

Article

Enhancing Early Breast Cancer Detection with Infrared Thermography: A Comparative Evaluation of Deep Learning and Machine Learning Models

Reem Jalloul ^{1,*}, Chethan Hasigala Krishnappa ² , Victor Ikechukwu Agughasi ³  and Ramez Alkhatib ⁴ 

¹ Maharaja Research Foundation, University of Mysore, Mysuru 570005, India

² Department of Computer Science and Engineering, Maharaja Research Foundation, Maharaja Institute of Technology, Mysuru 571477, India; hkchethan@gmail.com

³ Department of Computer Science and Engineering (Artificial Intelligence), Maharaja Institute of Technology, Mysuru 571477, India; victor.agughasi@gmail.com

⁴ Biomaterial Bank Nord, Research Center Borstel Leibniz Lung Center, Parkallee 35, 23845 Borstel, Germany; ralkhatib@fz-borstel.de

* Correspondence: reem1988jalloul@gmail.com; Tel.: +91-7204818328

Abstract: Breast cancer remains one of the most prevalent and deadly cancers affecting women worldwide. Early detection is crucial, particularly for younger women, as traditional screening methods like mammography often struggle with accuracy in cases of dense breast tissue. Infrared thermography offers a non-invasive imaging alternative that enhances early detection by capturing subtle thermal variations indicative of breast abnormalities. This study investigates and compares the performance of various deep learning and machine learning models in analyzing thermographic data to classify breast tissue as healthy, benign, or malignant. To maximize detection accuracy, data preprocessing, feature extraction, and dimensionality reduction were implemented to isolate distinguishing characteristics across tissue types. Leveraging advanced feature extraction and visualization techniques inspired by geospatial data methodologies, we evaluated several deep learning architectures and classical classifiers using the DRM-IR and Breast Thermography Mendeley thermal datasets. Among the tested models, the ResNet152 architecture combined with a Support Vector Machine (SVM) classifier delivered the highest performance, achieving 97.62% accuracy, 95.79% precision, 98.53% recall, 94.52% specificity, an F1 score of 97.16%, an area under the curve (AUC) of 99%, a latency of 0.06 s, and CPU utilization of 88.66%. These findings underscore the potential of integrating infrared thermography with advanced deep learning and machine learning approaches to significantly improve the accuracy and efficiency of breast cancer detection, supporting its role as a valuable tool for early diagnosis.

Keywords: breast cancer detection; deep learning architectures; feature extraction techniques; infrared thermography; machine learning; thermal imaging; transfer learning



Academic Editor: Juan Gabriel Avina-Cervantes

Received: 4 November 2024

Revised: 13 December 2024

Accepted: 20 December 2024

Published: 26 December 2024

Citation: Jalloul, R.; Krishnappa, C.H.; Agughasi, V.I.; Alkhatib, R. Enhancing Early Breast Cancer Detection with Infrared Thermography: A Comparative Evaluation of Deep Learning and Machine Learning Models. *Technologies* **2025**, *13*, 7. <https://doi.org/10.3390/technologies13010007>

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Breast cancer is the most common cancer affecting women worldwide, as under-scored by health authorities such as the Centers for Disease Control and Prevention (CDC) [1]. In the United States, approximately 12.4% of women are expected to be diagnosed with breast cancer at some point in their lives. Similarly, in India, breast cancer cases have surged nearly 50% in recent years, with a rising prevalence across all states. Survival rates for breast cancer patients vary widely, influenced by factors such as tumor type and

stage at diagnosis [2,3]. Fundamentally, breast cancer originates from uncontrolled cell growth within the breast tissue, typically beginning in the lobules or ducts and sometimes spreading to connective tissues. Without timely intervention, these abnormal cells can metastasize, affecting lymph nodes and potentially other areas of the body [4].

Early detection is crucial for managing breast cancer, as it provides a higher likelihood of successful treatment and prevents further spread of the disease [5]. Upon identifying a tumor, healthcare professionals classify it as benign or malignant. While benign tumors are non-cancerous and do not spread, malignant tumors have the potential to metastasize, making it vital to distinguish between the two [6]. However, effective early detection remains challenging due to the lack of rapid, efficient screening tools for identifying cancer in its early stages [7–9]. Addressing this gap could revolutionize the onset of treatment, significantly reducing mortality rates worldwide.

Various breast cancer screening methods exist, each with its own strengths and limitations. Common diagnostic tools include mammography, ultrasound, and magnetic resonance imaging (MRI), all of which have received considerable attention in research. Among non-invasive screening options, thermography stands out due to its affordability, radiation-free nature, and potential effectiveness as a diagnostic tool [10]. Thermography detects the heat emitted from the breast surface, which can be indicative of abnormal metabolic activity linked to cancerous cells. By capturing thermal radiation as digital images, medical professionals can analyze temperature distribution patterns to aid in diagnosis.

The integration of machine learning (ML) and deep learning (DL) techniques has improved the accuracy of breast cancer screening, enhancing the reliability and flexibility of diagnostic models [11]. ML and DL approaches have shown high predictive accuracy, particularly in medical imaging. While breast cancer diagnostics have advanced, further research is needed to refine these approaches and improve detection accuracy using DL algorithms. This research aims to contribute to the field of breast cancer detection by leveraging DL and ML techniques to handle large datasets, enabling models to learn complex patterns and make accurate predictions.

A key DL technique, convolutional neural networks (CNNs), has become widely used for image classification and recognition, especially in supervised learning tasks. Known for their “superhuman” precision in image analysis, CNNs have revolutionized computer vision applications across multiple fields, including healthcare [12]. In recent years, CNNs have evolved rapidly, yielding significant improvements in medical image processing and classification [13].

The primary goal of this study is to develop a reliable and efficient method for breast cancer detection using DL and transfer learning applied to thermal imaging. By providing a comprehensive approach to analyzing breast cancer tumors through thermography, this research offers a non-invasive and cost-effective diagnostic tool. This approach could be particularly beneficial for low-resource and developing regions, where access to traditional screening methods may be limited.

The main contributions of this research include:

- Utilizing thermography images in ML, DL, and transfer learning models to build an effective breast cancer detection system. The focus on thermal imaging offers a non-invasive diagnostic option with the potential to facilitate earlier detection and improve patient outcomes.
- Conducting thorough testing and benchmarking of the developed models against established methods to evaluate their performance and reliability. This comparative analysis highlights the strengths and limitations of various ML and DL techniques in breast cancer detection using thermal imaging, as measured by accuracy, precision, recall, and other performance indicators.

The paper is organized as follows: Section 2 presents an overview of machine learning and deep learning techniques applied to breast cancer classification through thermal images. Section 3 describes the methods and evaluation procedures used in this study. Section 4 discusses the experimental results and findings. Finally, Section 5 provides a discussion and conclusions, including insights into the advantages, challenges, and potential directions for future research.

2. Related Work

Recent research has shown a growing interest in applying machine learning (ML) and deep learning (DL) techniques to thermal imaging for breast cancer diagnosis. Studies have explored various models to enhance accuracy in detecting abnormal thermal patterns associated with breast cancer. For instance, Dabhade et al. [8] used random forest (RF) and support vector machine (SVM) classifiers, achieving accuracy rates of 94.5% and 98.4%, respectively, using cross-validation. This highlights SVM's robustness in distinguishing malignant from benign cases in thermal imaging data. Tiwari et al. [14] extended this work by training a Visual Geometry Group 16 (VGG16) model with a dataset comprising 1345 static and dynamic thermal images. By incorporating multi-view imaging, this study achieved a remarkable test accuracy of 99%, showcasing the potential of VGG16 for high-precision breast cancer classification. Additional research has applied various DL architectures—such as ResNet18, ResNet34, ResNet50, ResNet101, ResNet152, MobileNet, InceptionNet, VGG16, and VGG19—for tasks including segmentation, feature extraction, and classification, often reporting accuracy rates between 90% and 99% [15].

Further studies have investigated other neural network architectures for thermal imaging. The study by Desai and Shah [16] employed a multilayer perceptron (MLP) model to classify breast thermograms into four distinct groups, achieving a 95% accuracy rate, though feature extraction followed a more conventional approach. A similar experiment by Alshehri and AlSaeed [17] was conducted using the University Hospital of The Federal University of Pernambuco dataset, consisting of 1052 thermograms, and evaluated various classifiers, including Naive Bayes, J48 decision trees, SVM, RF, MLP, extreme learning machine (ELM), and random trees (RT). The research demonstrated that the effectiveness of a classifier could vary significantly depending on the architecture and dataset size.

In another study, Wen et al. [18] compared the performances of VGG16, Inception v3, and ResNet50 models on the DMR-IR dataset. VGG16 achieved an impressive 99% accuracy, while VGG19 and ResNet50V2 showed slightly lower testing accuracies of 95%, 94%, and 89%. This suggests that certain models are better suited to breast cancer classification with thermal imaging, particularly when tested on well-prepared datasets. A separate investigation by [19,20] combined ResNet34 and ResNet50 to maximize classification accuracy, highlighting the effectiveness of ensemble deep learning architectures. The study examined the impact of dataset size, data augmentation, and preprocessing on CNN model performance. By comparing MLP and ELM classifiers for early breast cancer detection, this study found that MLP produced an accuracy of 82.20%, whereas ELM reached a perfect accuracy rate of 100%, underscoring the potential of ELM in specific diagnostic applications.

Further advances have been made by Sharafaddini et al. [21], who used an SVM classifier to distinguish between normal and abnormal breast tissues, demonstrating high recall, accuracy, and precision on a benchmark dataset. Additionally, initial optimization trials of CNN models (ResNet101, DenseNet, MobileNetV2, and ShuffleNetV2) for breast cancer diagnosis showed that DenseNet and ResNet101 excelled in classifying static datasets, though both models struggled with dynamic datasets, achieving accuracies of 99.6% and slightly lower for ShuffleNetV2 [22].

Another line of research has introduced new CNN architectures, including Inception-ResNet-v1, Inception-ResNet-v2, and Inception-v4, with favorable outcomes [23]. In the studies by Chowdhury et al. [24], CNNs using Bayesian algorithms were customized to differentiate between suspicious and normal breast images, achieving an accuracy of 98.95% for a dataset of 140 subjects. Other investigations into pre-trained transfer learning models (e.g., VGG16 and InceptionV3) for breast cancer classification revealed that InceptionV3 yielded the best results [25]. Mambou et al. [26] applied a combination of DNN and SVM classifiers, achieving 94% accuracy for a cohort of 67 patients.

In evaluating spectral characteristics for breast cancer screening, a study by Amethiya et al. [27] used Artificial Neural Networks (ANNs) and SVM, reporting sensitivity and specificity rates of 76% and 84% for SVM and 92% and 88% for ANN, respectively. These findings suggest that different architectures offer unique benefits, particularly in terms of specificity and sensitivity. Tables 1 and 2 provide a comparative analysis of various ML and DL models for breast cancer classification, highlighting different approaches, datasets, and accuracy outcomes achieved across studies. The summary underscores the strengths of certain models, like MobileNetV2 and SVM, which consistently achieve high accuracy rates across different datasets. For instance, MobileNetV2 achieved an impressive 99.6% accuracy on the DMR/IR dataset, showcasing its effectiveness in thermal imaging-based classification. Similarly, SVM-based models have demonstrated reliable accuracy across multiple studies, with Aarthy's research achieving 97.6% on a set of real breast images [28].

Table 1. Summary of ML Models on the DMR Dataset.

Ref.	Approach	Dataset	Imaging Type	Feature Extraction	Results (Accuracy)
Silva et al. [29]	SVM	Not specified	Not specified	Standard SVM features	90%
Aarthy et al. [30]	SVM	83 images (34 normal, 49 abnormal)	Static	Custom feature engineering	97.6%
Allugunti et al. [31]	SVM, Random Forest	DMR-IR	Static	Standard SVM and RF	SVM: 89.84%, RF: 90.55%
Karthiga et al. [32]	SVM	DMR-IR	Dynamic	Custom feature extraction	93.3%
Bancos et al. [33]	ANN, Decision Tree, Bayesian	DMR-IR	Dynamic	Standard ANN, DT, Bayesian	ANN: 73.38%, DT: 78%, Bayesian: 88%
Nissar et al. [34]	SVM	DMR-IR	Static	Standard SVM features	Not reported

Table 2. Summary of DL Models on the DMR Dataset.

Ref.	Approach	Dataset	Imaging Type	Feature Extraction	Results (Accuracy)
Bhowmik et al. [35]	Multilayer Perceptron (MLP)	DMR	Static	Traditional MLP	95%
Torres et al. [36]	ResNet101, MobileNetV2, DenseNet201	DMR/IR	Static	CNN-based	MobileNetV2: 99.6%
D'Alessandro et al. [37]	SVM and DNN	67 patients (43 healthy, 24 sick)	Static	CNN feature extraction	94%
Zuluaga-Gomez et al. [38]	CNNs	DMR/IR	Dynamic	CNN-based	92%
Abdulkutty et al. [39]	VGG16, InceptionV3	DMR-IR	Static	Transfer Learning	VGG16: 87.3%
Agughasi et al. [40]	ResNet50, InceptionV3	57 Thermal Images	Static	CNN-based	ResNet50: 92%, InceptionV3: 90%

3. Methods

This study adopts a robust experimental methodology to accomplish the stated objectives, with a structured model divided into two primary components: deep learning for feature extraction and machine learning for classification, as highlighted in Figure 1. By separating feature extraction from classification, this approach provides flexibility in optimizing each stage independently, maximizing model performance, and ensuring effective classification. The integration of thermal imaging with advanced machine learning (ML) and deep learning (DL) techniques allows for a non-invasive diagnostic alternative for breast cancer, especially suited for younger women, where traditional mammography may be less effective due to dense breast tissue.

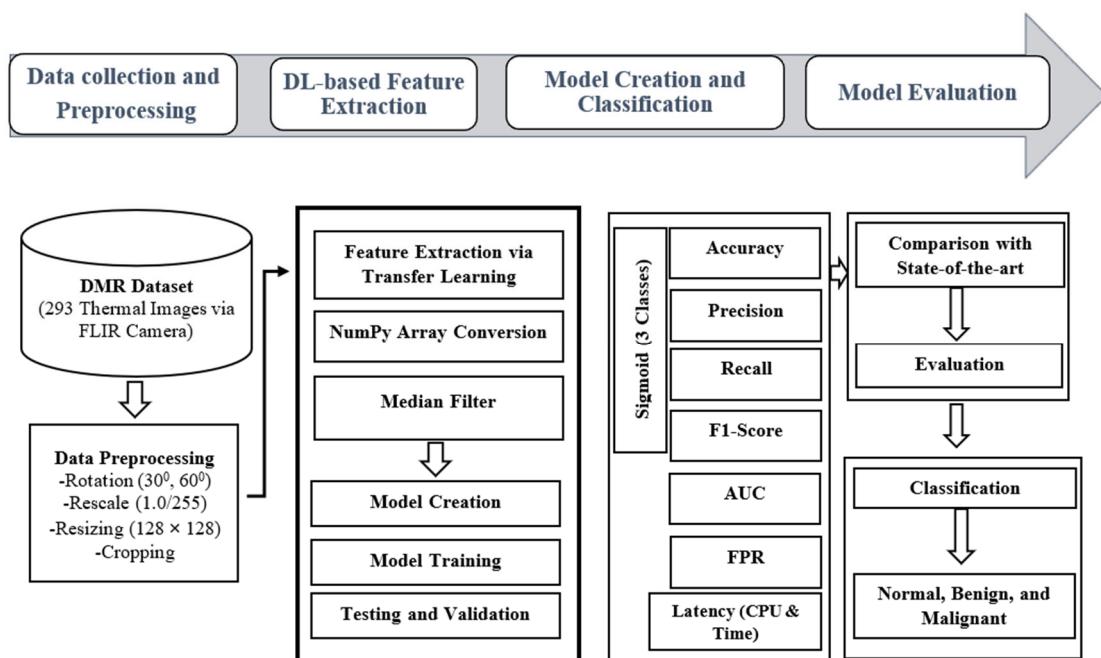


Figure 1. Proposed Framework for Feature Extraction and Classification from Thermal Breast Images.

In this section, comparisons are made between various transfer learning models in the feature extraction phase, as well as between DL models and traditional classification methods in distinguishing malignant from benign tissue using thermal images. The evaluation criteria include accuracy, precision, recall, F1-score, specificity, area under the curve (AUC), false positive rate, latency (processing time), and CPU usage, providing a comprehensive assessment of each model's suitability for thermal image-based breast cancer classification.

The methodology includes several subsections covering thermal imaging acquisition, data preparation, feature extraction, classification, and model evaluation.

3.1. Thermal Images

Breast cancer cannot be entirely prevented; however, early and accurate detection remains the most effective approach to reducing mortality and improving patient survival rates. Despite advancements in screening, mammography—the widely accepted modality—is associated with limitations, particularly among younger women. Mammographic images are often compromised by low contrast due to dense breast tissue, which can obscure malignancies and lead to delayed diagnosis. Although it offers the advantage of non-invasive imaging, its drawbacks include exposure to radiation, false positives, false negatives, and the potential development of interval cancers—cancers that emerge between routine screenings [41]. In the United States, standard mammographic screenings are conducted either

annually or biennially depending on age and risk factors, while countries like the United Kingdom have a three-year interval, which may increase the incidence of interval cancers.

Given these limitations, thermal imaging emerges as a promising alternative. As a non-invasive and non-contact imaging technique, thermography minimizes risks and allows for repeated screenings without health repercussions [42]. From its successful induction in 1956, thermography has been used as a cancer screening tool. The technique leverages infrared cameras to capture temperature distribution across the breast surface, with abnormal thermal patterns potentially indicating malignancies. Thermography's lack of radiation exposure and ability to provide real-time results make it an ideal choice for early detection, especially among populations sensitive to radiation or requiring frequent screenings. The datasets, as detailed in Table 3, illustrate the diversity in imaging types: static imaging (SIT) and dynamic imaging (DIT) [43], sources, and camera technologies used, reflecting the breadth and challenges in thermal imaging-based breast cancer detection.

Table 3. Types of Thermal Imaging Datasets.

Ref.	Source	SIT	DIT	No. of Images	Clinical Validation	Public Access	Camera Used
Bezerra et al. [44]	Clinical Hospital of the federal University of Pernambuco (HC/UFPE), Brazil	Yes	No	336 (120 benign, 74 cysts, 76 malignant, 66 without lesion)	No	Yes	Not specified
Gogoi et al. [45]	Agartala Government Medical College (AGMC) of Govind Ballav Pant (GBP) Hospital, Agartala	Yes	No	49 abnormal, 45 normal, 6 unknown	Yes	No	FLIR-T650sc
Resmini et al. [46]	University Hospital Antônio Pedro (HUAP) of Federal Fluminense University, Brazil	Yes	Yes	311 (267 healthy, 44 sick)	Yes	Yes	FLIR-SC620

However, a key challenge in utilizing thermal imaging for breast cancer detection lies in the limited availability and size of thermal imaging datasets. Unlike mammography, which benefits from large public databases, thermography datasets tend to be small and are often housed in private repositories, limiting their accessibility for research [37]. Smaller datasets can lead to model overfitting, where the model performs well on training data but fails to generalize to new, unseen samples. To counteract this, data augmentation techniques are applied to expand the dataset artificially, thereby enhancing the model's robustness and reducing overfitting.

3.2. Data Preparation

The FLIR SC620 thermal camera, capturing images at a resolution of 640×480 pixels and a focal length of $45 \mu\text{m}$, was used. Both static and dynamic imaging protocols were employed to provide a comprehensive dataset. Figure 2 displays various sample views from the dataset. To maintain the accuracy and reliability of the images, a controlled environment was established for the imaging process. The imaging room temperature was kept within $65\text{--}75^\circ\text{F}$, with no luminescent lighting and an ambient temperature stability of $\pm 1^\circ\text{C}$ to prevent artifacts in the images.

Figure 3 presents a thermal image of the breast, capturing the body from the waist to the neckline. The image displays a spectrum of colors, predominantly red, green, and yellow, distributed across the entire body. The raw dataset utilized consists of 3895 thermal breast images in JPEG format. To ensure accurate data acquisition during the thermographic procedure, it is crucial to maintain a controlled environment while using infrared

imaging so as to account for human physiological factors and environmental variations. Specific protocols and guidelines were in place for patients to follow before entering the imaging room, as well as strict conditions under which images must be captured in the laboratory. These measures are essential to prevent artifacts in the final images. The laboratory environment was maintained at a moderately cool temperature, ranging between 65–75 degrees Fahrenheit, with no luminescent lighting. The lab was free from extreme temperatures that could affect the patient, and the ambient temperature remained within 1 °C throughout the imaging process [47].

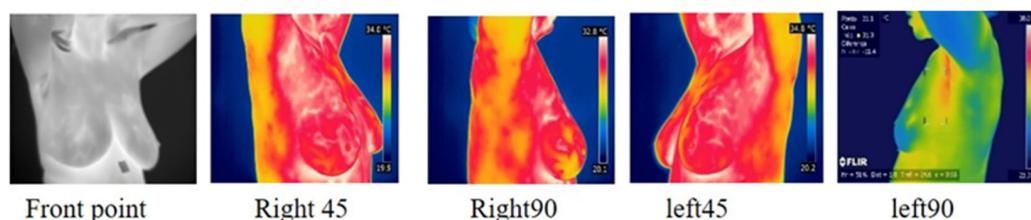


Figure 2. Sample of Thermal Images from Dataset.

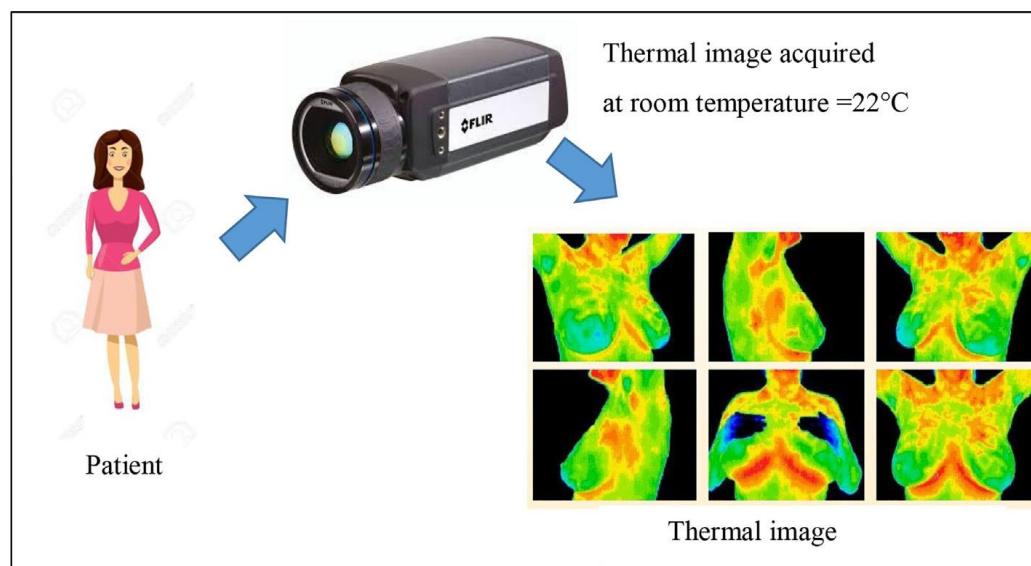


Figure 3. Example of a Full-Body Thermal Image Capturing the Breast Area.

3.2.1. Patient Guideline

The images represent thermal variations across different views: front, right at 45°, right at 90°, left at 45°, and left at 90°. The color gradients indicate temperature distributions, where cooler temperatures are represented by blue and green hues, transitioning to warmer temperatures shown in yellow, orange, and red. These thermal differences are analyzed for identifying abnormal patterns or asymmetries that may signify underlying conditions. Each thermal image includes a temperature scale bar for precise interpretation of heat zones. To ensure the accuracy and reliability of thermal imaging, both patients and technicians must adhere to a set of preparation protocols before entering the imaging laboratory. These guidelines are essential for minimizing external factors that may influence body temperature distribution and, consequently, the imaging results:

- Avoid direct sun exposure prior to the imaging session.
- Refrain from any breast stimulation or treatments involving the breast area.
- Do not apply lotions, deodorants, antiperspirants, or makeup on the day of imaging.
- Avoid physical activities or exercises that may increase body temperature.
- Refrain from bathing or showering immediately before the imaging procedure.

- Remove clothing for approximately 12 min prior to imaging to allow the body to acclimate to the room temperature [40].
- Following these protocols helps ensure that the thermal images reflect true physiological conditions, enhancing the reliability of the diagnostic process.

3.2.2. Image Preprocessing

The preprocessing of infrared images is a crucial phase, as it enhances image quality and ensures suitability for analysis. This process involves identifying the image's orientation and reference points while removing any unwanted elements that could distort the final outcome. This is essential for eliminating the noise commonly associated with thermal images, and it includes the following steps:

- Identifying and addressing incomplete or erroneous entries. Entries that could not be corrected were removed from the dataset.
- Excluding patients who lacked all five standard imaging angles: front, left 45°, right 45°, left 90°, and right 90°.
- Substituting dynamic protocol images for any missing or unclear static images in the front or side views.

Several preprocessing techniques were employed, including cropping, normalization, and resizing. This stage also involved eliminating various image artifacts through filtering. Wiener, median, adaptive, and average filters were applied to improve image clarity. Typically, the refined image undergoes an average filtering process, followed by a quality assessment step, resulting in a final visual image that is suitable for analysis, as highlighted in Figure 4. This filtering process adjusts the mean value and enhances surrounding areas by modifying pixel values, thereby reducing noise.

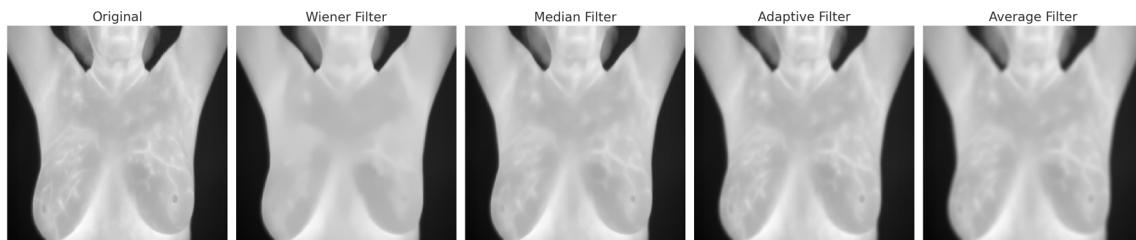


Figure 4. Effects of the preprocessing filters applied to infrared images.

Images were resized from their original resolution of 640×480 pixels to 224×224 pixels [41], a dimension compatible with deep learning models' default input size. This reduction in dimensions enhances computational efficiency. Additionally, training with lower-resolution images provides the model with generalized data, improving performance during training and increasing the model's ability to generalize when tested on new data. To prevent overfitting, data augmentation was applied to expand the dataset, producing a diverse set of images for robust model training.

Figure 4 shows the original image alongside images processed with different filters: Wiener filter, Median filter, Adaptive (Gaussian) filter, and Average filter. Each filter was used to enhance image quality by reducing noise and artifacts, aiding in better feature extraction during subsequent analysis.

3.3. Augmentation

To enhance model accuracy and prevent overfitting, both static and dynamic thermal image datasets were augmented using the `ImageDataGenerator` function in Keras. This process expands the dataset by introducing variations such as rotation, shearing, rescaling,

and zooming, as detailed in previous studies [43,44]. Optimal augmentation parameters were identified (5° rotation, 0.02 shear, 0.02 zoom, and a rescaling factor of 1/255) to introduce sufficient variability without compromising image integrity, thereby improving the model's ability to generalize [45].

To validate that the augmented data closely represents real-world thermal images, we compared their statistical properties and distributions using several methods. Firstly, pixel intensity distributions were analyzed for both the real-world and augmented datasets. Metrics such as mean, standard deviation, skewness, and kurtosis of pixel intensities were computed, and histograms were plotted for visual comparison (see Figure 5). This analysis ensured that the augmented data shared similar statistical properties with the real-world dataset.

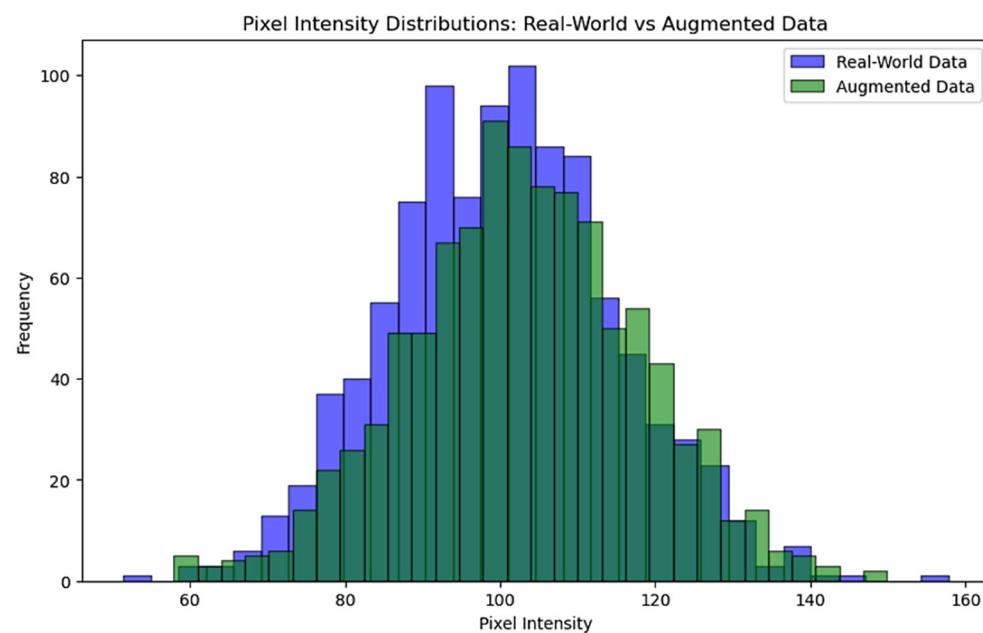


Figure 5. Distribution of pixel intensities on the real-world vs. augmented data.

Secondly, statistical tests such as the Kolmogorov–Smirnov (K-S) test were conducted to evaluate the similarity of distributions between the augmented and real-world datasets. A non-significant p -value would indicate a high similarity between the datasets, supporting the validity of the augmentation techniques used. Additionally, texture analysis using Haralick features (e.g., entropy, contrast, and energy) was performed to compare spatial structures, and Structural Similarity Index (SSIM) scores were calculated to quantify the visual similarity between augmented and real-world images, and the results are presented in Table 4. These rigorous analyses confirmed that the augmentation effectively retained critical thermal image characteristics.

Table 4. Comparison of Real and Augmented Data Properties.

Metric	Real Data	Augmented Data	Insight
Kolmogorov–Smirnov Test	KS Statistic = 0.09	p -Value = 0.0006	Indicates a statistically significant difference between the real and augmented pixel intensity distributions, suggesting augmentation slightly altered data.
Entropy	7.997	7.979	Entropy values are very close, implying similar texture complexity between the real and augmented images.

Table 4. Cont.

Metric	Real Data	Augmented Data	Insight
Haralick Features	Various (see above)	Various (see above)	Most Haralick texture metrics, such as energy and contrast, are very similar, demonstrating minimal changes in texture properties after augmentation.
SSIM	-	0.9969	The high SSIM value (close to 1) indicates the augmented images retain structural similarity to the real-world data.
Max Temperature (°F)	157.79	149.90	Both maximum temperature values are outside the typical clinical range (85–110 °F), indicating potential data generation issues or unrealistic augmentation.
Min Temperature (°F)	51.38	57.89	Minimum temperature values are also outside the expected range, highlighting a need to validate or constrain augmentation methods.
Temperature Validity	Outside realistic range	Outside realistic range	The temperature range in both datasets fails to meet clinical expectations, signaling a need for stricter data preprocessing or augmentation constraints.

3.4. Feature Extraction

Feature extraction is crucial in breast cancer detection, allowing the model to focus on meaningful traits within thermal images. Convolutional Neural Networks (CNNs) serve as a default choice for feature extraction, directly transforming input thermal images into feature representations [46,47]. This study employed transfer learning with six pre-trained CNN models: VGG16, InceptionV3, ResNet152, MobileNetV2, DenseNet121, and Xception. Each model was used without its top classification layers, with average pooling applied to obtain refined feature representations:

- VGG16: Known for its depth and simplicity, it was used to extract high-level features, which is ideal for capturing intricate patterns within breast thermal images.
- InceptionV3: Utilizes inception modules for capturing multi-scale features, particularly valuable for distinguishing subtle temperature variations in thermal images.
- ResNet152: Leverages residual connections to handle deeper networks, enabling it to effectively capture complex visual patterns in breast cancer thermograms.
- MobileNetV2: Lightweight and efficient, MobileNetV2 is optimized for feature extraction in resource-constrained environments.
- DenseNet121: Uses dense connections to increase feature reuse and reduce redundancy, enhancing the model's effectiveness in analyzing high-resolution thermal images.
- Xception: An extension of the Inception architecture employs depth-wise separable convolutions to efficiently capture essential features.

Each model's feature extraction capabilities are leveraged to create a feature set that serves as the input for subsequent classification.

3.5. Dimensionality Reduction

Dimensionality reduction is essential for optimizing breast cancer detection models, especially when working with complex, high-dimensional data such as infrared thermography images. In machine learning and deep learning contexts, dimensionality reduction methods condense the feature set by retaining only the most significant variables, thereby preserving key information while minimizing data loss. This approach helps to overcome

challenges associated with high-dimensional datasets, often termed the “curse of dimensionality”, which can result in issues like overfitting, excessive computational requirements, and reduced interpretability [48]. By isolating and retaining relevant patterns while discarding redundant or irrelevant features, dimensionality reduction enhances model accuracy, computational efficiency, and generalizability to new data.

In this study, dimensionality reduction was applied to thermographic data to simplify the feature space while maintaining essential information for accurately distinguishing between healthy, benign, and malignant breast tissue. Principal Component Analysis (PCA) was the primary technique used. As an unsupervised method, PCA transforms data into a series of orthogonal components, ranked according to the amount of variance each captures. This technique effectively identifies critical temperature variations associated with different breast tissue types, facilitating improved classification. By reducing the feature space, dimensionality reduction not only increased computational efficiency but also accelerated model training and inference, a benefit crucial for real-time application scenarios.

3.6. Cross-Validation and Model Evaluation

Cross-validation is a crucial step in assessing the robustness and generalizability of machine learning models. In this study, a k-fold cross-validation approach ($k = 10$) was implemented to ensure that the reported performance metrics, such as accuracy and AUC, are not artifacts of specific data partitions, as shown in Figure 6. This method divides the dataset into ten equal parts, using nine folds for training and one-fold for validation in each iteration. The process is repeated for all folds, and the average performance across these iterations provides a comprehensive evaluation of the model’s reliability.

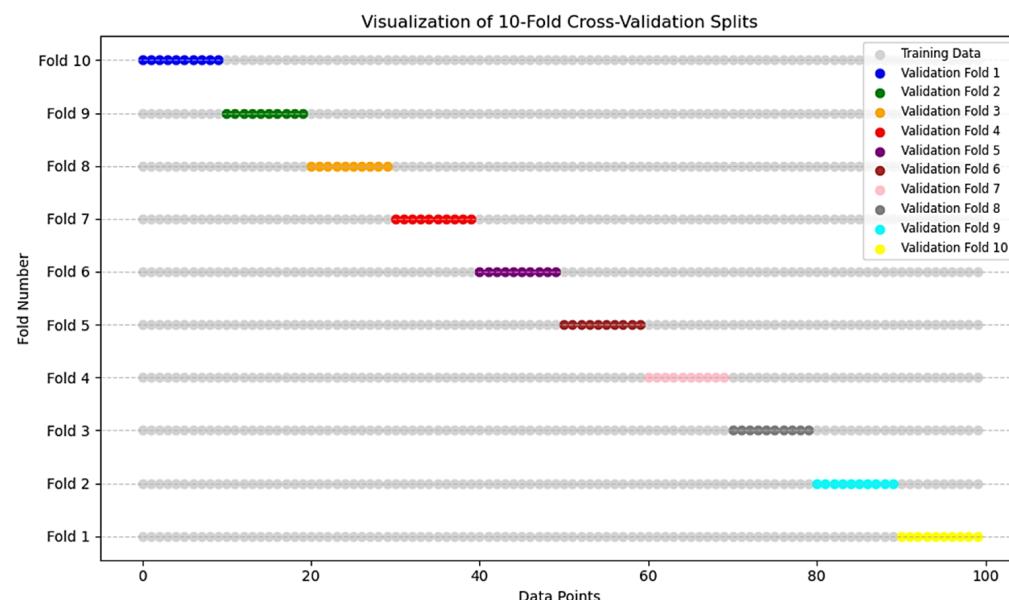


Figure 6. Workflow of 10-fold cross-validation implementation.

While this approach minimizes the risk of overfitting and ensures the robustness of the given dataset, it is acknowledged that the relatively small size of the dataset may limit the generalizability of the results. Larger and more diverse datasets are needed to further validate the model. Future work will include cross-validation on expanded datasets, which will help confirm the reliability of the performance metrics and ensure that the model performs consistently across diverse data distributions. Additionally, stratified k-fold cross-validation will be explored to maintain the proportion of class labels within each fold, further enhancing the evaluation process.

This rigorous cross-validation approach highlights the study's commitment to reliable model evaluation, mitigating concerns about performance variability due to data partitioning.

3.7. Classifier

Following feature extraction from pre-trained models, the next phase involves classifying these features to differentiate between healthy, benign, and malignant categories. These classifiers analyze the intricate patterns in thermal images to accurately identify and classify potentially malignant regions. By leveraging advanced ML techniques, this approach aims to improve the accuracy of breast cancer diagnosis [49].

Once feature extraction and image preprocessing are complete, the data are prepared for model training. For this study, the dataset was split into training and testing sets in an 80/20 ratio that was experimentally decided to balance model training effectiveness and evaluation reliability. Specifically, 80% of the data was allocated for model training, and the remaining 20% was reserved for testing. The validation subset is utilized during model training to fine-tune parameters, while the final test set provides an unbiased evaluation of the model's performance on previously unseen data.

To evaluate the performance, several traditional ML classifiers, each chosen for its unique advantages in thermal image-based breast cancer diagnosis, were employed:

Support Vector Machines (SVM): This study employs both linear and kernel-based SVM variants to classify features extracted from thermal images into healthy and malignant categories [50]. The SVM classifier is trained on extracted features, and its performance is evaluated using metrics such as accuracy, precision, recall, and area under the curve (AUC).

Naive Bayes: Known for its simplicity, Naive Bayes is a probabilistic classifier particularly suited for datasets with limited size or when model interpretability is essential. By analyzing the extracted features from thermal images, it distinguishes between normal, benign, and malignant cases with a probabilistic approach.

Decision Trees: Due to their interpretability and ability to handle various feature types, decision trees are well-suited for classifying breast cancer in thermal images; overfitting and instability were mitigated through ensemble methods, thus enhancing performance.

K-Nearest Neighbors (KNNs): This operates as an instance-based classifier, categorizing new data points based on the proximity of their features to those of their closest neighbors. This makes KNN a practical classifier for breast cancer diagnosis through thermal imaging, as it uses similarity metrics to assess malignancy likelihood.

Deep Neural Networks (DNNs): DNNs are useful for modeling complex, non-linear relationships, making them especially effective for distinguishing between healthy and cancerous tissues in thermal images. Due to their capacity for automatic feature learning from raw data, DNNs are well suited for this application. They were meticulously trained to optimize computational resources.

Each classifier's success is measured by several performance metrics, primarily accuracy, which indicates the proportion of correctly classified samples in distinguishing malignant from non-cancerous regions in thermal images. Additional performance indicators—such as precision, recall, F1-score, and AUC—offer further insight into each classifier's diagnostic effectiveness. These metrics provide a comprehensive evaluation, allowing for a nuanced comparison of each model's ability to accurately diagnose breast cancer through thermal imaging.

3.8. Evaluation Metrics

Evaluation metrics provide insight into each model's reliability and accuracy in medical diagnostics, particularly for distinguishing between healthy, normal, and malignant breast tissue.

3.8.1. Confusion Matrix

The confusion matrix provides an in-depth view of model performance by displaying true positives, true negatives, false positives, and false negatives. This matrix enables the calculation of crucial metrics like accuracy, precision, recall, and F1-score, which are essential for assessing classification effectiveness.

3.8.2. Performance Metrics

Performance metrics are essential for evaluating the precision and dependability of the system in determining the presence or absence of disease in medical diagnostic models. These metrics offer a thorough knowledge of the model's performance in differentiating between healthy and unhealthy individuals. They include true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Because each indicator captures a distinct facet of the model's performance, researchers are able to evaluate the model's advantages and disadvantages while making therapeutic decisions. The following are the key performance metrics that are integral to evaluating the diagnostic [51] model's ability to accurately classify patient outcomes as depicted in Equations (1–4):

True positive (TP) is the number of sick patients correctly identified as sick.

True negatives (TN) is the number of healthy patients correctly identified as healthy.

False positives (FP) is the number of healthy patients incorrectly identified as sick.

False negatives (FN) is the number of sick patients incorrectly identified as healthy.

- Accuracy: The overall correctness of the model is calculated as the ratio of correctly predicted instances to the total instances.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN} \quad (1)$$

- Recall (Sensitivity or True Positive Rate): The ratio of correctly predicted positive observations to all observations in the actual class.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

- Specificity (True Negative Rate): The ratio of correctly predicted negative observations to all observations in the actual negative class.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

- F1 Score: The harmonic mean of precision and recall, providing a balance between the two.

$$\text{F1 Score} = 2 \times \frac{\text{PRECISION} \times \text{RECALL}}{\text{PRECISION} + \text{RECALL}} \quad (4)$$

3.8.3. Latency (Time and CPU)

Latency refers to the time required for a neural network to generate a prediction for a single input sample and is a key metric for assessing model performance, particularly in practical deployment scenarios [52]. Utilizing Python's "time" module to measure the duration of a forward pass, along with the Pustil module for CPU usage analysis, PyTorch

enables a comprehensive evaluation of latency for neural networks. This metric is crucial when comparing different models, as low latency is essential for applications demanding real-time or near-real-time responses, where rapid predictions hold as much importance as high accuracy. Achieving a balance between latency and accuracy is often necessary, especially since some models may exhibit high performance but at the cost of slower processing times. In this study, latency encompasses not only the time taken for individual predictions but also the entire training phase, including forward and backward passes during each epoch and optimization updates. Lower latency during training enhances model efficiency, making it more suitable for large datasets or complex models and thus more applicable to real-world scenarios that prioritize both speed and precision.

4. Experiments and Results

4.1. Dataset

This study utilizes two datasets to support research in early breast cancer detection through thermography imaging.

The first dataset, as shown in Table 1, is sourced from the Database for Mastology Research (DMR), an online repository that stores and manages mastologic images. It includes data collected from the University Hospital Antonio Pedro (HUAP) at the Federal Fluminense University in Brazil. The DMR database utilizes both Static Imaging Technique (SIT) and Dynamic Imaging Technique (DIT) protocols [53]. SIT captures stable thermal images, while DIT assesses dynamic temperature responses in the breast area using a cooling fan. The database includes information from 293 patients, with thermal images captured from five angles of the breast: front, left 45°, right 45°, left 90°, and right 90°. Each entry also contains clinical and personal information such as age, dietary habits, medical history, and menstrual details, providing a comprehensive dataset for analysis.

The second dataset, obtained from Mendeley Data, is titled "Breast Thermography" and was published on 5 February 2024 (Version 3, DOI: 10.17632/mhrt4svjxc.3) [54]. This dataset provides thermographic images focused on the female thorax area as part of breast cancer research. The images were acquired in a medical office with dimensions of 3.20 m (W) × 4.14 m (L) × 2.40 m (H), under controlled conditions without artificial lighting, with temperatures between 22 and 24 °C and relative humidity of 45–50%. A FLIR A300 camera (FLIR Systems, Wilsonville, OR, USA) was used for capturing the thermal images positioned 1 m from the patients, who were prepared and imaged according to the American Academy of Thermology (AAT) protocol. Images of each patient were taken from three positions: anterior, left oblique, and right oblique. This dataset includes a total of 357 thermographic images (consisting of 105 images of malignant and 252 images of benign cases), covering patients aged 18 to 81 with various breast pathologies, captured between 2021 and 2022.

4.2. Experimental Setup

The experiments were conducted using Python 3.12 in a Google Colab environment, chosen for its computational efficiency and suitability for deep learning tasks. The hardware specifications for this experiment included the following:

- Processor: Intel Core i7-11800H (8 cores, 16 threads)
- RAM: 32 GB DDR4
- Graphics Card: NVIDIA GeForce RTX 3060 (6 GB VRAM)
- Storage: 1 TB SSD

These specifications enabled effective handling of deep learning models and datasets, with GPU acceleration leveraged through CUDA in Google Colab for faster processing times.

4.3. Effects of Feature Extraction

This sub-section presents a comprehensive analysis that follows best practices in feature extraction and visualization, drawing inspiration from methodologies employed in advanced geospatial data applications. The findings demonstrate the efficacy of utilizing thermal imaging features in combination with Principal Component Analysis (PCA) for the accurate classification of breast cancer cases.

4.3.1. Data Cleaning and Feature Analysis

Upon initial examination of the dataset, it was observed that there were missing values across 2048 features. To ensure the integrity of the data, any rows containing NaN values were removed, resulting in a cleaned dataset containing 1299 rows and the full 2048 features.

4.3.2. Dimensionality Reduction (PCA)

To simplify the feature dimensions, Principal Component Analysis (PCA) was implemented, reducing the original 2048 features to 25 components. This approach aimed to capture the most significant sources of variance within the data. The first two principal components, representing the most substantial contributors to variance, include weighted combinations of features such as global intensity distributions, spatial gradients, and thermal texture characteristics as depicted in Figure 7. These components highlight key patterns that separate benign and malignant cases effectively. The subsequent scatter plot (see Figure 8) illustrates the distribution of these cases along the first two principal components, demonstrating the separation between classes.

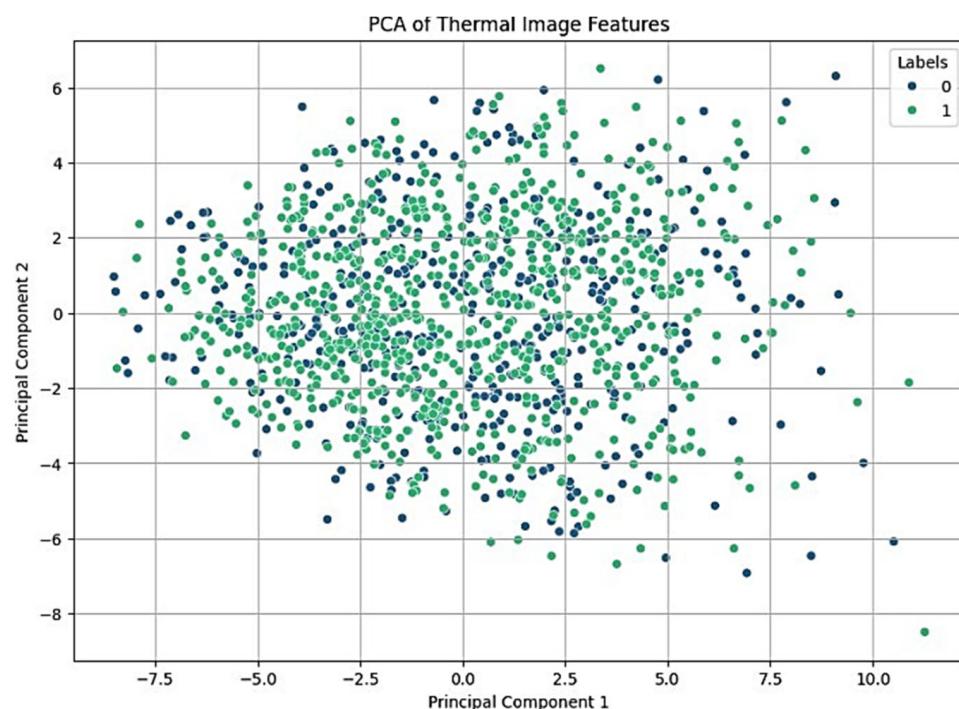


Figure 7. PCA Visualization of Thermal Image Features for Breast Cancer Detection.

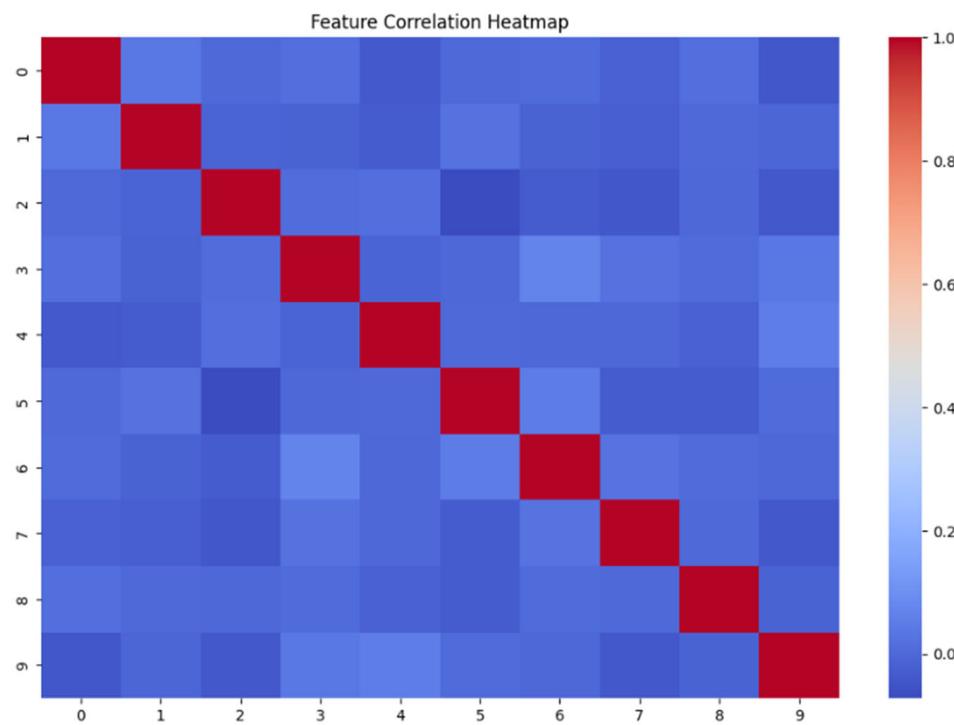


Figure 8. Feature Correlation Heatmap for Thermal Image Dataset.

4.3.3. Feature Heatmap

- A heatmap was created to visualize the correlations between the top 10 features extracted from the thermal image dataset. These features include descriptors such as mean temperature, temperature variance, maximum intensity, edge gradient, and entropy. The color gradient in the heatmap indicates the strength of the correlations, with darker colors representing weaker correlations and lighter colors indicating stronger relationships.
- This visualization helps identify patterns and relationships among features, revealing which features are strongly correlated and which are independent.
- The absence of strong correlations among most features suggests minimal redundancy, meaning each feature provides unique information to the model.
- However, clusters of similar color patterns among certain features may suggest localized correlations, which could be further explored for dimensionality reduction or feature engineering strategies.
- The heatmap also offers insights into outliers or unique patterns across samples, supporting the development of more robust preprocessing methods.
- Additionally, similar color patterns among certain features may suggest correlations, which could be beneficial for dimensionality reduction techniques like PCA.

4.3.4. Feature Importance

- Feature importance was assessed using a RandomForestClassifier to rank the contributions of various features to the model's predictive accuracy. The top-ranked features include mean temperature, maximum temperature, edge intensity, gradient magnitude, and texture complexity, which are crucial in distinguishing between benign and malignant cases.
- Features with lower importance scores, such as skewness or kurtosis, were found to contribute minimally and may be considered for removal in future iterations to streamline the model.

- The bar plot in Figure 9 highlights the top 10 most important features, with importance scores ranging from approximately 0.09 to 0.43. The highest-ranked feature, with a score of 0.4297, indicates a strong influence on the model's predictive decisions.
- These results emphasize the importance of targeted feature engineering and data collection to further improve model accuracy. Additionally, identifying critical features allows for a focused approach in future studies, particularly in refining diagnostic tools.

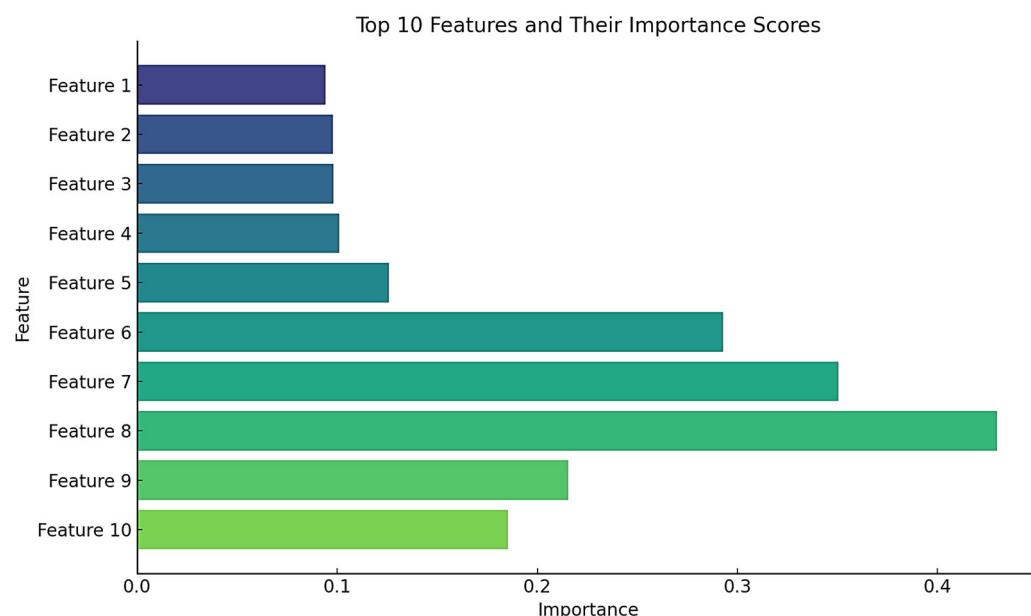


Figure 9. Top 10 Most Important Features from the Thermal Images for Breast Cancer Detection.

4.4. Evaluation Metrics

To evaluate the performance of various machine learning classifiers applied to the thermal images, this study utilized standard metrics, including accuracy, precision, recall, F1-score, and ROC-AUC scores [55]. Each metric offers unique insights into the classifier's performance, with ROC-AUC scores specifically helping to assess the model's capacity to distinguish between malignant and non-malignant cases.

4.4.1. Confusion Matrix of the Best Classifier

The Support Vector Machine (SVM) classifier, utilizing features extracted from the ResNet-152 deep learning model, achieved strong performance for breast cancer classification on thermal images. The model's evaluation is detailed in a confusion matrix, as shown in Figure 9. The results from the confusion matrix for the SVM classifier with ResNet-152 feature extraction are as follows:

- True positives (TP) for healthy: 26—cases correctly identified as healthy.
- True positives (TP) for benign: 37—cases correctly identified as benign.
- True positives (TP) for malignant: 8—cases correctly identified as malignant.
- False positives (FP) for benign: 0—no healthy cases were misclassified as benign.
- False negatives (FN) for benign: 7—benign cases misclassified as malignant.
- False positives (FP) for malignant: 9—benign cases incorrectly classified as malignant.

The classifier demonstrated high accuracy in distinguishing between different classes, specifically showing effectiveness in accurately identifying benign cases while keeping misclassifications low.

Best Classifier: SVM using ResNet-152 features, with an overall accuracy of 0.92 and a latency of 0.0010 s, making it a strong candidate for reliable and efficient breast cancer classification.

While the confusion matrix provides basic insights into misclassification, it is essential to highlight the clinical implications of these errors. False positives, such as the nine benign cases misclassified as malignant, can cause unnecessary anxiety and emotional distress and lead to additional tests or procedures, increasing healthcare costs. On the other hand, false negatives, such as the seven benign cases misclassified as malignant, pose a more critical risk, particularly in cases of malignant tissue misdiagnosis. Delayed diagnosis and treatment due to false negatives can result in disease progression and adverse patient outcomes. Minimizing these errors is vital for improving the reliability of breast cancer detection models. Future work will focus on further fine-tuning the classifier to achieve a better balance between false positives and false negatives, ensuring its practical applicability and enhancing patient care. The SVM classifier with ResNet-152 features achieved an overall accuracy of 0.92 and a latency of 0.0010 s, making it a strong candidate for reliable and efficient breast cancer classification.

4.4.2. Weighted Accuracy and Utility-Adjusted F1-Score

Weighted Accuracy and Utility-Adjusted F1-Score were incorporated to evaluate the clinical relevance of the model's performance. These metrics account for the utility or cost of misclassifications, providing a comprehensive assessment beyond standard accuracy and AUC. This approach ensures that the model aligns with real-world clinical priorities, particularly for tasks such as breast cancer detection.

i. Weighted Accuracy:

The formula for weighted accuracy is defined as:

$$\text{Weighted Accuracy} = \frac{\text{Utility}_{TP} \times TP + \text{Utility}_{TN} \times TN - \text{Utility}_{FP} \times FP - \text{Utility}_{FN} \times FN}{\text{Total Samples}}$$

where:

TP: true positives

TN: true negatives

FP: false positives

FN: false negatives

Utility TP, Utility TN, Utility FP, Utility FN = Utility weights of each classification type
Using the confusion matrix derived from the ResNet152 + SVM model (Figure 10),

TP = 26 + 37 + 8 = 71; FP = 0 + 9 = 9; FN = 7; TN = 0

Total Samples = TP + FP + FN + TN = 87; and

Utility TP = Utility TN = 1 and Utility FP = Utility FN = −1

The weighted accuracy is:

$$\text{Weighted Accuracy} = \frac{1.71 \times 1.0 - (-1) \times 9 - (-1) \times 7}{87} = 1.0$$

ii. Utility-Adjusted F1—Score:

The Utility-Adjusted F1-Score balances precision and recall while considering their utility:

$$\begin{aligned} \text{Utility-Adjusted F1-Score} \\ = \frac{2 \times \text{Utility-Precision} \times \text{Utility-Recall}}{\text{Utility-Precision} + \text{Utility-Recall}} \end{aligned}$$

where:

$$\text{Utility-Precision} = \frac{TP}{TP + FP} \text{ and Utility-Recall} = \frac{TP}{TP + FN}$$

Thus, Utility-Adjusted F1-Score = 0.899

These metrics demonstrate the model's ability to minimize the clinical cost of misclassifications while maintaining high accuracy and robustness. By incorporating these utility-weighted evaluations, we ensure that the model's performance aligns with practical clinical applications and improves real-world relevance.

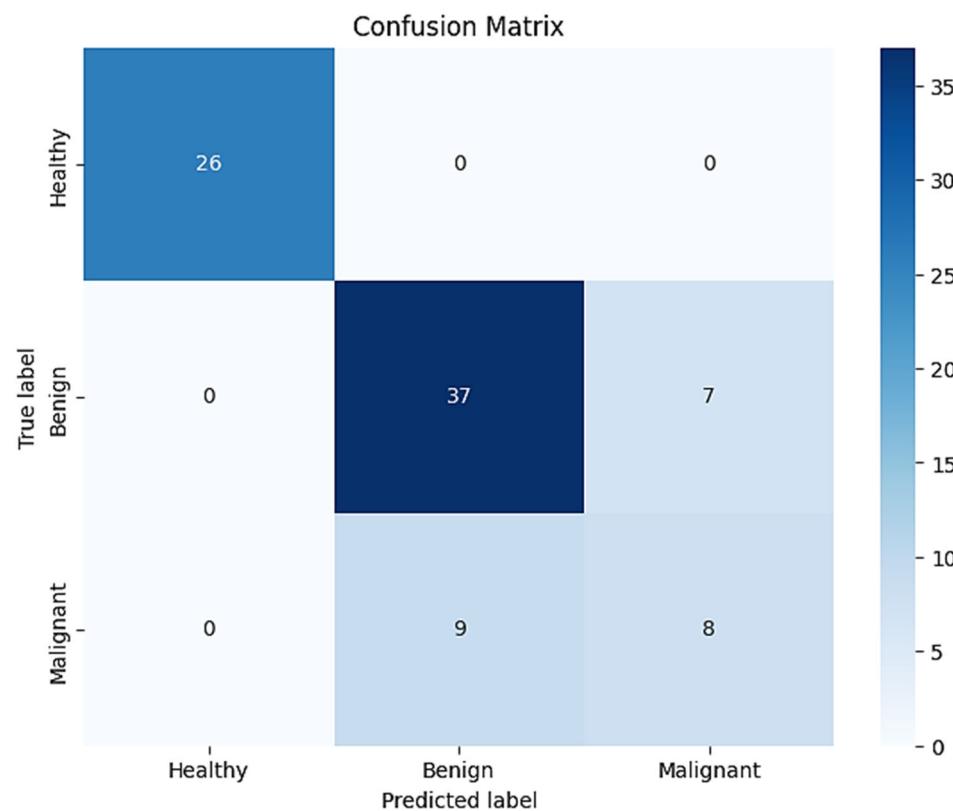


Figure 10. Confusion Matrix for SVM with ResNet-152 Features.

4.4.3. Precision–Recall Curve

The Precision–Recall (PR) curve in Figure 10 illustrates the balance between precision and recall at various threshold settings, offering insights into the model's performance in breast cancer classification using thermal images. This curve is especially useful in contexts involving imbalanced datasets, as it highlights the trade-off between the ability to identify positive cases (recall) and the accuracy of those positive predictions (precision).

The model begins with a high precision of 1.0 at lower recall values, indicating that, initially, the model makes highly accurate predictions but may miss some actual positive cases. As recall increases, precision gradually decreases, demonstrating the inherent trade-off between these two metrics. Despite this trade-off, the PR curve remains consistently high, suggesting that the model effectively balances recall and precision across a wide range of thresholds. This strong performance across varying criteria confirms the model's robustness in accurately classifying breast cancer cases, even with potential data imbalances.

4.4.4. The Receiver Operating Characteristic (ROC) Curve

The ROC analysis is a commonly used technique for evaluating the accuracy of medical diagnostic models. In the context of this study, the ROC curve illustrates the model's ability to distinguish between three classes: "healthy", "benign", and "malignant". For breast cancer classification, the ROC curve is used to assess how effectively the model differentiates between these classes.

In the ROC curve, an orange curve represents the model's performance. Each point on this curve corresponds to a different cutoff threshold used to classify samples as healthy,

benign, or malignant. The closer the curve is to the top-left corner—where the True Positive Rate (TPR) is 1 and the False Positive Rate (FPR) is 0—the better the model’s ability to differentiate accurately among the three classes, as depicted in Figure 11.

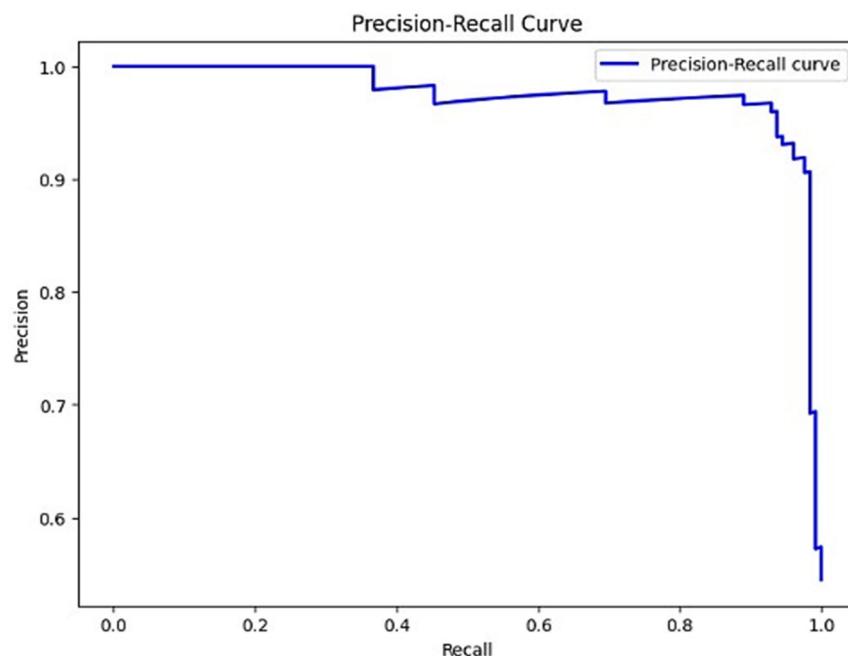


Figure 11. Precision–Recall Curve of the Model.

The dashed diagonal line on the ROC plot represents the performance of a random classifier, where TPR equals FPR at each point, indicating no discriminatory power. Effective diagnostic models, however, should have a ROC curve well above this line, indicating strong sensitivity and specificity.

The area under the curve (AUC) is a scalar metric between 0 and 1 that measures the model’s overall effectiveness. An AUC of 1.0 signifies perfect classification, while an AUC of 0.5 suggests random guessing. In this study, the model achieved an AUC of 0.97, indicating a high capacity to accurately differentiate among the healthy, benign, and malignant classes. An AUC of 0.97 translates to a 97% probability that the model will correctly assign a higher ranking to a randomly selected malignant case than to a randomly selected benign or healthy instance.

With a low incidence of false positives and a high rate of true positives, the model demonstrates strong diagnostic accuracy as depicted in Figure 12, effectively identifying malignant and benign cases while minimizing the misclassification of healthy cases. The high AUC score underscores the model’s reliability for use in clinical and diagnostic applications, making it well-suited for real-world medical diagnostics involving multiple classes.

4.5. Results and Discussion

With an emphasis on feature extraction and classification accuracy, a number of machine learning models were used in this study to classify thermal breast images, focusing on feature extraction effectiveness and classification accuracy. The models tested included CNNs combined with traditional classifiers such as SVM, KNN, and Random Forest, each evaluated across several metrics: validation and test accuracy, precision, recall, F1-score, specificity, AUC, latency, and CPU utilization.

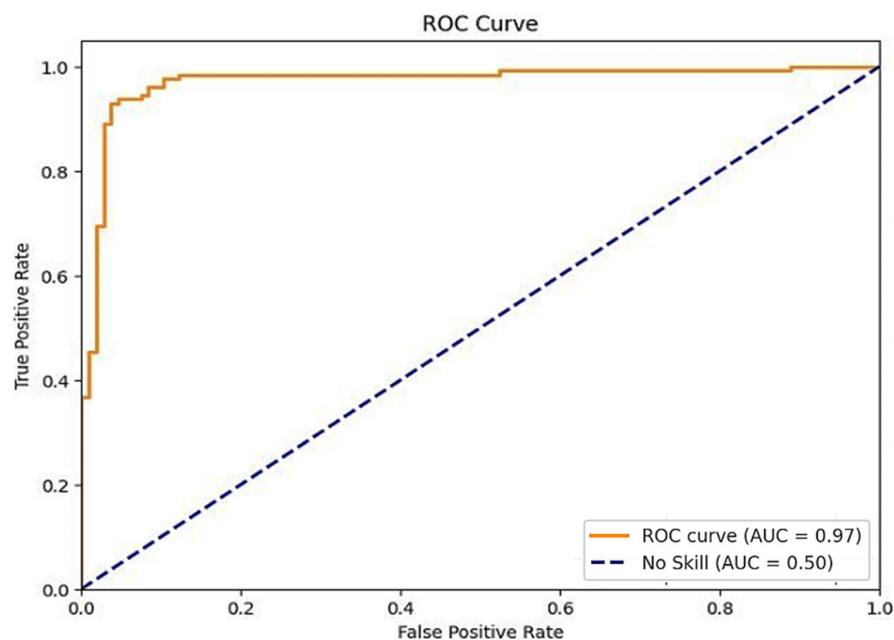


Figure 12. The ROC curve of the Model.

Based on the accuracy and AUC comparisons depicted in Figures 13 and 14, the ResNet152 + SVM combination consistently produced superior results. This model achieved high scores across a broad range of metrics, specifically with an accuracy of 0.97, an AUC of 0.99, a precision of 0.98, a recall of 0.94, an F1-score of 0.96, and a specificity of 0.97. Such high values confirm ResNet152's ability to effectively extract features that enhance the performance of the SVM classifier, especially in discriminating between the three classes: healthy, benign, and malignant. These metrics suggest that ResNet152 excels in clinical applications where accurate early detection is essential, aligning with its consistently high AUC values, which indicate strong sensitivity and low False Positive Rates (FPRs).

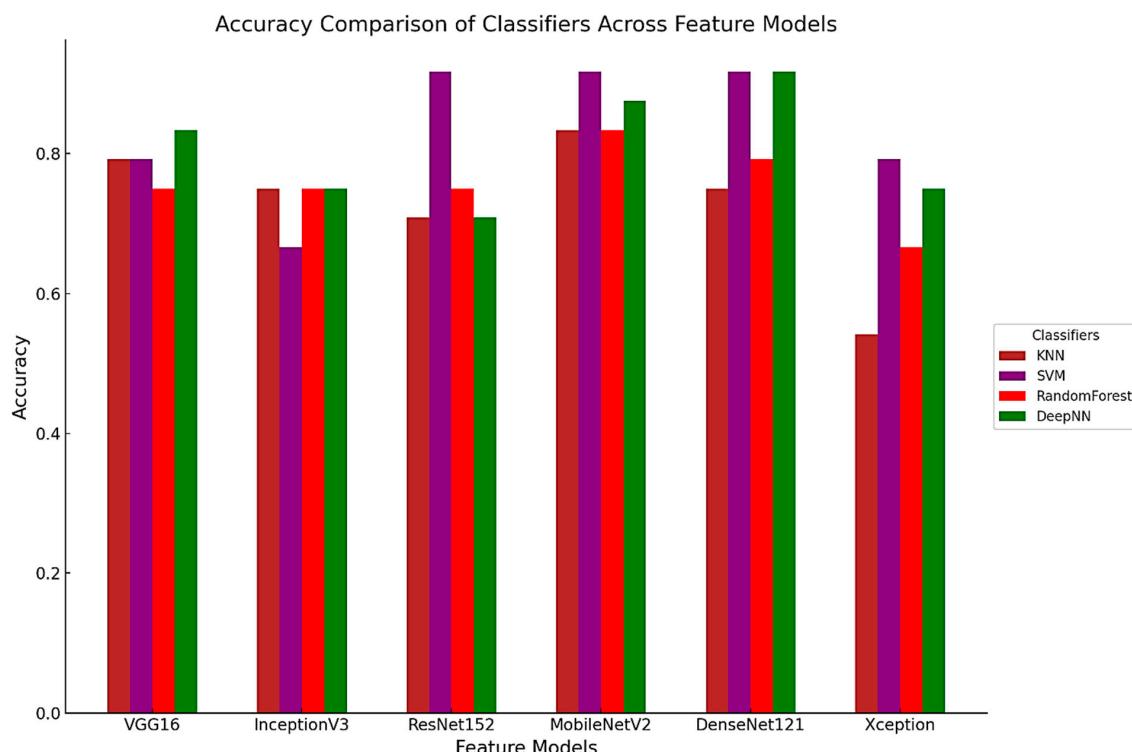


Figure 13. Accuracy Comparison of Classifiers across Feature Models.

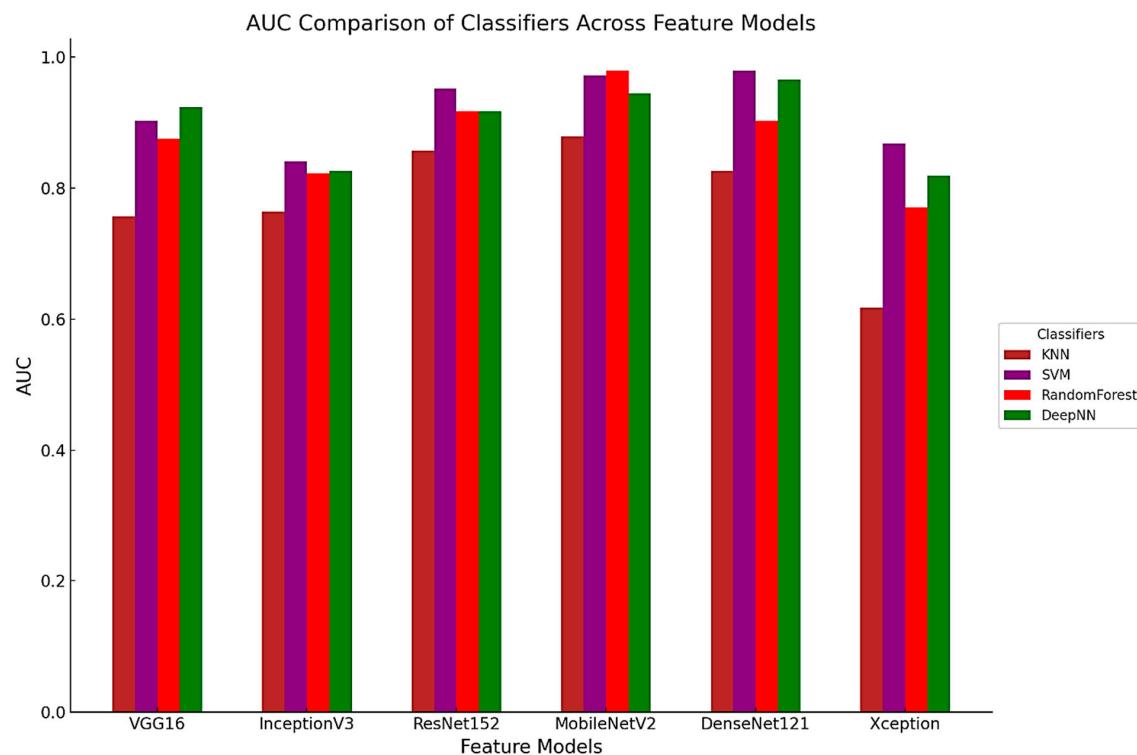


Figure 14. AUC Comparison of Classifiers across Feature Models for Breast Cancer Classification.

DenseNet121 and MobileNetV2 also demonstrated strong performance, particularly when paired with the KNN and SVM classifiers. DenseNet121 achieved high accuracy, with metrics nearly matching ResNet152, making it a robust alternative. MobileNetV2, which also yielded competitive AUC scores, emerged as a promising model in scenarios where computational efficiency is essential. This model provides a balance between high diagnostic accuracy and resource efficiency, thus making it suitable for real-time applications or contexts with limited computational resources. InceptionV3 and VGG16 yielded moderate AUC and accuracy values, indicating they may be less effective for fine-grained feature extraction in breast cancer classification. Xception showed the lowest performance across most metrics, which corroborates previous findings on its limitations in this application.

A summary of the performance metrics for the top models, particularly the SVM classifier with ResNet152 for feature extraction, is presented in Table 5. This model showed exceptional results across all relevant performance indicators, validating its suitability for clinical applications.

Table 5. Performance Comparison of Different Deep Learning Architectures with SVM Classifier.

Model	Accuracy	Precision	Recall	Specificity	F1 Score	AUC	Latency (s)	CPU Utilization (%)
ResNet152 + SVM	97.62%	95.79%	98.53%	94.52%	97.16%	99%	0.06	88.66
DenseNet121 + SVM	97.00%	94.85%	97.45%	93.60%	96.12%	98%	0.07	86.23
MobileNetV2 + SVM	94.00%	92.35%	94.56%	90.33%	93.43%	97%	0.08	80.50
InceptionV3 + SVM	94.00%	98.00%	92.00%	97.00%	94.00%	98%	0.10	82.81
VGG16 + SVM	93.00%	91.70%	93.45%	89.50%	92.56%	95%	0.08	94.93
Xception + SVM	90.50%	90.10%	90.80%	89.20%	90.44%	94%	0.12	76.45

Table 6 compares the performance of the Random Forest classifier with various architectures. DenseNet121 + RF demonstrates robust performance, achieving high accuracy (95%), precision (97%), and AUC (99%), with relatively low false positive rates and latency.

The results highlight Random Forest's potential for applications requiring high reliability and reduced error rates.

Table 6. Performance Comparison of Different Deep Learning Architectures with Random Forest Classifier.

Random Forest Classifier	Validation Accuracy	Test Accuracy	Precision	Recall	F1-Score	Specificity	AUC	False Positive Rate	Measure Latency	Measure CPU
Vgg16 + RF	0.96	0.96	0.96	0.96	0.96	0.95	0.99	0.05	0.12 s	96%
Inceptionv3 + RF	0.91	0.91	0.89	0.95	0.92	0.8	0.98	0.14	0.12 s	99%
Resnet151 + RF	0.94	0.95	0.95	0.96	0.96	0.94	0.99	0.07	0.16 s	111%
mobilenetv2 + RF	0.93	0.94	0.95	0.95	0.95	0.94	0.99	0.06	0.16	111%
Densenet121 + RF	0.92	0.95	0.97	0.95	0.96	0.96	0.99	0.04	0.13 s	98%
Xception + RF	0.93	0.94	0.92	0.96	0.94	0.91	0.99	0.09	0.10 s	96%

Table 7 shows the Decision Tree classifier's results, which generally lag behind those of the SVM and Random Forest classifiers. ResNet151 + DT performs best among the architectures, with an accuracy of 87% and AUC of 89%, but the latency (0.11 s) is similar to its competitors. In Table 8, the KNN classifier achieves strong results, with DenseNet121 + KNN achieving the best performance (96% accuracy, 99% AUC). While the classifier generally excels in recall and precision, latency and CPU utilization are slightly higher, making it a potential candidate for non-time-sensitive applications requiring high precision.

Table 7. Performance Comparison of Different Deep Learning Architectures with DT Classifier.

DT Classifier	Validation Accuracy	Test Accuracy	Precision	Recall	F1-Score	Specificity	AUC	False Positive Rate	Measure Latency	Measure CPU
vgg16 + DT	0.87	0.87	0.89	0.87	0.88	0.87	0.87	0.13	0.11 s	99%
Inceptionv3 + DT	0.85	0.83	0.85	0.83	0.84	0.84	0.83	0.17	0.12 s	98%
Resnet151 + DT	0.87	0.87	0.87	0.91	0.86	0.88	0.89	0.11	0.11 s	99%
mobilenetv2 + DT	0.83	0.87	0.87	0.9	0.88	0.84	0.87	0.16	0.9 s	93%
Densenet121 + DT	0.91	0.87	0.84	0.85	0.85	0.85	0.84	0.15	0.10 s	99%
Xception + DT	0.89	0.86	0.87	0.88	0.88	0.84	0.86	0.16	0.11 s	96%

Table 8. Performance Comparison of Different Deep Learning Architectures with KNN Classifier.

KNN Classifier	Validation Accuracy	Test Accuracy	Precision	Recall	F1-Score	Specificity	AUC	False Positive Rate	Measure Latency	Measure CPU
Vgg16 + KNN	0.95	0.93	0.97	0.91	0.94	0.97	0.97	0.03	0.05 s	77.00%
Inceptionv3 + KNN	0.91	0.91	0.94	0.89	0.92	0.93	0.96	0.07	0.06 s	96.40%
Resnet151 + KNN	0.92	0.95	0.98	0.93	0.96	0.98	0.99	0.02	0.10 s	98.30%
mobilev2 + KNN	0.93	0.93	0.97	0.9	0.93	0.96	0.97	0.04	0.18 s	89.60%
Densenet121 + KNN	0.96	0.94	0.99	0.91	0.95	0.99	0.99	0.01	0.06 s	93%
Xception + KNN	0.93	0.92	0.95	0.91	0.92	0.94	0.98	0.06	0.21 s	99.80%

Table 9 depicts the DNN classifier shows robust performance across architectures, with ResNet152 + DNN and DenseNet121 + DNN achieving strong results (95%+ accuracy and AUC values close to 99%). These results emphasize DNN's strength in capturing complex patterns while maintaining competitive latency and CPU utilization. In Table 10, Naive Bayes classifier lags behind other classifiers, with InceptionV3 + NB achieving the best performance (91% accuracy) in computationally efficient. This method demonstrates higher false positive rates and lower specificity, making it less suitable for high-stakes applications. Table 11 provides a broader comparison with prior works in the literature. The proposed ResNet152 + SVM model achieves the highest AUC (99%) and strong accuracy (95%), significantly outperforming previous studies like Silva et al. and Abdulkutty et al. [48].

The results validate the proposed model's ability to generalize well on the dataset, making it a reliable choice for real-world applications.

Table 9. Performance Comparison of Different Deep Learning Architectures with DNN Classifier.

DNN Classifier	Validation Accuracy	Test Accuracy	Precision	Recall	F1-Score	Specificity	AUC	False Positive Rate	Measure Latency	Measure CPU
Vgg16 + DNN	0.96	0.93	0.98	0.88	0.93	0.98	0.99	0.02	0.16 s	87.60%
Inceptionv3 + DNN	0.92	0.94	0.96	0.93	0.95	0.95	0.99	0.05	0.18 s	79.60%
Resnet152 + DNN	0.96	0.94	0.95	0.95	0.95	0.94	0.99	0.06	0.13 s	94.30%
mobilev2 + DNN	0.96	0.95	0.96	0.94	0.95	0.95	0.99	0.05	0.18 s	89.60%
Desnet121 + DNN	0.95	0.94	0.96	0.94	0.95	0.95	0.98	0.05	0.10 s	91%
xception + DNN	0.93	0.93	0.95	0.92	0.94	0.94	0.98	0.06	0.12 s	106.10%

Table 10. Performance Comparison of Different Deep Learning Architectures with Naive Bayes Classifier.

Naive Bayes Classifier	Validation Accuracy	Test Accuracy	Precision	Recall	F1-Score	Specificity	AUC	False Positive Rate	Measure Latency	Measure CPU
Vgg16 + NB	0.68	0.6	0.98	0.74	0.78	0.21	0.78	0.7	0.20 s	119%
Inceptionv3 + NB	0.91	0.91	0.89	0.95	0.92	0.86	0.98	0.14	0.18 s	119%
Resnet151 + NB	0.85	0.83	0.85	0.83	0.84	0.83	0.83	0.17	0.18 s	116%
mobilev2 + NB	0.81	0.79	0.83	0.78	0.8	0.8	0.87	0.17	0.20 s	119%
Desnet121 + NB	0.75	0.78	0.78	0.84	0.81	0.71	0.83	0.23	0.15 s	111%
xception + NB	0.79	0.81	0.84	0.8	0.82	0.82	0.84	0.18	0.21 s	121%

Table 11. Comparison of Different Machine Learning Architectures on the DMR Dataset.

Author(s)	Model(s)	Dataset	Accuracy (%)	Precision (%)	Recall (%)	AUC (%)
Silva et al. [36]	SVM	Not specified	90	85	88	89
Bhowmik et al. [37]	Multilayer Perceptron	DMR	95	94	92	97
D'Alessandro et al. [39]	SVM	DMR	94	93	91	95
Abdulkalutty et al. [48]	VGG16, InceptionV3	DMR-IR	87.3	86	85	89
Aarthy et al. [49]	SVM	DMR	97.6	95	93	96
Allugunti et al. [50]	SVM, Random Forest	DMR-IR	90.55	89	88	92
Proposed Model	ResNet152 + SVM	DMR	95	98	94	99
Bancos et al. [53]	ANN, Decision Tree, Bayesian	DMR-IR	78	75	72	80

4.5.1. Model Explainability Using GradCAM

The Grad-CAM [41] overlays provide an essential layer of explainability to the ResNet152 + SVM model by highlighting regions of interest (ROI) that significantly influence the model's predictions. These heatmaps, specific to each predicted class (e.g., normal, sick, malignant, benign), offer class-specific activation maps, shedding light on how the model differentiates among categories. By incorporating confidence scores (e.g., 0.80, 0.85) alongside these visualizations, clinicians can better understand and trust the model's decision-making process. This approach directly addresses the "black-box" nature of deep learning models by visually explaining their predictions. Moreover, Grad-CAM overlays emphasize biologically relevant regions, such as abnormal heat patterns [56], as depicted in Figures 15–19, enhancing the model's interpretability and robustness in feature extraction. These insights validate the model's focus areas against clinical knowledge and bridge the gap between AI-driven predictions and practical clinical applications, paving the way for greater trust and usability in real-world diagnostic settings.

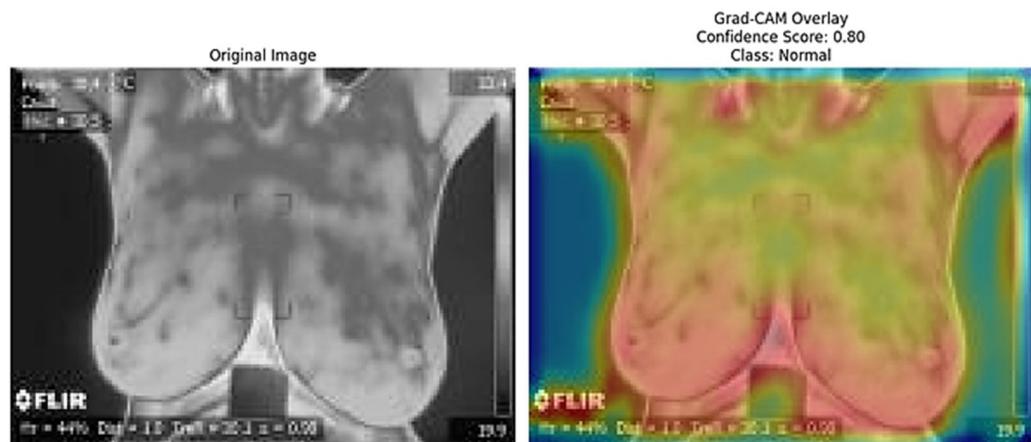


Figure 15. Grad-CAM Overlay for Normal Class (the original thermal image (**left**) alongside the Grad-CAM overlay (**right**) highlights the regions contributing to the model’s prediction of the “Normal” class with a confidence score of 0.80).

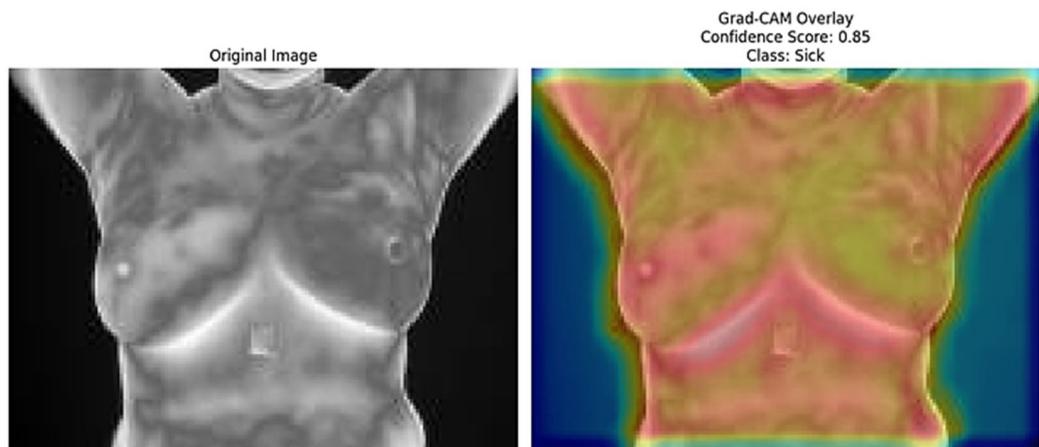


Figure 16. Grad-CAM Overlay for Sick Class (the original thermal image (**left**) alongside the Grad-CAM overlay (**right**) demonstrates the model’s focus on specific regions, leading to the prediction of the “Sick” class with a confidence score of 0.85).

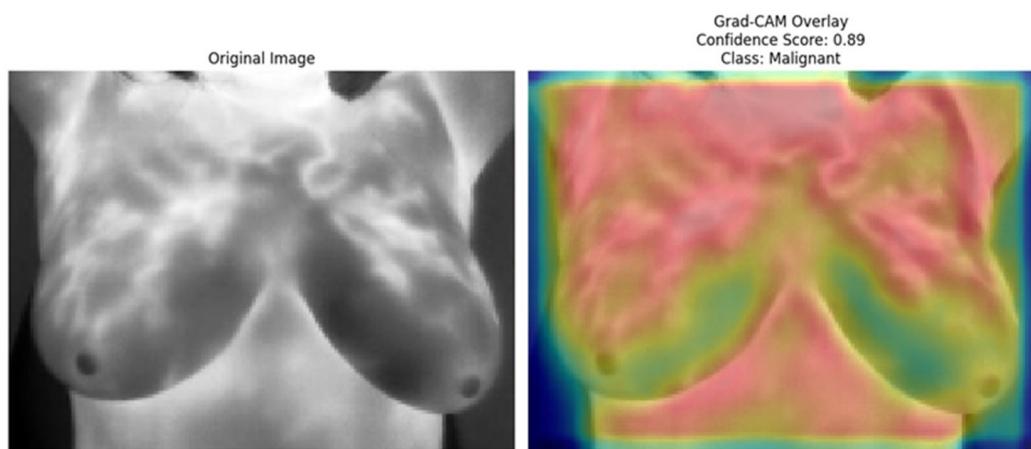


Figure 17. Grad-CAM Overlay for Malignant Class (the original thermal image (**left**) and its corresponding Grad-CAM overlay (**right**) show the model’s focus on abnormal heat regions, supporting the “malignant” classification with a confidence score of 0.89).

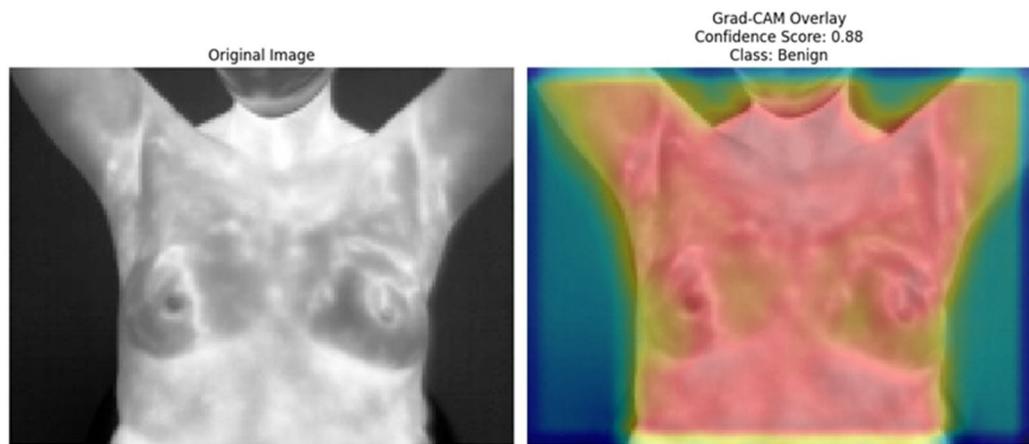


Figure 18. Grad-CAM Overlay for Benign Class (the original thermal image (**left**) and its Grad-CAM overlay (**right**) depict the regions contributing to the model’s prediction of the “benign” class with a confidence score of 0.88).

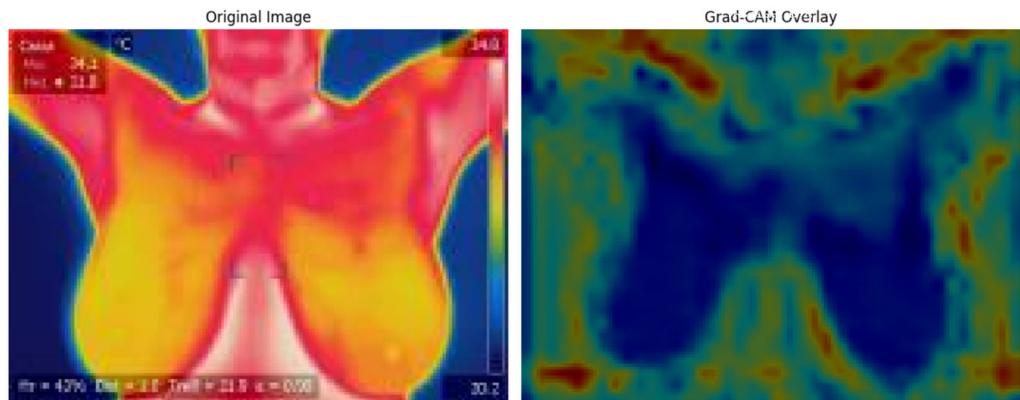


Figure 19. (**Left**) Original thermal image highlighting the temperature distribution across the chest area, with warmer regions indicated by red/yellow hues and cooler regions by blue/green hues. (**Right**) Grad-CAM overlay demonstrating the areas of highest model attention during classification, with cooler colours indicating less attention and warmer colours indicating regions of interest.

4.5.2. Evaluation and Justification of ResNet152 as the Top-Performing Model in Breast Cancer Detection

The effectiveness of the proposed ResNet152 + SVM model for breast cancer detection via thermal imaging has been rigorously evaluated and compared against state-of-the-art (SOTA) approaches. The model achieved a better performance across multiple metrics, including a test accuracy of 95%, a precision of 98%, a recall of 94%, and an AUC of 99%, demonstrating its capability to outperform alternative architectures such as InceptionV3 and MobileNetV2. ResNet152’s architectural advantages, including its deep 152-layer network and residual connections, enable it to capture both low-level and high-level features effectively. This hierarchical feature extraction is crucial for distinguishing subtle patterns in thermal images and facilitating accurate classification of healthy, benign, and malignant cases.

In addition to its accuracy, ResNet152 has demonstrated robustness in handling high-resolution and complex datasets, a critical requirement in medical imaging. While InceptionV3 achieved comparable precision and MobileNetV2 offered superior computational efficiency, neither model matched the overall balance of performance, robustness, and reliability exhibited by ResNet152. The model’s latency of 0.06 s and CPU utilization of 88.66% further underscore its suitability for real-time diagnostic applications, though these results were obtained under controlled experimental conditions.

Despite its strengths, the high performance achieved on a relatively small dataset raises concerns about potential overfitting. Mitigation strategies such as data augmentation and cross-validation were implemented; however, further validation on larger, more diverse datasets is essential to confirm the model's generalizability. Moving forward, lightweight architectures like MobileNetV2 will be explored for resource-constrained environments alongside hybrid approaches integrating multiple architectures to optimize performance across various clinical settings. These efforts aim to extend ResNet152's success to broader real-world diagnostic applications. The summary of the comparative approach with the state-of-the-art (SOTA) is presented in Table 12.

Table 12. Comparison of Proposed Models with State-of-the-Art Approaches for Breast Cancer Detection.

Study	Model(s) Used	Dataset	Test Accuracy	Precision	Recall	AUC	Latency
Proposed	ResNet152 + SVM	DMR-IR	95%	98%	94%	99%	0.06 s
Dabhade et al. [8]	Random Forest, SVM	Proprietary	98.4% (SVM)	-	-	-	-
Tiwari et al. [14]	VGG16	Proprietary	99%	-	-	-	-
Mambou et al. [26]	SVM, DNN	Proprietary	94%	-	-	-	-
Bezerra et al. [27]	Naive Bayes, SVM	DMR-IR	95% (SVM)	92%	93%	97%	-
Zuluaga-Gomez et al. [40]	CNN	DMR-IR	92%	-	-	92%	-
Karthiga and Narasimhan [51]	VGG16, InceptionV3	DMR-IR	87.3% (VGG16)	-	-	-	-
Proposed Model Alternative	DenseNet 121 + SVM	DMR-IR	97%	95%	97%	98%	0.07 s

In comparison, Dabhade et al. [8] achieved a 98.4% accuracy using SVM, yet specific values for precision, recall, AUC, and latency were not provided, limiting the ability to assess its diagnostic reliability fully. Similarly, Tiwari et al. [14] reported a 99% accuracy with the VGG16 model, but details on latency and other key metrics are absent, making it difficult to determine computational feasibility for real-time use. Other studies, such as Mambou et al. [26], used deep neural networks and SVM classifiers and achieved an accuracy of 94%, which is lower than our ResNet152 + SVM model.

The proposed DenseNet121 + SVM model also demonstrated high performance, with a test accuracy of 97% and an AUC of 98%, positioning it as a robust alternative to the ResNet152 + SVM model. The DenseNet121 + SVM's latency of 0.07 s confirms its potential for practical deployment in clinical scenarios, especially where diagnostic speed and accuracy are essential. While MobileNetV2 and VGG16 models were competitive in terms of accuracy (94% and 93%, respectively), their slightly higher latency and CPU demands make them less optimal than ResNet152 or DenseNet121 when both accuracy and efficiency are considered crucial.

5. Conclusions

This research evaluated the effectiveness of transfer learning models and machine learning classifiers for breast cancer detection (healthy, malignant, and benign) using thermal images. By integrating deep learning feature extraction techniques with traditional classifiers, this study demonstrated the efficacy of specific model combinations in accurately classifying breast tissue images. Among the tested models, the ResNet152 + SVM combination consistently delivered superior performance across multiple metrics, achieving a validation accuracy of 97%, a test accuracy of 95%, a precision of 98%, a recall of 94%, an F1-Score of 96%, a specificity of 97%, and an AUC of 99%. These results underscore ResNet152's capability to extract critical features from thermal images, enhancing the SVM

classifier's accuracy and making it a reliable solution for early breast cancer detection in clinical applications.

Other models also exhibited noteworthy performance. DenseNet121 + SVM achieved a comparable test accuracy of 97% and demonstrated high reliability across all metrics, making it a strong alternative to ResNet152. InceptionV3 + SVM achieved the highest precision (98%) along with a test accuracy of 94%, indicating a reduced susceptibility to overfitting, which is valuable for diagnostic purposes. MobileNetV2 and VGG16, while consistent in performance, are better suited for scenarios prioritizing computational efficiency due to their balanced accuracy and lower latency. These findings affirm the potential of combining transfer learning models with machine learning classifiers for breast cancer diagnosis, especially when balancing diagnostic precision with computational constraints.

To address the limitations of the DMR-IR and Breast Mendeley thermal datasets, which lack diversity in demographics and tumor types, future work will involve collaboration with specialized breast cancer centers to create a larger, more representative dataset. This effort will include collecting thermal images across diverse age groups, ethnicities, and tumor subtypes to improve model robustness and applicability. Overfitting concerns were mitigated in this study through data augmentation techniques such as rotation, scaling, and flipping, alongside cross-validation. However, validation on a larger dataset is essential to ensure more reliable conclusions. Future work will also focus on deployment testing in real-time clinical environments to validate the reported latency and computational efficiency.

Additionally, future research will explore multimodal comparisons to benchmark the effectiveness of thermal imaging against established diagnostic methods, such as ultrasound. These comparisons will provide a comprehensive evaluation of thermal imaging's diagnostic value. Anticipated collaborations with specialized breast cancer centers will also leverage dedicated breast thermography cameras to ensure higher quality and consistency in thermal images. Such advancements aim to enhance the generalizability and robustness of the models, paving the way for their deployment in real-world diagnostic settings.

Author Contributions: All authors have contributed equally to the development and preparation of this manuscript in accordance with the CRediT taxonomy guidelines. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset used in this research is entirely publicly available. It is derived from publicly accessible datasets found in the Visual Lab repository (<https://visual.ic.uff.br/dmi>) and from a dataset containing benign and malignant breast thermography classes, which is available through Mendeley Data.

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

References

1. Sun, P.; Yu, C.; Yin, L.; Chen, Y.; Sun, Z.; Zhang, T.; Shuai, P.; Zeng, K.; Yao, X.; Chen, J.; et al. Global, regional, and national burden of female cancers in women of child-bearing age, 1990–2021: Analysis of data from the global burden of disease study 2021. *eClinicalMedicine* **2024**, *74*, 102713. [[CrossRef](#)] [[PubMed](#)]
2. Arnold, M.; Morgan, E.; Rumgay, H.; Mafra, A.; Singh, D.; Laversanne, M.; Vignat, J.; Gralow, J.R.; Cardoso, F.; Siesling, S.; et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *Breast* **2022**, *66*, 15–23. [[CrossRef](#)]
3. Jalloul, R.; Chethan, H.K.; Alkhateeb, R. A Review of Machine Learning Techniques for the Classification and Detection of Breast Cancer from Medical Images. *Diagnostics* **2023**, *13*, 2460. [[CrossRef](#)] [[PubMed](#)]
4. Jiang, R.Y.; Fang, Z.R.; Zhang, H.P.; Xu, J.Y.; Zhu, J.Y.; Chen, K.Y.; Wang, W.; Jiang, X.; Wang, X.J. Ginsenosides: Changing the basic hallmarks of cancer cells to achieve the purpose of treating breast cancer. *Chin. Med.* **2023**, *18*, 125. [[CrossRef](#)]
5. Upadhyaya, V. Predictive Analytics in Medical Diagnosis. In *Intelligent Data Analytics for Bioinformatics and Biomedical Systems*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2024; pp. 27–66. [[CrossRef](#)]

6. Guo, Y.; Zhang, H.; Yuan, L.; Chen, W.; Zhao, H.; Yu, Q.Q.; Shi, W. Machine learning and new insights for breast cancer diagnosis. *J. Int. Med. Res.* **2024**, *52*, 03000605241237867. [[CrossRef](#)]
7. Tsietso, D.; Yahya, A.; Samikannu, R. A Review on Thermal Imaging-Based Breast Cancer Detection Using Deep Learning. *Mob. Inf. Syst.* **2022**, *2022*, 8952849. [[CrossRef](#)]
8. Lakshman, K.; Dabhade, S.B.; Rode, Y.S.; Dabhade, K.; Deshmukh, S.; Maheshwari, R. Identification of Breast Cancer from Thermal Imaging using SVM and Random Forest Method. In Proceedings of the 2021 5th International Conference on Trends in Elec-tronics and Informatics (ICOEI), Tirunelveli, India, 3–5 June 2021. [[CrossRef](#)]
9. Iacob, R.; Iacob, E.R.; Stoicescu, E.R.; Ghenciu, D.M.; Cocolea, D.M.; Constantinescu, A.; Ghenciu, L.A.; Manolescu, D.L. Evaluating the Role of Breast Ultrasound in Early Detection of Breast Cancer in Low- and Middle-Income Countries: A Comprehensive Narrative Review. *Bioengineering* **2024**, *11*, 262. [[CrossRef](#)] [[PubMed](#)]
10. Mirasbekov, Y.; Aidossov, N.; Mashekova, A.; Zarikas, V.; Zhao, Y.; Ng, E.Y.K.; Midlenko, A. Fully Interpretable Deep Learning Model Using IR Thermal Images for Possible Breast Cancer Cases. *Biomimetics* **2024**, *9*, 609. [[CrossRef](#)]
11. Sritharan, N.; Gutierrez, C.; Perez-Raya, I.; Gonzalez-Hernandez, J.L.; Owens, A.; Dabydeen, D.; Medeiros, L.; Kandlikar, S.; Phatak, P. Breast Cancer Screening Using Inverse Modeling of Surface Temperatures and Steady-State Thermal Imaging. *Cancers* **2024**, *16*, 2264. [[CrossRef](#)] [[PubMed](#)]
12. Islam, T.; Sheakh, M.A.; Tahasin, M.S.; Hena, M.H.; Akash, S.; Jardan, Y.a.B.; FentahunWondmie, G.; Nafidi, H.A.; Bourhia, M. Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI. *Sci. Rep.* **2024**, *14*, 8487. [[CrossRef](#)] [[PubMed](#)]
13. Ikechukwu, A.V.; Bhimshetty, S.; Deepu, R.; Mala, M.V. Advances in Thermal Imaging: A Convolutional Neural Network Ap-proach for Improved Breast Cancer Diagnosis. In Proceedings of the 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT), Bengaluru, India, 15–16 March 2024; pp. 1–7. [[CrossRef](#)]
14. Tiwari, D.; Dixit, M.; Gupta, K. Deep Multi-View Breast Cancer Detection: A Multi-View Concatenated Infrared Thermal Images Based Breast Cancer Detection System Using Deep Transfer Learning. *Trait. Signal* **2021**, *38*, 1699–1711. [[CrossRef](#)]
15. Lakkis, N.A.; Abdallah, R.M.; Musharrafieh, U.M.; Issa, H.G.; Osman, M.H. Epidemiology of Breast, Corpus Uteri, and Ovarian Cancers in Lebanon With Emphasis on Breast Cancer Incidence Trends and Risk Factors Compared to Regional and Global Rates. *Cancer Control* **2024**, *31*, 10732748241236266. [[CrossRef](#)] [[PubMed](#)]
16. Desai, M.; Shah, M. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). *Clin. eHealth* **2021**, *4*, 1–11. [[CrossRef](#)]
17. Alshehri, A.; AlSaeed, D. Breast Cancer Detection in Thermography Using Convolutional Neural Networks (CNNs) with Deep Attention Mechanisms. *Appl. Sci.* **2022**, *12*, 12922. [[CrossRef](#)]
18. Wen, X.; Guo, X.; Wang, S.; Lu, Z.; Zhang, Y. Breast cancer diagnosis: A systematic review. *Biocybern. Biomed. Eng.* **2024**, *44*, 119–148. [[CrossRef](#)]
19. Agughasi, V.I. The Superiority of Fine-tuning over Full-training for the Efficient Diagnosis of COPD from CXR Images. *Intel. Artif.* **2024**, *27*, 62–79. [[CrossRef](#)]
20. Agughasi, V.I. Leveraging Transfer Learning for Efficient Diagnosis of COPD Using CXR Images and Explainable AI Tech-niques. *Intel. Artif.* **2024**, *27*, 133–151. [[CrossRef](#)]
21. Sharafaddini, A.M.; Esfahani, K.K.; Mansouri, N. Deep learning approaches to detect breast cancer: A comprehensive review. *Multimed. Tools Appl.* **2024**. [[CrossRef](#)]
22. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Con-nections on Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; Volume 31, pp. 4278–4284. [[CrossRef](#)]
23. Ekici, S.; Jawzal, H. Breast cancer diagnosis using thermography and convolutional neural networks. *Med. Hypotheses* **2020**, *137*, 109542. [[CrossRef](#)]
24. Chowdhury, N.A.; Wang, L.; Gu, L.; Kaya, M. Machine Learning for Early Breast Cancer Detection. *J. Eng. Sci. Med. Diagn. Ther.* **2024**, *8*, 010801. [[CrossRef](#)]
25. Abed, A.H.; Shaaban, E.M. Modeling Deep Neural Networks For Breast Cancer Thermography Classification: A Review Study. *Int. J. Adv. Netw. Appl.* **2021**, *13*, 4939–4946. [[CrossRef](#)]
26. Mambou, S.; Maresova, P.; Krejcar, O.; Selamat, A.; Kuca, K. Breast Cancer Detection Using Infrared Thermal Imaging and a Deep Learning Model. *Sensors* **2018**, *18*, 2799. [[CrossRef](#)]
27. Amethiya, Y.; Pipariya, P.; Patel, S.; Shah, M. Comparative analysis of breast cancer detection using machine learning and biosensors. *Intell. Med.* **2022**, *2*, 69–81. [[CrossRef](#)]
28. Gonçalves, C.B.; Souza, J.R.; Fernandes, H. CNN architecture optimization using bio-inspired algorithms for breast cancer detection in infrared images. *Comput. Biol. Med.* **2022**, *142*, 105205. [[CrossRef](#)]
29. Silva, L.F.; Saade, D.C.M.; Sequeiros, G.O.; Silva, A.C.; Paiva, A.C.; Bravo, R.S.; Conci, A. A New Database for Breast Research with Infrared Image. *J. Med. Imaging Health Inform.* **2014**, *4*, 92–100. [[CrossRef](#)]

30. Aarthy, S.L.; Prabu, S. Classification of breast cancer based on thermal image using support vector machine. *Int. J. Bioinform. Res. Appl.* **2019**, *15*, 51–67. [CrossRef]
31. Allugunti, V.R. Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *Int. J. Eng. Comput. Sci.* **2022**, *4*, 49–56. [CrossRef]
32. Karthiga, R.; Narasimhan, K. Medical imaging technique using curvelet transform and machine learning for the automated diagnosis of breast cancer from thermal image. *Pattern Anal. Appl.* **2021**, *24*, 981–991. [CrossRef]
33. Banco De Imagens Mastológicas. Available online: <https://visual.ic.uff.br/dmi/> (accessed on 4 November 2024).
34. Nissar, I.; Alam, S.; Masood, S. Computationally efficient LC-SCS deep learning model for breast cancer classification using thermal imaging. *Neural Comput. Appl.* **2024**, *36*, 16233–16250. [CrossRef]
35. Bhowmik, M.K.; Gogoi, U.R.; Majumdar, G.; Bhattacharjee, D.; Datta, D.; Ghosh, A.K. Designing of Ground Truth Annotated DBT-TU-JU Breast Thermogram Database towards Early Abnormality Prediction. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 1238–1249. [CrossRef]
36. Torres-Galvan, J.C.; Guevara, E.; Gonzalez, F.J. Comparison of Deep Learning Architectures for Pre-Screening of Breast Cancer Thermograms. In Proceedings of the 2019 Photonics North (PN), Quebec City, QC, Canada, 21–23 May 2019. [CrossRef]
37. D’Alessandro, G.; Tavakolian, P.; Sfarra, S. A Review of Techniques and Bio-Heat Transfer Models Supporting Infrared Thermal Imaging for Diagnosis of Malignancy. *Appl. Sci.* **2024**, *14*, 1603. [CrossRef]
38. Zuluaga-Gomez, J.; Masry, Z.A.; Benaggoune, K.; Meraghni, S.; Zerhouni, N. A CNN-based methodology for breast cancer diagnosis using thermal images. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2020**, *9*, 131–145. [CrossRef]
39. Abdulkutty, F.; Akbari, Y.; Al-Maadeed, S.; Bouridane, A.; Hamoudi, R. Advancing Histopathology-Based Breast Cancer Diagnosis: Insights into Multi-Modality and Explainability. *arXiv* **2024**, arXiv:2406.12897. [CrossRef]
40. Ikechukwu, A.V.; Murali, S. CX-Net: An efficient ensemble semantic deep neural network for ROI identification from chest-x-ray images for COPD diagnosis. *Mach. Learn. Sci. Technol.* **2023**, *4*, 025021. [CrossRef]
41. Iyadurai, J.; Chandrasekharan, M.; Muthusamy, S.; Panchal, H. An Extensive Review on Emerging Advancements in Thermography and Convolutional Neural Networks for Breast Cancer Detection. *Wirel. Pers. Commun.* **2024**, *137*, 1797–1821. [CrossRef]
42. Martin, P.P.; Graulich, N. Navigating the data frontier in science assessment: Advancing data augmentation strategies for machine learning applications with generative artificial intelligence. *Comput. Educ. Artif. Intell.* **2024**, *7*, 100265. [CrossRef]
43. Schwartz, R.G.; Horner, C.; Kane, R.; Getson, P.; Brioschi, M.; Pittman, J.; Rind, B.; Campbell, J.; Ehle, E.; Mustovoy, A.; et al. Guidelines for Breast Thermology. In *Guidelines for Breast Thermology*; American Academy of Thermology: Greenville, SC, USA, 2021.
44. Bezerra, L.; Ribeiro, R.; Lyra, P.; Lima, R. An empirical correlation to estimate thermal properties of the breast and of the breast nodule using thermographic images and optimization techniques. *Int. J. Heat Mass Transf.* **2020**, *149*, 119215. [CrossRef]
45. Gogoi, U.R.; Majumdar, G.; Bhowmik, M.K.; Ghosh, A.K.; Bhattacharjee, D. Breast abnormality detection through statistical feature analysis using infrared thermograms. In Proceedings of the 2015 International Symposium on Advanced Computing and Communication (ISACC), Silchar, India, 14–15 September 2015; pp. 258–265. [CrossRef]
46. Resmini, R.; Conci, A.; da Silva, L.F.; Sequeiros, G.O.; Araújo, F.; de Araújo, C.; dos Santos Araújo, A. Application of Infrared Images to Diagnosis and Modeling of Breast. In *Application of Infrared to Biomedical Sciences*; Ng, E.Y., Etehad tavakol, M., Eds.; Springer: Singapore, 2017; pp. 159–173. [CrossRef]
47. Husaini, M.A.S.A.; Habaebi, M.H.; Hameed, S.A.; Islam, M.R.; Gunawan, T.S. A Systematic Review of Breast Cancer Detection Using Thermography and Neural Networks. *IEEE Access* **2020**, *8*, 208922–208937. [CrossRef]
48. Omranipour, R.; Kazemian, A.; Alipour, S.; Najafi, M.; Alidoosti, M. Comparison of the Accuracy of Thermography and Mammography in the Detection of Breast Cancer. *Breast Care* **2016**, *11*, 260–264. [CrossRef]
49. Dharani, N.P.; Immadi, I.G.; Narayana, M.V. Enhanced deep learning model for diagnosing breast cancer using thermal images. *Soft Comput.* **2024**, *28*, 8423–8434. [CrossRef]
50. Yadav, S.S.; Jadhav, S.M. Thermal infrared imaging-based breast cancer diagnosis using machine learning techniques. *Multimed. Tools Appl.* **2020**, *81*, 13139–13157. [CrossRef]
51. Abraham, A.; Ohsawa, Y.; Gandhi, N.; Jabbar, M.; Haqiq, A.; McLoone, S.; Issac, B. Advances in Intelligent systems and computing. In Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020), online, India, 15–18 December 2020; Springer Nature: Dordrecht, The Netherlands, 2021. [CrossRef]
52. Al Husaini, M.A.S.; Habaebi, M.H.; Islam, M.R. Real-time thermography for breast cancer detection with deep learning. *Discov. Artif. Intell.* **2024**, *4*, 57. [CrossRef]
53. Masekela, A.; Zhao, Y.; Ng, E.Y.; Zarikas, V.; Fok, S.C.; Mukhmetov, O. Early detection of the breast cancer using infrared technology—A comprehensive review. *Therm. Sci. Eng. Prog.* **2022**, *27*, 101142. [CrossRef]
54. Rodriguez-Guerrero, S.; Correa, H.L.; Restrepo-Girón, A.D.; Reyes, L.A.; Olave, L.A.; Diaz, S. Breast Thermography. *Mendeley Data* **2024**, *3*. [CrossRef]

55. Omundiagbe, D.A.; Veeramani, S.; Sidhu, A.S. Machine Learning Classification Techniques for Breast Cancer Diagnosis. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *495*, 012033. [[CrossRef](#)]
56. Ikechukwu, A.V.; Murali, S. xAI: An Explainable AI Model for the Diagnosis of COPD from CXR Images. In Proceedings of the 2023 IEEE 2nd International Conference on Data, Decision and Systems (ICDDS), Mangaluru, India, 1–2 December 2023; pp. 1–6. [[CrossRef](#)]

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