer detection in magnetic resonance imaging," *Commun. Med.*, vol. 5, 2025, [Online]. Available: https://www.nature.com/articles/s43856-024-00722-5

[18] Julia Amann, Alessandro Blasimme, Effy Vayena, Dietmar Frey, "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective," *BMC Med. Inform. Decis. Mak.*, vol. 20, 2020, [Online]. Available: https://link.springer.com/article/10.1186/s12911-020-01332-6