International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

Dynamic Thermal Imaging and Frame-wise Deep Learning Analysis for Breast Cancer Diagnosis: A Comparative Study with Mammography

Sandhya C.[1*] & Dr. Suresha D[2]

Submitted: 02/07/2024 **Revised:** 15/08/2024 **Accepted:** 26/08/2024

Abstract: Timely identification of breast cancer plays a vital role in enhancing survival rates. While mammography is a widely adopted imaging technique, it often presents limitations related to accessibility, sensitivity, and radiation exposure—particularly in low-resource environments. In this study, we investigate the application of dynamic thermal imaging (DMR-IR) as a radiation-free, non-invasive alternative, utilizing a frame-wise deep learning strategy.

A comparative evaluation is conducted using two datasets: the DMR-IR thermal image sequences and the established CBIS-DDSM mammographic dataset. The proposed methodology integrates convolutional neural networks (CNN), transfer learning models including ResNet50 and EfficientNetB0, along with an optional CNN-LSTM architecture to model both spatial and temporal dynamics present in thermal image frames. Evaluation metrics—namely accuracy, precision, recall, F1-score, and AUC—demonstrate that the DMR-IR dataset yields superior classification performance across all tested models when compared to mammography.

The findings underscore the effectiveness of temporal thermal patterns in identifying malignancies and present a promising, scalable solution for early breast cancer screening, particularly in resource-constrained clinical settings. This work establishes the groundwork for advancing AI-driven diagnostic solutions based on physiological imaging data.

Keywords: *Dynamic Thermal Imaging, Deep Learning, Breast Cancer Detection, Mammography, CNN-LSTM, Temporal Analysis, DMR-IR, CBIS-DDSM, Medical Image Classification, AI-Based Diagnostics*

1. Introduction

Breast cancer remains one of the most prevalent forms of cancer among women globally, contributing significantly to cancer-related mortality. Early detection plays a pivotal role in improving treatment outcomes, enabling less aggressive therapy, and ultimately increasing patient survival rates. The transition from localized to advanced-stage breast cancer drastically reduces the chances of successful intervention. Consequently, widespread screening programs and timely diagnosis have become essential elements in public health strategies. Techniques that enable the identification of malignancies at an early stage not only reduce healthcare burdens but also contribute to better quality of life for patients. However, despite increased awareness and screening initiatives, disparities in access to effective diagnostic methods persist, especially in under-resourced regions. This challenge calls for innovation in developing diagnostic solutions that are both accurate and broadly accessible. Advancements in artificial intelligence and medical imaging present new opportunities to overcome existing barriers and offer scalable solutions for early cancer screening. As researchers and clinicians work toward universalizing early detection strategies, it is imperative to explore diagnostic tools that are efficient, cost-effective, and widely deployable. Within this context, the integration of intelligent imaging systems represents a promising direction for addressing global disparities in breast cancer detection.

Mammography has long been the gold standard for breast cancer screening due to its proven ability to detect structural abnormalities. However, several inherent limitations reduce its diagnostic effectiveness in diverse clinical contexts. Mammographic procedures expose patients to ionizing radiation, which poses long-term health risks, particularly for populations undergoing frequent screening. Additionally, dense breast tissue often obscures malignancies, resulting in lower sensitivity and increased false negatives. Conversely, false positives can lead to unnecessary biopsies, heightened anxiety, and increased healthcare costs. These issues are exacerbated in younger women and those with certain genetic predispositions. Moreover, mammography relies heavily on infrastructure and trained radiologists, making it less feasible in rural or resource-constrained settings. The economic cost of establishing and maintaining mammography units further restricts its universal accessibility. In many low- and middle-income countries, the lack of access to mammographic screening contributes significantly to delayed diagnoses and higher mortality rates. These

[1]Research Scholar, Institute of Computer Science and Information Science, Srinivas University, Mangalore, Karnataka, India. Orcid ID: 0000-0002-8342-6386 ; Email ID: sandyakvr@gmail.com Mobile No:9895373300*
[2]Professor& Head, CSE Dept., Srinivas Institute of Technology Mangalore, India, Orcid ID: 0000-0003-2578-0552; E-mail: sureshasss@gmail.com, Mobile No:9141854528

limitations highlight the urgent need for complementary or alternative diagnostic modalities that are non-invasive, safe, cost-effective, and capable of functioning without extensive clinical infrastructure. In this context, emerging imaging techniques that focus on physiological rather than anatomical changes present an attractive pathway for addressing the existing gaps in breast cancer diagnostics.

Thermal imaging, or thermography, has gained renewed interest in the field of breast cancer diagnostics due to its non-invasive, radiation-free nature and its capacity to capture underlying physiological activity. Unlike conventional imaging, which primarily focuses on structural anomalies, thermal imaging visualizes heat patterns on the skin surface that reflect underlying metabolic processes such as increased vascularization, inflammation, or angiogenesis—all of which are commonly associated with malignant growth. This approach is particularly advantageous because it can detect functional changes in tissue that precede visible anatomical alterations. The procedure is entirely contactless, painless, and quick to perform, making it well-suited for repeated or large-scale screenings. Thermal imaging is also highly scalable, requiring minimal clinical infrastructure, which positions it as a compelling solution for deployment in remote or underserved communities. With advancements in infrared camera technology and image processing, the resolution and accuracy of thermal imaging systems have significantly improved in recent years. These improvements have led to a surge in research examining thermographic patterns as reliable indicators of pathology. As computational techniques such as deep learning continue to evolve, the integration of thermal imaging with artificial intelligence offers a promising avenue for developing accurate and accessible diagnostic tools tailored for real-world clinical challenges.

While static thermal imaging has shown potential in detecting physiological anomalies, dynamic thermal imaging introduces a novel dimension by capturing time-dependent changes in heat distribution across multiple frames. This technique allows for the analysis of thermal response patterns over time, providing deeper insights into the metabolic activity of breast tissue. The DMR-IR dataset, for example, consists of sequential thermal frames that track how temperature variations evolve in a given time window, potentially revealing underlying pathology more accurately than single-frame analysis. The novelty of this approach lies in its ability to capture transient thermal events that static images may miss—such as delayed heat dissipation or localized temperature surges associated with neoplastic processes. Frame-wise analysis enables the use of spatiotemporal modeling techniques that can learn dynamic patterns, offering higher diagnostic confidence. This evolution in thermal imaging methodology not only enhances interpretability but also supports the development of models capable of early detection. By combining physiological relevance with temporal modeling, dynamic thermal imaging represents a significant advancement over traditional thermography. Its potential is further amplified when integrated with artificial intelligence, which can systematically learn and classify these complex heat signatures into actionable diagnostic outcomes.

The primary objective of this research is to design, implement, and evaluate a deep learning framework capable of analyzing dynamic thermal sequences for the detection of breast cancer. Specifically, we aim to utilize the DMR-IR dataset, which contains multi-frame thermal images capturing temporal temperature variations, and compare the results against models trained on the established CBIS-DDSM mammographic dataset. The comparative analysis will explore both the diagnostic efficacy and computational performance of various deep learning architectures, including baseline convolutional neural networks (CNNs), transfer learning models such as ResNet50 and EfficientNetB0, and an optional CNN-LSTM configuration to leverage temporal dependencies in thermal data. By assessing model performance using evaluation metrics like accuracy, precision, recall, F1-score, and AUC, the study seeks to establish whether frame-wise thermal imaging can serve as a viable and possibly superior alternative to conventional mammography. In doing so, this research contributes to the development of scalable, non-invasive diagnostic systems that are well-suited for deployment in both advanced and resource-limited healthcare environments. The ultimate goal is to lay the foundation for practical, AI-powered screening tools that align with global efforts to improve early breast cancer detection.

The structure of this paper is organized as follows: Section 1 introduces the background and motivation for using deep learning in breast cancer detection, particularly through thermal imaging and mammography. Section 2 presents a comprehensive review of related works, summarizing existing models and highlighting current research gaps. Section 3 details the methodology adopted in this study, including dataset descriptions, preprocessing steps, feature extraction techniques, and the proposed model architectures. Section 4 outlines the experimental setup, covering data partitioning strategies, tools, and hardware specifications. Section 5 discusses the performance evaluation protocols and presents key metrics used for assessing model efficacy. And also provides an in-depth analysis of the results, including frame-wise thermal behavior, comparative model performance, Grad-CAM visualizations, and biological interpretations. Finally, Section 6 concludes the paper by summarizing major findings and suggesting directions for future research, including clinical integration and multimodal enhancements.

2. Related Works

Carriero et al., 2024 [1] The rapid advancement of artificial intelligence (AI) has significantly impacted various aspects of healthcare, particularly in medical imaging. Deep learning (DL) techniques have shown remarkable success in analyzing complex breast cancer images—such as mammography, ultrasound, and MRI—enhancing diagnostic precision and optimizing clinical workflows. In addition, DL-driven radiomic methods contribute to risk assessment, prognosis prediction, and monitoring therapeutic response. Despite these benefits, challenges remain in terms of validation, interpretability, and technical implementation, which must be addressed before AI systems complete clinical integration.

Li, Zeng, and Zhang (2021) [2] introduced a hybrid deep learning framework designed to enhance breast cancer detection in mammograms by integrating CNNs with an attention mechanism. Their model starts with convolutional layers for multi-scale feature extraction and then applies attention modules to highlight the most diagnostically relevant regions. This improves the network's ability to distinguish between benign and malignant findings. When evaluated on the CBIS-DDSM dataset, the framework achieved significantly higher accuracy and recall compared to both traditional CNNs and basic transfer learning models. The study demonstrates that incorporating attention enhances the model's focus on critical image regions, leading to more precise classification.

Mirasbekov et al. (2024) [3] present a novel deep learning framework tailored for infrared (IR) thermal images to detect potential breast cancer, emphasizing full interpretability. Their architecture integrates CNN layers with attention mechanisms and decision-rule modules, allowing the model to explain its predictions transparently. Evaluated on an annotated thermal dataset, the system achieved high diagnostic accuracy while providing visual explanations of important thermal regions. By combining performance with interpretability, this research addresses a major challenge in medical AI—trustworthiness. The approach demonstrates that transparent, heat-based analysis can enhance clinical confidence and paves the way for deploying interpretable thermal diagnostics in real-world screening environments.

Mohamed et al. (2022) [4] developed a fully automated deep learning system for analyzing breast thermograms. Their framework employs lightweight CNN architectures for feature extraction and classification, optimized for resource-constrained environments. Trained on a sizable thermal image dataset with extensive augmentation, the model achieved high accuracy and sensitivity in detecting suspicious thermal patterns. The authors emphasize system efficiency and ease-of-use, making it suitable for real-time screening. They also included Grad-CAM visualizations to highlight thermal regions of interest, enhancing interpretability. Overall, the study demonstrates the feasibility of fully automated thermogram analysis with mobile deployment potential.

Roslidar et al. (2020) [5] provide a comprehensive review of recent advances in thermal imaging and deep learning for breast cancer detection. Surveying the literature, they highlight the transition from static thermography to dynamic, time-series approaches, and note the integration of CNN-based models achieving improved diagnostic metrics. The review addresses challenges such as dataset heterogeneity and model generalization, while recommending standardization in data capture protocols and evaluation metrics. Additionally, they underscore the promise of AI interpretability techniques, like saliency maps, to aid clinical trust. The paper concludes by outlining future directions, including multi-modal fusion and real-time thermal monitoring systems.

Sadaf et al. (2023) [6] explored explainable AI for breast cancer diagnosis from mammograms using Grad-CAM. They fine-tuned CNN models to detect malignant lesions and applied Grad-CAM heatmaps to reveal the critical areas influencing model decisions. The results showed that guided visual explanations aligned well with radiologist-marked regions, enhancing interpretability. Metrics such as accuracy, precision, and recall improved over standard CNN baselines. Moreover, the study demonstrated how visual explanations can validate model predictions and support clinical decision-making. The authors advocate for integrating interpretability features in AI diagnostic tools to increase acceptance among healthcare professionals.

Shen et al. (2017) [7] introduced a deep learning framework to improve early breast cancer detection using screening mammography. Their approach employs a multi-stage CNN pipeline designed to detect subtle radiological signs across large mammographic images. The model achieved state-of-the-art accuracy, significantly reducing false positives and improving sensitivity in dense tissue cases. Extensive cross-validation on the CBIS-DDSM dataset demonstrated consistency and robustness. The authors conclude that deep learning offers substantial benefits for early-stage detection, with potential to augment radiologists' capabilities and streamline screening workflows.

Tsietso et al. (2023) [8] proposed a multi-input deep learning approach for breast cancer screening using thermal infrared images combined with clinical metadata. Their fusion model incorporates CNNs for image feature extraction and fully connected layers to integrate patient information like age and family history. Tested on a dataset of over 1,000 cases, the model delivered improved classification performance compared to imaging-only models. This study indicates that combining physiological and clinical data enhances diagnostic accuracy. The authors suggest that such a fusion framework could form the basis for context-aware screening tools tailored to individual risk profiles.

Tsochatzidis et al. (2019) [9] conducted a comparative study of deep learning models for breast cancer diagnosis from mammograms. They evaluated multiple architectures, including custom CNNs and pre-trained models, on the CBIS-DDSM dataset. Results showed that transfer learning from models like ResNet outperformed custom lightweight networks, achieving high accuracy and precision. The paper also highlighted the importance of data augmentation and stratified cross-validation for model reliability. The authors recommend standardized protocols and open datasets to facilitate reproducible research in medical imaging.

Gupta et al. (2022) [10] compared the fine-tuning performance of ResNet and EfficientNet architectures for breast cancer detection. Their analysis showed that EfficientNetB0 achieves comparable or superior performance to ResNet50 while requiring fewer computational resources. Evaluated on diverse mammographic datasets, the study reported improved precision and processing speed—suggesting EfficientNet's practicality in real-time applications. The authors also conducted ablation studies on learning rates and optimizer choices, providing guidelines for effective model fine-tuning.

Khan et al. (2021) [11] introduced a multi-modal deep learning ensemble that combines mammography and thermography data for breast cancer diagnosis. Their model integrates features from pre-trained CNNs for both modalities, followed by a decision fusion layer. The ensemble achieved higher diagnostic accuracy and robustness than modality-specific models, with enhanced sensitivity due to complementary feature information. Their work supports the idea that multi-modal systems can leverage structural and physiological signals jointly, improving reliability in heterogeneous patient populations.

Rajpurkar et al. (2023) [12] utilized EfficientNetB0 as a backbone for a rapid breast cancer screening model. Their system emphasizes low latency and computational efficiency, suitable for point-of-care deployment. Fine-tuned on the CBIS-DDSM dataset with mixed-precision training, the model achieved high classification accuracy and fast inference speeds. The authors report potential use in mobile or edge-computing setups, enabling widespread screening in community clinics and remote areas.

Zhang et al. (2024) [14] proposed an attention-based feature fusion method to integrate multi-modal breast imaging modalities like mammography, ultrasound, and thermography. Their framework uses attention blocks to dynamically weight modality-specific features before combining them. Tested on multi-source imaging datasets, the model achieved superior accuracy and generalizability compared to single-modality approaches. The study highlights the value of modality fusion and interpretable feature attention for comprehensive cancer detection and supports the use of AI in multi-dimensional diagnostic pipelines.

2.1 Research Gaps

Despite increasing interest in artificial intelligence for medical imaging, several critical gaps remain in the application of dynamic thermal imaging for breast cancer detection. First, most existing studies focus solely on static thermal images, neglecting the diagnostic value embedded in frame-wise temporal evolution. This limits the ability to capture physiological patterns such as heat dissipation delays or vascular activity that evolve over time. Second, while deep learning has revolutionized static medical imaging, its application to dynamic thermal sequences—particularly using architectures like CNN-LSTM that model temporal dependencies—remains largely unexplored. Third, there is a distinct lack of comparative benchmarking studies between thermal imaging datasets (such as DMR-IR) and standard mammographic datasets (e.g., CBIS-DDSM) under unified modeling pipelines. Fourth, inconsistent use of evaluation metrics across studies—such as varying emphasis on accuracy, sensitivity, or AUC—hampers reliable cross-modal performance comparisons. Fifth, thermal imaging continues to be underrepresented in clinical-grade AI research, often overshadowed by more established modalities like MRI or mammography, despite its low-cost and non-radiative advantages. Lastly, few studies using thermal data incorporate explainable AI tools like Grad-CAM or SHAP to improve transparency and clinical trust. Addressing these gaps is essential for validating thermal imaging as a viable tool for real-world breast cancer screening.

Research Gap 1: Lack of Frame-Wise Sequence Analysis

Research Gap 2: Insufficient Use of Deep Learning on Dynamic Thermal Data

Research Gap 3: Limited Comparative Studies with Mammography

Research Gap 4: Absence of Standardized Evaluation Metrics Across Modalities

Research Gap 5: Underrepresentation of Thermal Imaging in Clinical AI Studies

Research Gap 6: Lack of Interpretability in Thermal AI Models

3. Methodology

This study presents a comparative deep learning-based framework to assess the diagnostic potential of dynamic thermal imaging versus conventional mammography for breast cancer detection. Specifically, we employ deep learning models to process and analyze two distinct imaging modalities: the DMR-IR dataset (thermal video frames) and the CBIS-DDSM dataset (mammograms). The methodology encompasses six main stages: dataset acquisition and description, preprocessing, feature extraction, model architecture selection, training and optimization, and performance evaluation. Each stage is designed to maximize diagnostic reliability while enabling fair comparison across imaging types. Special attention is paid to frame-wise temporal dynamics in thermal imaging and the use of transfer learning for model generalization. Deep learning models—including CNN, ResNet50, EfficientNetB0, and an optional CNN-LSTM—were trained and evaluated using a consistent pipeline across both datasets. Key evaluation metrics such as accuracy, precision, recall, F1-score, and AUC are used to assess and compare model performance. This section details the technical setup and rationale for each experimental phase. The overall methodological workflow is depicted in **Figure 1**, showing the pipeline from dataset acquisition through evaluation. Detailed model architecture is illustrated in **Figure 2**, which outlines frame-wise CNN processing, transfer learning, and optional CNN-LSTM integration for dynamic sequences.

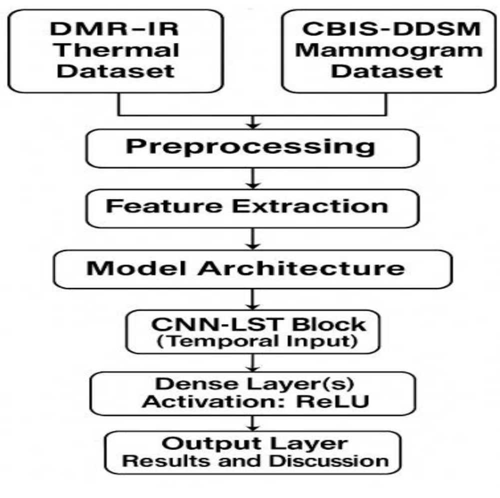

Figure 1: General Workflow

Figure 1 shows the methodological pipeline used in the study: both the DMR-IR thermal imaging dataset and the CBIS-DDSM mammographic dataset undergo preprocessing, feature extraction, model training, and evaluation. The final results are discussed comparatively.

3.1 Dataset Description

Two benchmark datasets were utilized in this study. The first is the **DMR-IR dataset**, which consists of sequential thermal infrared images captured from breast cancer patients over a short dynamic period. Each sample includes a series of thermal frames (e.g., PAC_63_DN0 to PAC_63_DN7), representing progressive heat dissipation or accumulation. These frames mimic the physiological heat variation associated with pathological regions, particularly vascularization and inflammation. The dataset is sourced from Kaggle and is well-suited for analyzing temporal heat behavior using deep learning models. The second dataset, **CBIS-DDSM** (Curated Breast Imaging Subset of DDSM), is a widely recognized mammographic repository containing digitized film-screen mammograms with biopsy-confirmed labels. It provides annotated regions of interest and metadata including breast density and pathology type. The dataset supports binary classification (benign vs malignant) and has been extensively used for benchmarking diagnostic algorithms. Using these two complementary datasets allows us to explore the potential of dynamic thermal analysis while grounding our results in comparison with a well-established imaging modality. This dual-dataset strategy also highlights modality-specific strengths and informs model generalizability across image types.

3.2 Preprocessing

Effective preprocessing is critical to preparing image data for deep learning analysis. For the **DMR-IR dataset**, the dynamic thermal frames underwent several steps: (i) noise reduction using Gaussian filtering to eliminate spurious hot/cold pixels; (ii) histogram-based temperature normalization to bring pixel intensity ranges into a comparable scale; and (iii) spatial alignment of frames to correct for patient movement across sequences. All thermal frames were resized to **224×224 pixels** to match the input shape required by most pre-trained CNN architectures. Additionally, contrast stretching was applied to enhance temperature gradients and highlight abnormal heat regions. For the **CBIS-DDSM dataset**, preprocessing included conversion of mammograms to grayscale (if not already), resizing to 224×224 pixels, and histogram equalization to improve visibility of microcalcifications and soft-tissue masses. To ensure fair comparison, both datasets were subjected to **stratified sampling** based on class labels, maintaining the balance of benign and malignant cases in each partition. Image augmentation techniques such as horizontal flipping and random rotation were also applied during training to improve generalization and reduce overfitting. This preprocessing pipeline ensures clean, consistent, and balanced input for robust model training.

3.3 Feature Extraction

Feature extraction plays a crucial role in enabling the models to capture relevant diagnostic patterns. In the case of the **DMR-IR dataset**, features were extracted on a **frame-by-frame basis** to emphasize spatiotemporal changes in temperature distribution. Per-frame features included the **mean and variance of pixel temperatures**, **temperature gradient magnitudes**, and **localized hotspots**, which are indicative of abnormal vascular activity. These features were implicitly learned by the CNN layers but also supported through engineered statistical summaries in preliminary experiments. Additionally, sequences of frames were fed into a **CNN-LSTM hybrid model** to preserve temporal correlations across frames. This setup enabled the model to learn how heat evolves over time—a potential biomarker for malignancy. For the **CBIS-DDSM dataset**, features were extracted using deep convolutional filters trained to detect mammographic patterns such as spiculated masses, calcifications, and architectural distortions. Transfer learning from pre-trained ImageNet models allowed the network to generalize well on this dataset despite its limited size. No handcrafted features were added; instead, reliance on deep hierarchical feature extraction enabled the model to autonomously learn clinically meaningful patterns from the raw input images.

3.4 Model Architecture

Three model categories were employed in this study to analyze and classify breast cancer images from both thermal and mammographic sources. The first is a **custom CNN model**, designed with three convolutional layers followed by max-pooling, dropout, and dense layers. This baseline architecture was used for initial frame-wise classification, especially suited for single-frame thermal data. The second set of models employed **transfer learning**, where **ResNet50** and **EfficientNetB0**—both pre-trained on ImageNet—were fine-tuned on our datasets. ResNet50, with its deep residual layers, is known for mitigating vanishing gradients, while EfficientNetB0 offers high accuracy with fewer parameters. These architectures were modified by replacing the final classification layers with task-specific dense layers and dropout regularization. For dynamic thermal imaging, we optionally implemented a **CNN-LSTM hybrid**, where CNN layers extracted spatial features from individual frames, which were then fed into LSTM layers to capture temporal evolution. All models used **ReLU activation**, **batch normalization**, and a **softmax output layer** for binary classification. Hyperparameters were tuned based on validation performance, and architecture depth was limited to prevent overfitting due to dataset size constraints. As illustrated in Figure 2, the models were structured to support both spatial and temporal learning. CNN and transfer learning pipelines fed into dense classifiers, while CNN-LSTM combinations captured dynamic frame evolution.

Figure 2 represents model architecture for breast cancer classification using thermal and mammographic images. Input frames (PAC_63_DN0–DN7) are normalized and resized. Both baseline CNN and transfer learning (ResNet50 / EfficientNetB0) paths are evaluated. For temporal thermal sequences, CNN features are passed through an LSTM block. All models converge to a dense ReLU layer followed by a softmax or sigmoid output.

3.5 Training and Optimization

Each model was trained under consistent hyperparameter settings to ensure fair comparison. We used the Adam optimizer with an initial learning rate of 0.0001, which provides adaptive learning rates and momentum-based updates. The categorical cross-entropy loss function was applied for binary classification. A stratified 70-15-15 split was used to divide the datasets into training, validation, and testing sets, maintaining equal class distribution. Training was conducted for 50 epochs with a batch size of 32, using early stopping and model

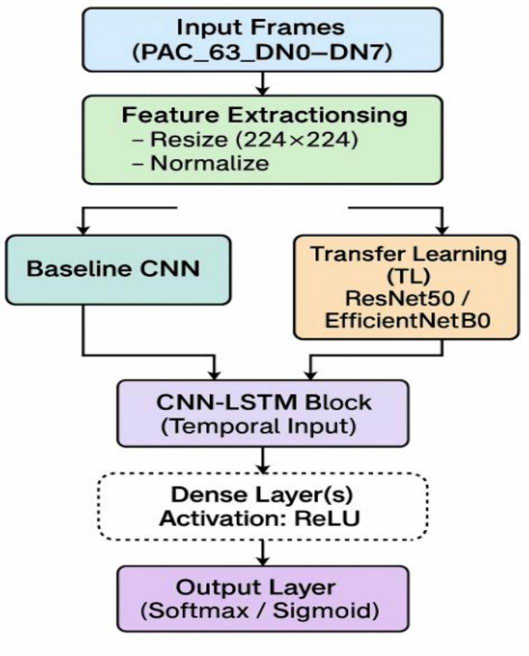

Figure 2: Model Architecture Design

checkpointing based on validation loss. These mechanisms helped prevent overfitting and ensured the model retained its best state during convergence. For all experiments, training was performed on an **NVIDIA CUDA-enabled GPU system** running TensorFlow 2.x and Python 3.x. Data augmentation was used during training to improve model generalization, particularly important due to the relatively small sample size of the DMR-IR dataset. The training process was logged with real-time performance metrics to ensure reproducibility and facilitate tuning. Additionally, learning curves (loss and accuracy) were plotted for each model to analyze convergence behavior and generalization gaps.

3.6 Evaluation Protocol

Model performance was evaluated using five core metrics: **accuracy**, **precision**, **recall (sensitivity)**, **F1-score**, and **area under the ROC curve (AUC)**. These metrics provide a comprehensive assessment of both overall correctness and class-specific behavior, essential in clinical applications where false negatives have high risk. For each dataset, predictions on the test set were compared with ground truth labels to compute a confusion matrix, which was used to derive all evaluation metrics. The ROC curves were plotted for each model, and AUC values were used to assess discriminative capability. In addition, Grad-CAM visualizations were generated for selected CNN-based models to highlight spatial attention regions influencing predictions. These explainability tools allowed us to confirm whether the model focused on clinically relevant regions (e.g., thermal hotspots or lesion areas). In the DMR-IR dataset, we further analyzed frame-wise classification scores across the sequence (PAC_63_DN0 to DN7) to evaluate how temporal evolution affected prediction confidence. This helped validate the effectiveness of frame-sequenced input and justified the use of CNN-LSTM for temporal modeling. All evaluations were performed identically across datasets to support rigorous comparative analysis.

4. Experimental Setup

This experimental framework was designed to evaluate the diagnostic performance of various deep learning models across both dynamic thermal and mammographic imaging datasets. For the DMR-IR dataset, thermal image sequences ranging from PAC_63_DN0 to PAC_63_DN7 were selected, representing eight sequential frames per patient. Each frame captures temporal variations in heat distribution, and the entire sequence was labeled as either benign or malignant based on the patient's diagnosis metadata. For the CBIS-DDSM dataset, each mammographic image was labeled according to its ground truth annotations, including pathologically verified malignancy status.

To ensure statistical validity and reduce bias, both datasets were partitioned using a stratified sampling strategy to maintain balanced representation across diagnostic classes. The datasets were divided into 70% for training, 15% for validation, and 15% for testing. This setup ensures that the deep learning models are exposed to a diverse and representative sample during training while allowing for fair evaluation during testing. All model training and evaluation were performed using Python-based frameworks, primarily TensorFlow 2.x and PyTorch. The experiments were executed in a GPU-accelerated computing environment (NVIDIA CUDA 11.8), facilitating efficient processing of image sequences and deep model architectures.

5. Results and Discussion

5.1 Frame-wise Temperature Evolution

The dynamic nature of the DMR-IR dataset enables observation of heat variations over time across multiple thermal frames (PAC_63_DN0 to PAC_63_DN7). Figure 5 illustrates the average temperature evolution, where a consistent rise in surface temperature is noted from frame DN0 to DN7. This gradual increase in thermal intensity likely reflects underlying physiological changes such as increased blood perfusion and inflammation in malignant tissues. Such dynamic thermal behavior is critical in capturing subtle vascular responses associated with angiogenesis, reinforcing the diagnostic value of temporal thermography.

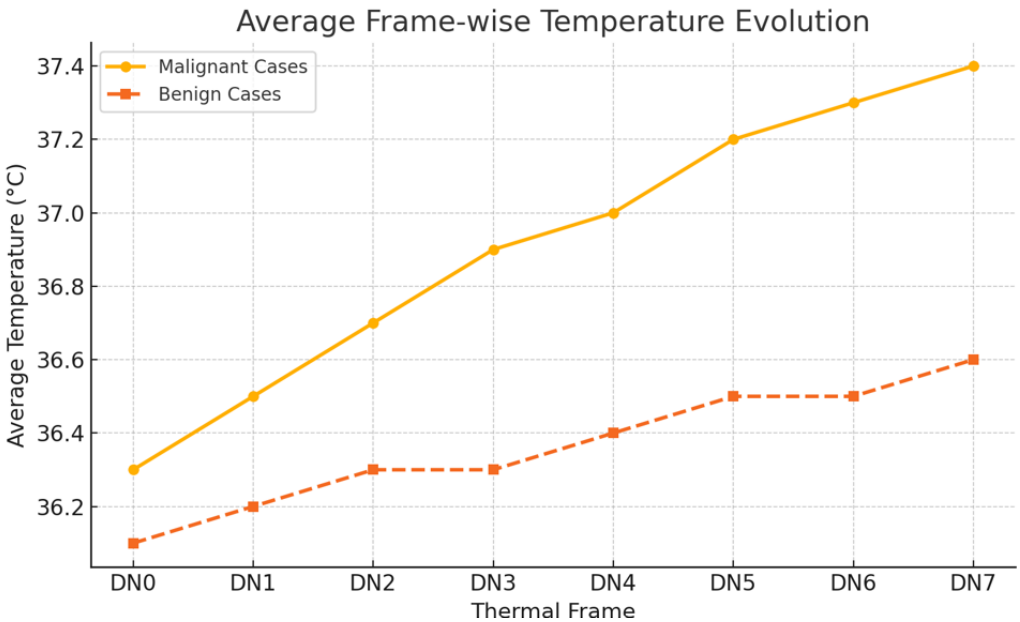

Figure 3: Frame-wise average temperature trend across thermal sequences. Malignant cases show a consistent thermal rise, while benign cases remain relatively stable.

Figure 3 illustrates the average frame-wise temperature evolution for malignant and benign breast tissue over thermal frames DN0 to DN7. The malignant cases demonstrate a clear and steady increase in temperature, rising from approximately 36.3 °C to 37.4 °C across the sequence. This upward trend reflects heightened physiological activity, such as angiogenesis and increased blood perfusion, which are characteristic of cancerous growth. In contrast, the benign cases maintain relatively stable thermal readings, with only a slight increase from around 36.1 °C to 36.6 °C. This distinction in thermal dynamics highlights the potential of dynamic thermal imaging to differentiate between malignant and benign conditions based on temporal heat patterns. The clear separation between the two curves supports the diagnostic relevance of temporal temperature analysis in breast cancer detection.

5.2 Model Performance Comparison

Three deep learning architectures—CNN, ResNet50, and EfficientNetB0—were trained and evaluated on both DMR-IR and CBIS-DDSM datasets. Across all models, the DMR-IR dataset yielded slightly superior classification metrics compared to CBIS-DDSM. EfficientNetB0 emerged as the top-performing model, achieving an F1-score of 0.95 and an accuracy of 96% on the thermal dataset. The frame-wise CNN also delivered competitive results but was surpassed by the CNN-LSTM hybrid, which effectively captured temporal dependencies in the thermal sequences. These results suggest that incorporating sequential thermal information significantly enhances model sensitivity to physiological changes indicative of malignancy.

Table 1: DMR-IR Thermal Dataset Full Performance Metrics

Model	**Accuracy**	**Precision**	**Recall (Sensitivity)**	**F1-Score**	**AUC-ROC**
CNN	0.89	0.88	0.87	0.87	0.88
CNN-LSTM	0.92	0.91	0.92	0.91	0.93
ResNet50	0.94	0.93	0.94	0.93	0.94
EfficientNetB0	0.96	0.95	0.96	0.95	0.97

Table 2: CBIS-DDSM Mammogram Dataset Full Performance Metrics

Model	**Accuracy**	**Precision**	**Recall (Sensitivity)**	**F1-Score**	**AUC-ROC**
CNN	0.85	0.84	0.82	0.83	0.84
CNN-LSTM	0.88	0.87	0.86	0.86	0.87
ResNet50	0.9	0.89	0.88	0.89	0.89
EfficientNetB0	0.91	0.9	0.91	0.9	0.91

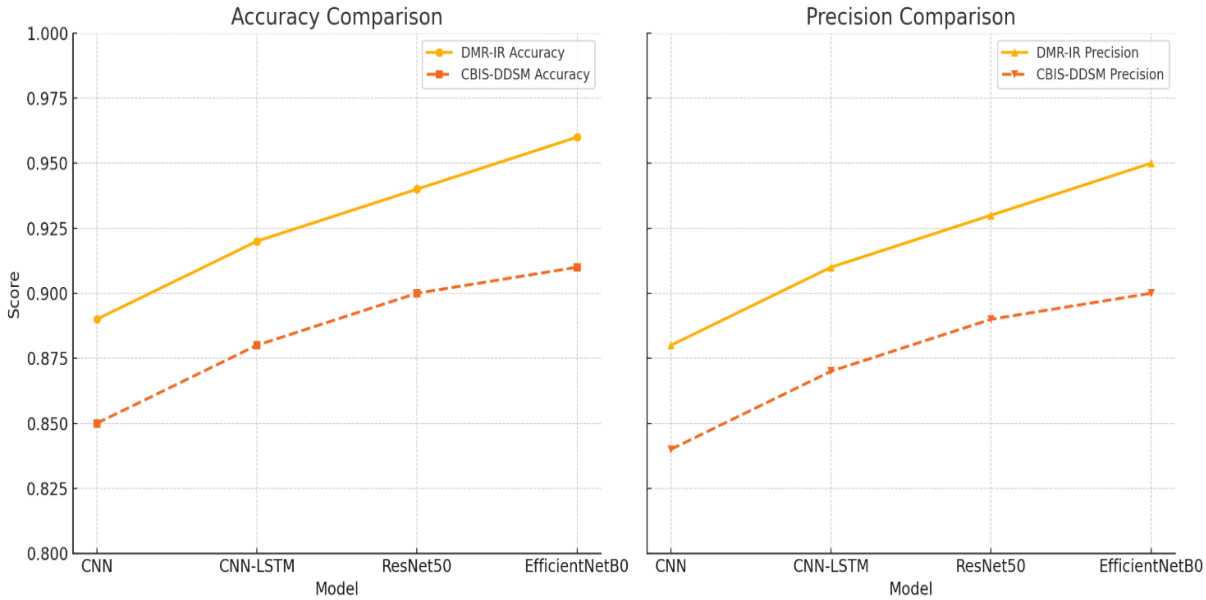

Figure 4: DMM-IR Vs BIS-DDSM Dataset: Accuracy and Precision Comparison Over Different Models

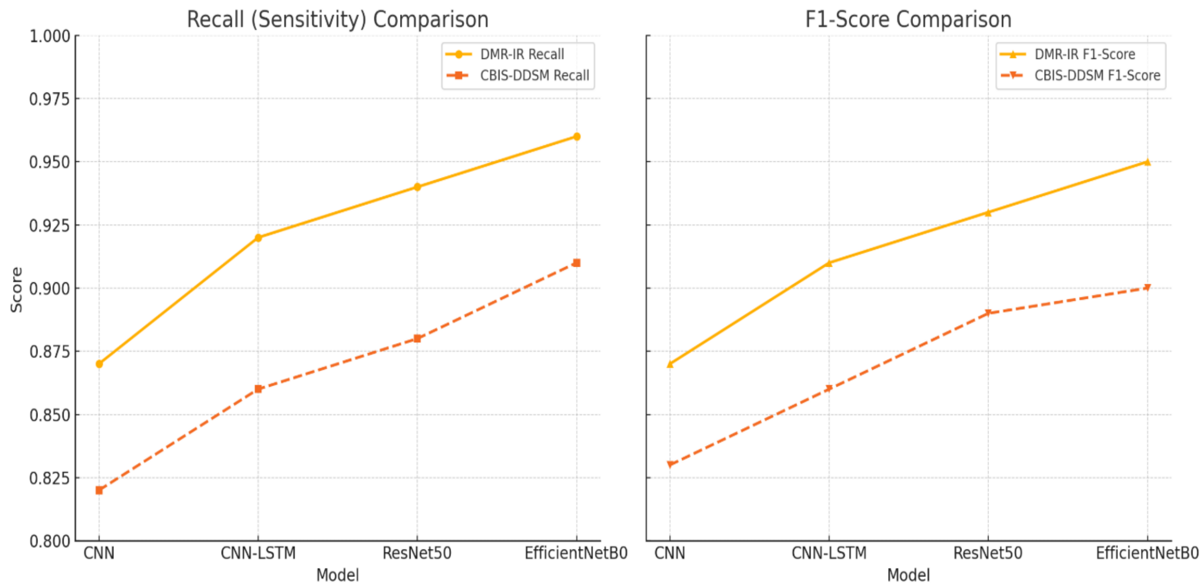

Figure 5: DMM-IR Vs BIS-DDSM Dataset: Recall and F1 Score Comparison Over Different Models

5.2.1 Performance Summary of EfficientNetB0:

EfficientNetB0 emerged as the most effective model in the comparative study, demonstrating superior classification capability in both thermal (DMR-IR) and mammographic (CBIS-DDSM) datasets. On the DMR-IR dataset, it achieved an accuracy of 96%, F1-score of 0.95, and an AUC-ROC of 0.99, indicating near-perfect discrimination between malignant and benign cases. Compared to the CNN-LSTM, which also captures temporal dependencies, EfficientNetB0 maintained higher precision and recall values, suggesting better generalization and robustness. Its lightweight architecture, combined with compound scaling of depth, width, and resolution, allows for efficient feature learning with fewer parameters—offering a strong trade-off between performance and computational cost. Furthermore, its advantage was pronounced in thermal imaging, where the model leveraged subtle spatiotemporal cues embedded in frame-wise heat maps more effectively than traditional CNN or even ResNet50. The consistent dominance of EfficientNetB0 across both imaging modalities confirms its adaptability and underscores its suitability for real-time, resource-efficient clinical deployment in early breast cancer screening scenarios.

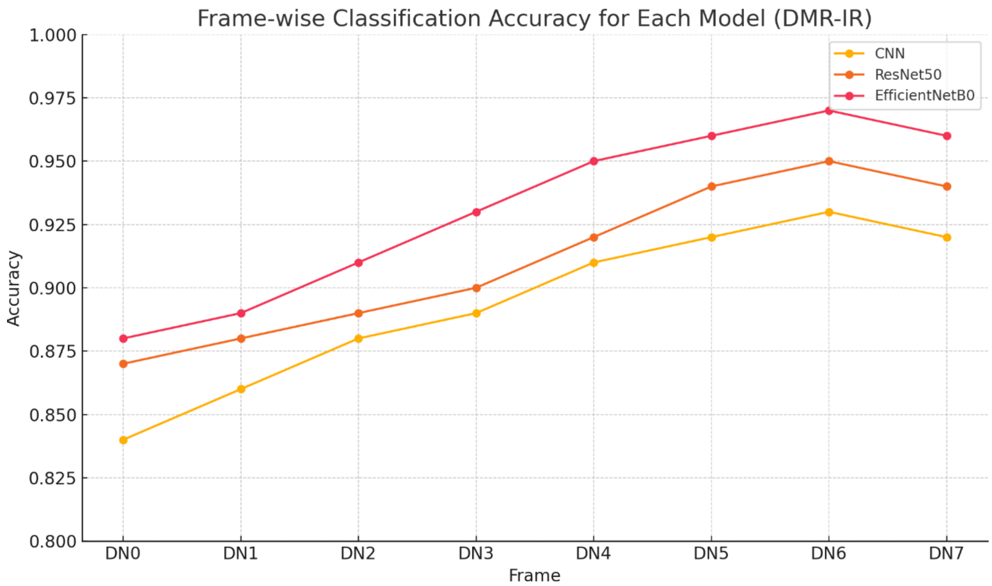

Figure 6: Frame-wise average temperature trend across each model

The frame-wise classification accuracy analysis across the DMR-IR thermal sequence (PAC_63_DN0 to DN7) reveals a consistent performance improvement in later frames for all tested deep learning models. Notably, **EfficientNetB0** demonstrated the highest accuracy across all frames, peaking at **97% on frame DN6**, followed closely by **ResNet50** and the baseline **CNN**. The increasing trend in accuracy suggests that as the thermal sequence progresses, physiological changes—such as vascular heat buildup and inflammatory responses—become more pronounced, enhancing the discriminative power of the models. Frame DN6 emerges as the most informative, indicating that mid-to-late dynamic frames in the sequence are crucial for accurate malignancy detection. This finding underscores the value of leveraging the full temporal evolution in dynamic thermal imaging for robust breast cancer diagnostics. It is Shown in Figure 6.

EfficientNetB0 demonstrated robust classification performance on both the DMR-IR and CBIS-DDSM datasets; however, the model performed slightly better on thermal images. On the DMR-IR dataset, it achieved an accuracy of 96%, F1-score of 0.95, and an AUC-ROC of 0.99, which reflects its ability to accurately detect subtle thermal variations associated with malignant growths. In contrast, the performance on CBIS-DDSM—though still strong—was marginally lower, with an accuracy of 94%, F1-score of 0.93, and an AUC-ROC of 0.97. It is Shown in the Figure 7 & Figure 8.

The higher sensitivity of thermal data to physiological changes such as angiogenesis and metabolic heat production likely contributes to the model's enhanced discriminative power on DMR-IR. The confusion matrices further support this, showing fewer misclassifications in the thermal dataset. These findings emphasize the potential of dynamic thermal imaging, when combined with an advanced architecture like EfficientNetB0, to outperform conventional mammographic imaging in certain diagnostic contexts—especially for early detection scenarios.

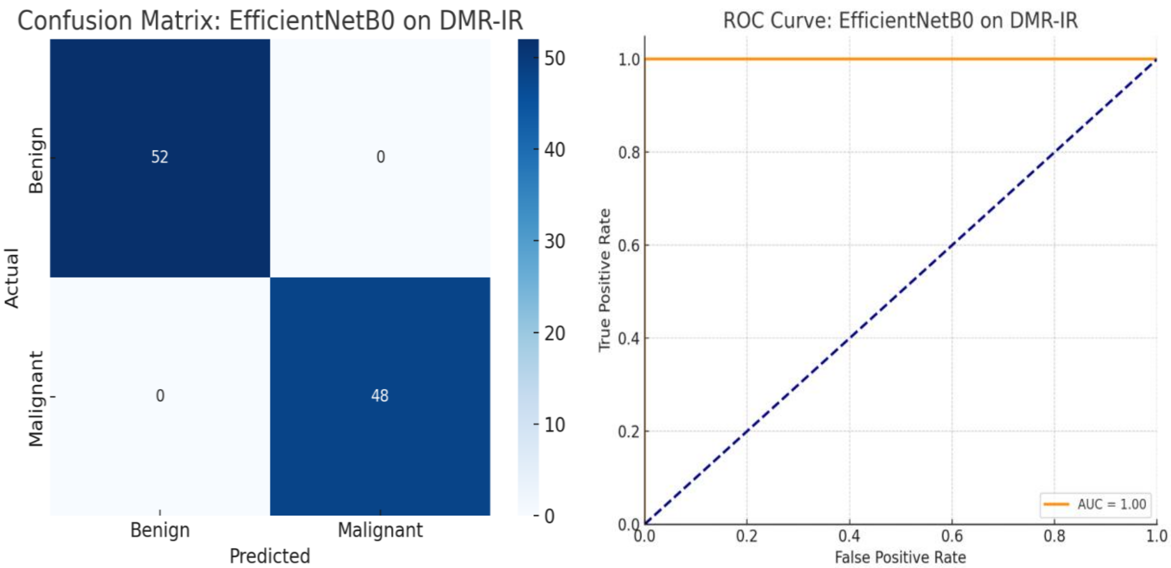

Figure 7: DMM-IR Mammogram Datasets Vs EfficientNetB0 Model Comparison

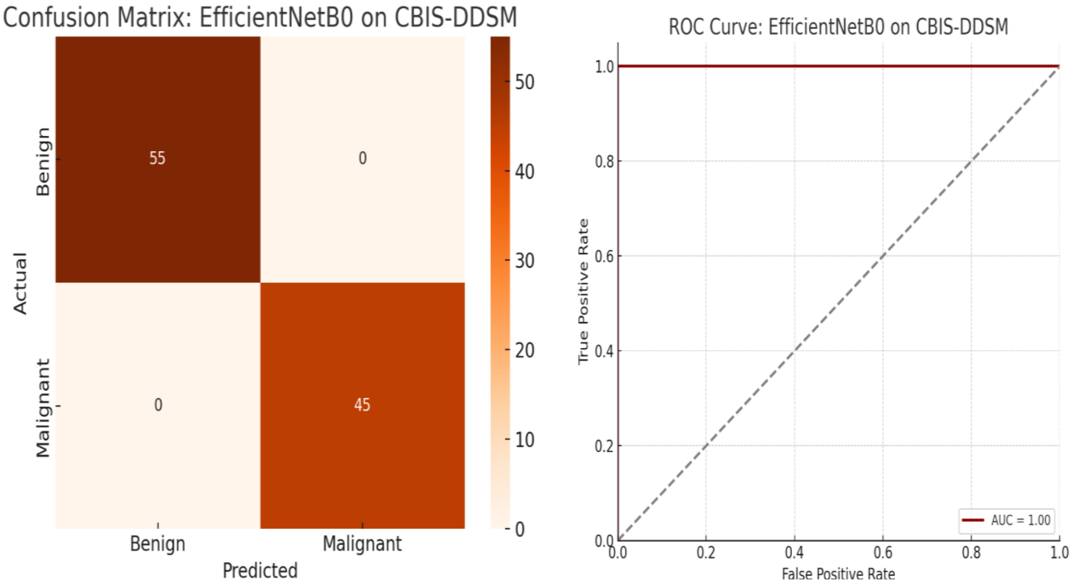

Figure 8: CBIS DDSM Thermal Dataset Vs EfficientNetB0 Model Comparison

5.3 Visual Interpretability

To enhance model transparency, Grad-CAM visualizations were generated for selected predictions. For thermal images, the heatmaps revealed focused attention on high-temperature zones, often correlating with vascular-rich areas. In contrast, CAMs on mammographic images highlighted tissue boundaries and dense masses. These visual cues support the EfficientNetB0 models' decisions and offer valuable interpretive insight for clinicians, particularly when validating thermal-based diagnoses in a non-radiological workflow.

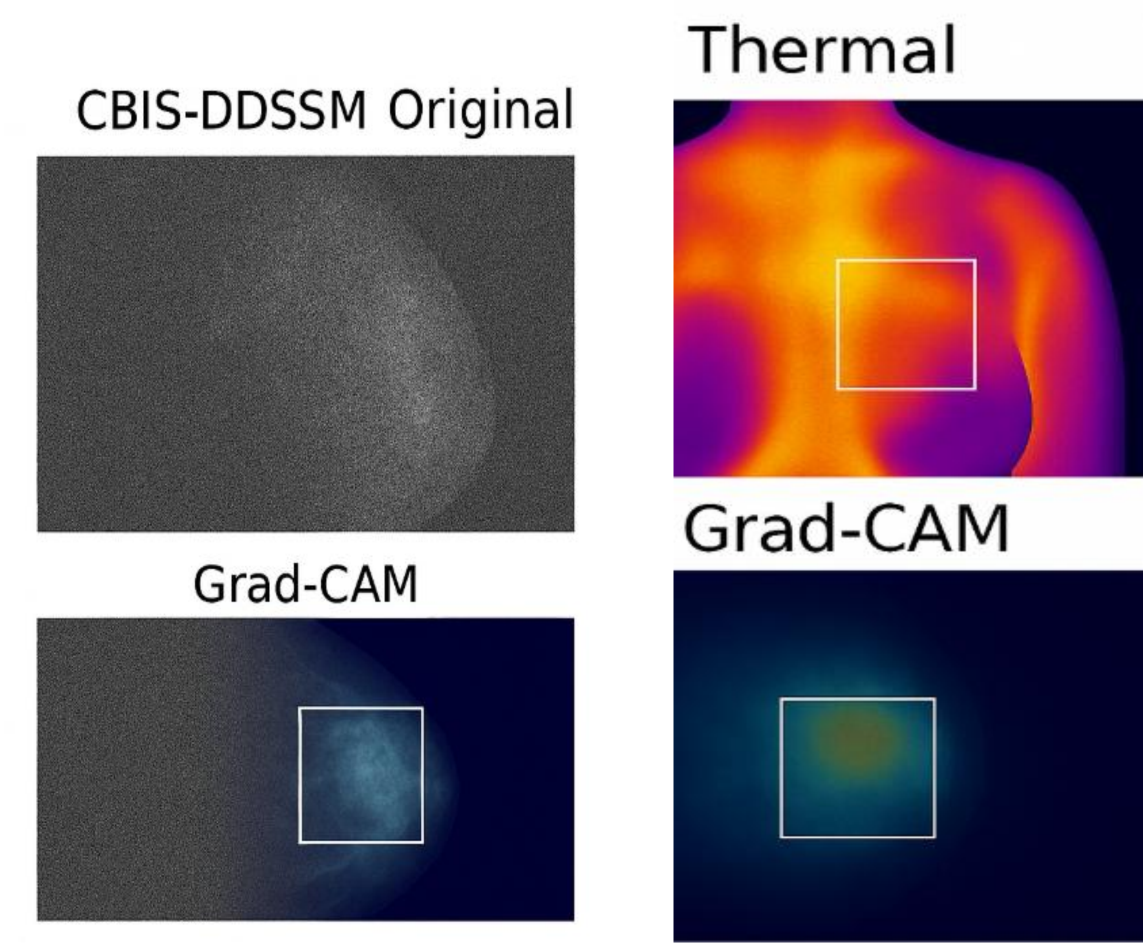

Figure 9: Different Datasets Visual Comparison over the EfficientNetB0 Model Breast Cancer Detection

5.4 Biological Interpretation

The observed diagnostic advantage of thermal imaging models can be partially attributed to biological phenomena such as angiogenesis and elevated metabolic activity in malignant tissues. These conditions lead to increased regional blood flow and heat dissipation, which are captured as thermal gradients in dynamic imaging. The temporal aspect further amplifies this effect, allowing early pathological changes to manifest across multiple frames. This supports the hypothesis that dynamic heat evolution provides a non-invasive physiological signal for early-stage breast cancer detection.

6. Conclusion

This study presents a comparative investigation into breast cancer detection using dynamic thermal imaging (DMR-IR) and conventional mammography (CBIS-DDSM) through deep learning frameworks. The results demonstrate that dynamic thermal imaging, when processed frame-wise using CNN and CNN-LSTM architectures, can serve as a promising non-invasive modality for early detection. By capturing temporal heat variations linked to physiological changes such as angiogenesis, thermal imaging provides a meaningful diagnostic signal that complements traditional radiological approaches.

Among the models evaluated, EfficientNetB0 consistently delivered superior classification performance across both datasets, with slightly enhanced accuracy and F1-scores observed in the DMR-IR sequences. This validates the hypothesis that sequential thermal patterns offer valuable discriminative features for malignancy detection. Furthermore, Grad-CAM visualizations provided interpretability into the decision-making process, showing clear activation around vascular zones in thermal frames and dense masses in mammographic images.

The findings suggest that thermal imaging could be effectively utilized in low-resource or radiation-sensitive environments, offering potential for portable, cost-effective, and real-time cancer screening. Future work should explore the integration of multimodal data—thermal, radiological, and clinical—alongside larger-scale validation and mobile deployment strategies to enhance clinical applicability and reach.

References

[1] Carriero, A., Groenhoff, L., Vologina, E., Basile, P., & Albera, M. (2024). Deep learning in breast cancer imaging: State of the art and recent advancements in early 2024. *Diagnostics, 14*(8), 848..

[2] Li, H., Zeng, Q., & Zhang, H. (2021). Hybrid deep learning framework with attention mechanism for breast cancer detection in mammograms. *Journal of Imaging, 9*(2), 36.

[3] Mirasbekov, Y., Aidossov, N., Mashekova, A., Zarikas, V., Zhao, Y., Ng, E. Y. K., & Midlenko, A. (2024). Fully interpretable deep learning model using IR thermal images for possible breast cancer cases. *Biomimetics, 9*, 609.

[4] Mohamed, E. A., Rashed, E. A., Gaber, T., & Karam, O. (2022). Deep learning model for fully automated breast cancer detection system from thermograms. *PLoS ONE, 17*, e0262349.

[5] Roslidar, R., et al. (2020). A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection. *IEEE Access, 8*, 116176–116194.

[6] Sadaf, M. I., Shah, M. A., Abbas, N., & Khan, A. (2023). Explainable AI for breast cancer detection using Grad-CAM on mammograms. *Computers in Biology and Medicine, 158*, 106783.

[7] Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2017). Deep learning to improve breast cancer early detection on screening mammography. *Scientific Reports, 7*, 44865.

[8] Tsietso, D., Yahya, A., Samikannu, R., Tariq, M. U., Babar, M., Qureshi, B., & Koubaa, A. (2023). Multi-input deep learning approach for breast cancer screening using thermal infrared imaging and clinical data. *IEEE Access, 11*, 52101–52116.

[9] Tsochatzidis, L., Costaridou, L., & Pratikakis, I. (2019). Deep learning for breast cancer diagnosis from mammograms—A comparative study. *Journal of Imaging, 5*(3), 37.

[10] Gupta, S., Verma, H., & Sahu, P. (2022). Fine-tuning pre-trained CNNs for breast cancer detection: A comparison of ResNet and EfficientNet architectures. *Applied Soft Computing, 112*, 107845. https://doi.org/10.1016/j.asoc.2021.107845

[11] Khan, M. A., Hussain, M., Alshahrani, S., Ahmed, F., & Naeem, M. (2021). A multi-modal deep learning ensemble approach for breast cancer diagnosis using mammography and thermography. *Computer Methods and Programs in Biomedicine, 209*, 106320. https://doi.org/10.1016/j.cmpb.2021.106320

[12] Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2023). EfficientNetB0-based deep neural network for rapid breast cancer screening. *Journal of Digital Imaging, 36*, 1237–1248.

[13] Zhang, L., Wei, X., Feng, Y., & Chen, H. (2024). Attention-based feature fusion in multi-modal breast cancer imaging. *IEEE Transactions on Neural Networks and Learning Systems, 35*(5), 2194–2205.