DEVELOPMENT OF AN AI-POWERED EARLY BREAST CANCER DETECTION SYSTEM USING THERMAL IMAGING, EFFICIENTNETB3, AND SUPPORT VECTOR MACHINE CLASSIFICATION

BY

NDUBUISI OGECHI JOSEPHINE

GOU/U23/CSC/1291

DEPARTMENT OF COMPUTER SCIENCE AND MATHEMATICS, FACULTY OF NATURAL SCIENCES AND ENVIRONMENTAL STUDIES, GODFREY OKOYE UNIVERSITY, ENUGU STATE. IN PARTIAL FULFILLMENT OF THE AWARD OF BACHELOR OF SCIENCE (B.SC), DEGREE IN COMPUTER SCIENCE.

SUPERVISOR: MR. NJOKU CAMILLUS

MAY, 2026.

APPROVAL PAGE

This research, titled "**AI-POWERED EARLY BREAST CANCER DETECTION SYSTEM USING THERMAL IMAGING, EFFICIENTNETB3, AND SUPPORT VECTOR MACHINE CLASSIFICATION**," has been assessed and approved by the Department of Computer Science and Mathematics Committees in the Faculty of Natural Sciences and Environmental Studies, Godfrey Okoye University, Enugu, Nigeria.

Engr. Camillus Njoku ....

Supervisor Signature Date

External Examiner

....

External Examiner Signature Date

Dr. Frank Okebanama

....

HOD, Department of Computer Science and Mathematics   Signature   Date

Prof. N. M. Ozofor

....

Dean, Faculty of Computing And Information Technology.   Signature   Date

CERTIFICATION

The researcher certify that she completed the research included therein and that it was primarily based on my observations and investigation. Consequently, I will fully accept the University's decision if I am found to have cheated on the assignment.

NDUBUISI OGECHI JOSEPHINE
GOU/U23/CSC/1291

DEDICATION

This project is dedicated to every woman in Nigeria and across the developing world who has faced a breast cancer diagnosis too late, not because medicine failed them, but because the tools for early detection were never within their reach. It is dedicated to the researchers, clinicians, and engineers who believe that artificial intelligence can change that — and to every future patient whose life may be saved because of work like this.

ACKNOWLEDGEMENTS

First and foremost, I give all glory and thanks to God Almighty for His grace, strength, wisdom, and guidance throughout the course of this research. His unfailing support made the successful completion of this work possible.
I sincerely appreciate my supervisor, Mr. Njoku Camillus, for his invaluable guidance, constructive feedback, and encouragement throughout this study. His expertise and commitment greatly contributed to the success of this project.
I am also grateful to Dr. Frank Okebanama, the Head of the Department of Computer Science and Mathematics, and to all the lecturers in the department for their knowledge, mentorship, and contributions to my academic development. I equally appreciate the management of Godfrey Okoye University for providing a conducive environment for learning and research.
My heartfelt gratitude goes to my husband, Mr. Gabriel Arinze M., my children, family members, friends, and colleagues for their unwavering support, prayers, encouragement, and understanding during this research journey.
Finally, I acknowledge the researchers and institutions that made their thermal imaging datasets publicly available, particularly the DMR-IR Database and the Mendeley Breast Thermography Dataset, whose contributions provided the foundation for this study.

ABSTRACT

Breast cancer remains one of the leading causes of cancer-related deaths among women worldwide, particularly in developing countries where access to early screening facilities is limited. This study was motivated by the need for a non-invasive, radiation-free, and accessible method for supporting early breast cancer detection using thermal imaging and artificial intelligence. To address this challenge, a hybrid EfficientNetB3–SVM model was developed for the classification of breast thermographic images as normal/benign or abnormal/malignant. Thermal images from the DMR-IR and Mendeley Breast Thermography datasets were combined, producing a total of 619 images that were preprocessed using contrast enhancement, resizing, normalization, and data augmentation techniques. EfficientNetB3 was employed for deep feature extraction, while a Support Vector Machine with an RBF kernel performed the final classification. The developed model was integrated into a full-stack web application named BreastCare AI, featuring user authentication, real-time prediction, Grad-CAM visual explanations, automated PDF report generation, scan history, and comparison functionalities. Experimental evaluation on the test dataset achieved an accuracy of 81.45%, sensitivity of 85.37%, specificity of 79.52%, and an AUC-ROC score of 0.87, while cross-validation demonstrated the model's robustness and reliability. The study contributes an explainable and deployable AI-powered breast cancer screening system that combines intelligent classification, visual interpretability, and patient monitoring within a single platform, thereby providing a practical decision-support tool for early breast cancer detection.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION

1.1 Background of the Study

Breast cancer is one of the biggest public health issues that affect women in this twenty-first century. According to the World Health Organization (2023), it is the most common diagnosed cancer among women globally, accounting for approximately 2.3 million new cases, along with 685,000 deaths. In Nigeria, however, the situation is especially grim. In their systematic review and meta-analysis on staging of cancers across sub-Saharan Africa, Jedy-Agba et al. (2016) report that the majority of cases diagnosed in Nigeria fall under Stages III and IV, during which chances of survival are very low while treatment is difficult. At stage I, five-year survival rate is above 99%, but at stage IV, it drops down to around 28% (Łukasiewicz et al., 2021). There is a huge difference between what modern medicine can achieve and what it can offer Nigerians with breast cancer, and it is all because of late detection.

Late diagnosis in most cases occurs because of the nature of mammography itself. This imaging test, considered the gold standard for screening breast cancer worldwide, needs sophisticated capital-intensive equipment, specialized radiographic technologists, power, and technical support (Naki et al., 2024). Previous research studies have shown that mammograms are available only in a few select tertiary health facilities in the cities of Lagos, Abuja, and Enugu; thus, the majority of Nigerians remain outside its scope (Amir et al., 2019). Furthermore, mammography uses radiation in the form of ionized X-rays and, hence, may not be a safe procedure in terms of safety risks due to exposure over time. Mammograms also do not work well in detecting abnormalities in cases of women with dense breasts, a characteristic common to young African women.

Thermography provides an altogether different solution to the detection challenge. Its biological basis was established by Lawson in 1956, where he showed that the surface temperature of the breast tissue lying over the tumor site is significantly higher than the rest of the surrounding healthy tissue (Goni-Arana et al., 2025). The process works on the principle of angiogenesis or the creation of new and abnormal blood vessels as the malignant tumor expands in size. As compared to mature vessels, these new and immature blood vessels leak and have increased metabolic activity, generating extra heat. Infrared cameras today can detect even small variations of less than 0.1 degrees celsius, which makes it possible for them to capture this signal without any physical intervention or exposure to any form of ionizing radiation. Furthermore, the changes taking place at the blood vessel level occur many years prior to tumor development, which gives thermography a definite advantage over other techniques like mammography.

Nonetheless, the application of thermal images for cancer screening has not become routine practice, mainly due to the challenge of interpreting the results correctly – differentiating the unique signature of an early malignancy from natural fluctuations, benign pathology, inflammation, and other factors demands pattern recognition expertise, which even experienced physicians lack (Ryan & Agaian, 2025). Deep learning and, more generally, artificial intelligence offer a way out in this challenging scenario. Litjens et al. (2017) performed an extensive review of more than 300 studies on the applications of deep learning in medicine and reported that CNN models demonstrate comparable diagnostic accuracy with that of specialists on almost all imaging tasks. Deep learning algorithms will facilitate thermography-based screening by eliminating the need for a specialist, whose availability may be a limiting factor in resource-poor locations.

The practical problem in using deep learning to classify thermographic breast cancer is the limited availability of labeled training data. Creating a dataset in medicine involves ethics review, patient consent, expertise in the field, and significant labor, thus publicly available datasets for thermography number in the hundreds rather than the millions required to train generic computer vision models (Mohi ud din et al., 2022). The solution proposed by transfer learning is clear. Instead of training the neural network from randomly initialized weights on the limited thermographic dataset, transfer learning trains the model first on a large generic image dataset such as ImageNet, which contains 1.2 million images, and then fine tunes it to perform the specific medical task (Gu & Lee, 2024). Using EfficientNets, Tan and Le (2019) showed how scaling up the network depth, width, and input image resolution simultaneously leads to greater performance on ImageNet while minimizing parameters, hence making EfficientNetB3 a suitable candidate for transfer learning in small medical datasets prone to overfitting.

Yet, CNNs have an inherent drawback when implemented as standalone classifiers that is especially critical in the context of small datasets in medical imaging. Softmax, which converts CNN features to probabilities, is designed to maximize the likelihood but not maximize margins between different classes and produces overconfident predictions as a result of training on insufficient amounts of data. One of the ways to overcome this problem is using the hybrid CNN-SVM approach, where CNN operates only as a feature extractor and provides the compressed numerical representation of an input image generated by its penultimate layer, whereas SVM makes the decision based on the output features. According to Cortes & Vapnik, SVM solves the problem of maximizing the margins between the points in high-dimensional spaces, thus providing better generalization properties than the empirical risk minimization. It is supported by several independent researchers, including Araújo and colleagues (2017) and Rodrigues and colleagues (2019), have demonstrated that CNN-SVM hybrids consistently outperform standalone CNNs on small medical imaging datasets.

Apart from its accuracy of prediction, another factor that needs to be considered for the integration of any AI diagnostic system in practice is that of explainability. It would be highly irrational to act upon predictions made when we cannot interrogate the basis behind such predictions. This need was fulfilled by Selvaraju et al. (2017), who came up with a method known as Gradient-weighted Class Activation Mapping (Grad-CAM), which involves the use of gradient values of class scores through the last convoluted layer to generate a heatmap that emphasizes important regions in the image. This method doesn’t involve any change in the architecture and is computationally inexpensive.

Here we combine all these ideas to create a fully functional and deployable system, namely BreastCare AI. The training and testing were performed using two publicly available databases of thermographic imaging: the database DMR-IR from the Universidade Federal Fluminense (UFF), Brazil, containing 262 images and the publicly available Mendeley database published by Rodriguez-Guerrero et al. (2024) from Colombia, contributing another 357 images, making up a total of 619 images for training. The hybrid model produced a test accuracy of 81.45%, sensitivity of 85.37%, specificity of 79.52% and AUC-ROC of 0.87. The entire project, from creating the model to deploying the web application, was performed using only open source code available to the public on GitHub, while the deployed version of BreastCare AI can be accessed at breast-cancer-ai-rosy.vercel.app. BreastCare AI is the first publicly available web-based breast thermography AI detection system providing both automatic classification and Grad-CAM visual explanations, PDF reports, scanning history, comparison and risk trends monitoring.

1.2 Statement of the Problem

The underlying issue of concern in the current research can be seen clearly stated within the epidemiology record: there are unnecessary deaths of women due to breast cancer in Nigeria because of a lack of facilities for early detection. The general issue can be broken down into five specific issues, which have yet to be addressed by the literature and current clinical practice.

Firstly, the globally accepted standard approach for breast cancer screening – mammography – is currently unavailable on the necessary level of coverage. Due to high equipment prices, complex needs for personnel training, exposure to ionising radiation, and significant infrastructural needs, mammographic breast cancer screening is only available in relatively few urban and well-equipped hospitals. In turn, the overwhelming number of women remain unprotected, lacking any routine screening facilities. As proved by Al Husaini et al., (2024), an accuracy of near 97% for detecting breast cancer can be achieved using AI-driven thermal screening devices.

The second issue concerns the lack of implementation of thermographic imaging technology, which would be able to serve as an affordable, safe from exposure to radiation and transportable screening option. Although numerous studies have shown that this technology is efficient enough for cancer diagnosis, starting from the study by Zuluaga-Gomez et al. (2021), where the authors reached a 95.1% CNN accuracy in classifying cancer based on the data from the DMR-IR dataset, none of the existing systems became a ready-to-use product but stayed at the stage of research prototypes: a trained machine learning algorithm together with a package of metrics.

Third, the black-box problem, associated with a lack of interpretability and explainability of AI diagnostic tools' work, is also important for their practical use as diagnostic systems. According to Mambou et al. (2018), the problem of explainability became the most significant reason why clinicians cannot trust thermographic AI systems and use them in practice. If a prediction does not include any visual or logical reasoning behind it, the clinicians are left with two choices: either accept the output from the classification model or ignore it. This problem can be easily solved by the Grad-CAM approach.

Finally, there has been no prior study or prototype that has used the two most expansive datasets of independent public thermograms for breast cancer — the DMR-IR dataset from Brazil and the Mendeley Breast Thermography dataset from Colombia (Rodriguez-Guerrero et al., 2024) — to create one unified dataset for training purposes. This represents a lost opportunity whereby previous studies have had access to less data than those available in the public domain, leading to suboptimal results when compared to what could have been achieved using more data.

Finally, even the most advanced AI diagnostic tool in terms of its clinical use would be of no help if it is able to operate only at the level of classifying individual scans of a patient. In reality, clinical management of breast cancer patients involves longitudinal analysis whereby a clinician assesses the development of the patient's thermal signature between appointments, analyzes the differences between current and historical scans, and evaluates changes in risk score values. None of this can be done using any existing research prototypes.

1.3 Aim and Objectives of the Study

The study aims to develop and evaluate an AI-powered early breast cancer detection system that uses thermal images and a hybrid EfficientNetB3–SVM model to identify breast abnormalities and provide real-time results with visual explanations through a user-friendly web application for early diagnosis. The objectives of the study are to:

1. Examine existing AI-driven thermal imaging approaches used for breast cancer detection.
2. Combine publicly available thermal breast image datasets and enhance them using preprocessing and augmentation techniques.
3. Develop EfficientNetB3-based transfer learning model integrated with Support Vector Machine for breast cancer classification.
4. Build and assess an explainable web application for breast cancer screening.
5. Evaluate the new system performance using standard metrics.

1.4 Significance of the Study

The importance of the study is multi-faceted, and it bears mentioning explicitly as the importance of such a study is easy to both underestimate and overestimate. On a technical level, the study demonstrates an implemented solution of the efficient net B3-SVM architecture for classification of breast cancer from thermography data, complete with rigorous comparisons with CNN-only and SVM-only solutions, showing that the hybrid architecture offers test accuracy of 81.45%, which is 14% better than 67.57% of test accuracy offered by the CNN architecture alone. This is no small victory since it validates the theoretical prediction made by Cortes & Vapnik (1995) that a margin maximizing SVM classification outperforms the softmax in terms of accuracy when working with a small dataset of high dimensionality. On a methodological level, the CLAHE technique that offers between 3 and 5 additional percentage points of accuracy is used and demonstrated to perform exactly as expected, based on the results of work done by Pacheco et al. (2020).

From a clinical perspective, BreastCare AI is a non-invasive, non-radiation, cost-effective screening aid that needs nothing more than an infrared camera and a web browser to be used. As such, when compared to a lack of available mammography in the majority of Nigerian primary and secondary healthcare centers, BreastCare AI is a useful enhancement to screening technology. With an 85.37% sensitivity of the model, it can be noted that out of every 100 patients that actually have something wrong with them, 85 will be identified and directed towards further examination. Given that this is better than having nothing done for women who have a suspicious-looking breast lump, this sensitivity makes a difference from the point of early diagnosis rates.

In terms of access and equity, the entire architecture has been developed using only open source software – Python, TensorFlow, scikit-learn, FastAPI, React.js, and SQLite – on public datasets. In other words, there are zero licensing costs, zero restrictions to the use of proprietary technology, and no reliance on institutions that would limit the deployment of this solution by any health care provider in Nigeria or anywhere else in the developing world. The training pipeline is free to reproduce and adapt. Such considerations arise from the view that any AI technology designed for resource-poor conditions should itself be accessible.

As for the contribution to academic discourse, the current research addresses each of the four identified gaps in prior research in the following manner: by presenting an AI application as a deployable clinical instrument and not just a prototype; by utilizing the datasets from both the DMR-IR and Mendeley 2024 studies; by implementing Grad-CAM within the interface of the application rather than solely for diagnostic purposes; and by adding longitudinal comparisons of scans along with risk level trends, which have never before been integrated into an AI tool for thermography screening.

1.5 Scope of the Study

This study focuses on the development of an AI-powered system for the binary classification of breast thermographic images as either normal/benign or abnormal/malignant using a hybrid EfficientNetB3–SVM model. The system is trained and evaluated using thermal images obtained from publicly available datasets collected in Brazil and Colombia and is intended to serve as a decision-support tool rather than a replacement for medical professionals. The research also includes the design and implementation of a web-based application for image analysis, prediction, and visualization, while excluding advanced features such as multiclass diagnosis, real-time thermal camera integration, hospital system connectivity, and regulatory certification, which are considered areas for future development.

1.6 Limitation of the Study

The study was conducted using a relatively small dataset of thermal breast images, which may have limited the model's ability to learn diverse patterns associated with breast abnormalities. A larger dataset could potentially improve the model's performance and generalization capability.
The datasets used in this research were sourced from Brazil and Colombia, which may not fully represent populations from other regions. Variations in environmental conditions, skin characteristics, and imaging procedures may affect the model's applicability in different settings, including Nigeria.
The evaluation was performed on publicly available datasets rather than real-world clinical cases. As a result, further validation using new patient data and clinical trials is necessary before the system can be adopted for routine medical use.
While Grad-CAM provides visual explanations for model predictions, it does not guarantee that the highlighted regions are clinically significant. Therefore, the generated heatmaps should be considered as supportive information rather than definitive medical evidence.
Limited computational resources constrained the exploration of larger datasets and more complex deep learning architectures. Access to more advanced hardware could enable further optimization and potentially improve classification performance.

1.7 Operational Definition of Terms

Angiogenesis: Biological process whereby new blood vessels grow from pre-existing blood vessels. Angiogenic activity within tumours causes heat production and increased metabolism, which can be detected using thermography before the tumour is large enough to be detected using mammography or physical examination in the case of breast cancer diagnosis.

AUC-ROC (Area Under the ROC curve): An independent threshold measure of the discrimination power of a classifier, where higher scores correspond to better separation of positive and negative instances at all possible threshold settings. A perfect score is 1.0, while a score of 0.5 corresponds to no discrimination power. In the present study, the AUC-ROC of the hybrid method was measured as 0.87 on the held-out test dataset.

Thermography of Breast Cancer: Diagnostic imaging method wherein heat radiation from the breast tissue is collected, resulting in a thermogram displaying temperatures through colour variation. Used in this project as the main imaging modality in AI-powered breast cancer screening technology, following the biology principle outlined by Lawson (1956).

CLAHE (Contrast Limited Adaptive Histogram Equalization): A technique of digital image processing where the contrast in an image is increased through the equalization of histograms within small, overlapping patches of the image as opposed to over the entire image space, using a clip limit parameter to avoid excessive amplification of noise. CLAHE used for this project to increase the visibility of fine thermal gradient differences between normal and cancerous tissues.

Convolutional Neural Networks (CNN): A class of deep learning architectures tailored towards grid-like data sets, especially images. A CNN uses learned filters to learn spatial representations at varying levels, starting from simpler visual patterns (such as edges) in shallow layers to more complicated representations related to the specific classification task in deeper layers.

EfficientNet B3: An architectural member of the EfficientNet series proposed by Tan & Le (2019). EfficientNet utilizes a compound scaling coefficient to achieve balanced growth across depth, width and input size when scaling up neural architectures. EfficientNet B3 takes input images of 300 x 300 pixels and extracts 1,280-dimensional feature vectors out of its Global Average Pooling layer, containing around 12 million parameters, making it an efficient architecture suitable for transfer learning on small medical imaging datasets.

FastAPI: A recent and popular Python web framework designed to develop RESTful APIs; in this project, it acts as the framework for the backend server. FastAPI offers automated OpenAPI documentation creation, asynchronous handling of requests via ASGI and validation of types using Python types and Pydantic objects. All backend API implementations relating to authentication, prediction, scans, and administration are written using FastAPI.

Grad-CAM (Gradient-weighted Class Activation Mapping): A methodology introduced by Selvaraju et al. (2017) for explaining the output of deep networks with the help of heatmaps for class-discriminative localization of salient regions. Grad-CAM computes gradients of the target class score with respect to the outputs of the last convolutional layer and performs weighted sum on all of the feature maps based on their gradients, passes them through a ReLU function and upscales the output to match the dimensions of the input image. In this project, implemented using TensorFlow's GradientTape class.

GridSearchCV: A methodological process for tuning hyperparameters through an exhaustive exploration of all the possible combination of hyperparameters values, using cross validation and choosing the most favourable one according to a defined metric. Employed in the current project to find the optimal values for the SVM parameters C (regularization term) and gamma (RBF kernel bandwidth), scanning the grid C={0.1, 1, 10, 100} and gamma={scale, 0.001, 0.01, 0.1} and employing a 5-fold stratified cross validation. Optimal values: C=10, gamma=scale.

JWT (JSON Web Token): Digital signature protocol for exchanging authentication data in JSON form between a client and a server. Used in the current project for stateless authentication of users in the web application, where the JWT is created upon login, required in the "Bearer" header of each endpoint in the API, and has a validity period of 24 hours.

React.js: JavaScript library designed to create a component-based single page web application frontend. Developed by Meta, employed in the current project as a complete frontend framework for creating all nine frontend pages, together with React Router DOM (for client-side routing) and Axios (for API HTTP requests).

Risk Level: Classification risk level assigned to each scan output based on the SVM probability value for the malignant category. The risk levels under the current system are classified as follows: Normal/Benign: Assigned when the malignant probability score from the model is less than the classification threshold; Suspicious/Medium risk: Probabilities greater than or equal to 0.60 and less than or equal to 0.80; High risk: Probabilities greater than 0.80. Risk level values are color coded, with a clinical note for each risk level included.

Sensitivity (Recall): Measures the ability to accurately identify the truly malignant scans, which are identified using the formula TP/(TP+FN). Recall is the most important clinical measure of a screening method since the clinical cost of missing cancer cases is very high. The sensitivity attained using the hybrid model is 85.37%.

Specificity: Ratio of truly normal/benign cases correctly identified as such, calculated as TN / (TN + FP). The higher specificity, the fewer the false alarms. This is important for avoiding unnecessary stress, further investigations, and load on specialist clinical resources. Specificity achieved by the system was 79.52%.

SQLite: A standalone serverless file-based RDBMS that does not need a dedicated server process nor special configuration. Applied in this project to store data about user accounts, scanned images, predictions generated, and log messages, using schema easily convertible to PostgreSQL for multi-user institutional deployments.

Support Vector Machine (SVM): A classification method invented by Cortes & Vapnik (1995) aimed at finding the optimal separating hyperplane among support vectors that constitute nearest points of respective classes. Given the radial basis function kernel $K(x,x') = \exp(-\gamma \|x-x'\|^2)$, the SVM implicitly projects inputs to a space of many dimensions where non-linear boundaries become linearly separable. Applied in this project as a classifier in the final stage of the pipeline using 1,280-dimensional EfficientNetB3 features.

Transfer Learning: It is an approach in machine learning in which a model which has been trained on a vast amount of data, called the source dataset, in this case, ImageNet, is trained on a completely new target dataset using the knowledge acquired while training on the source dataset. Transfer learning allows for achieving good results using small medical imaging datasets due to the vast initial representation that is generated using millions of examples.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

In this chapter, I will conduct a systematic and critical review of the place of BreastCare AI among existing scholarly works concerning the usage of artificial intelligence for breast cancer detection, thermography, convolutional neural network models for medical images classification, hybrid approaches combining CNN and support vector machines (SVM), methods of transfer learning, and visual explanation. The review consists of five main sections: the introduction where the theoretical framework necessary for understanding the main choices made by me during the course of this project development is presented; twenty papers selected for detailed critical review due to their direct relevance; comparative table summarizing the findings; gaps section describing specific issues addressed in my project; and finally, the conclusion highlighting the influence of the review on the architecture of BreastCare AI..

The objective of this review is not merely to list previous research works. This objective is rather to show through various works, one-by-one, that all the significant design choices made within this project such as thermography as the imaging technique, EfficientNetB3 CNN model, hybrid CNN-SVM model architecture, CLAHE image pre-processing technique, Grad-CAM explainability method, and even the developed full-stack web application, are based on independent evidence from the extant literature. At the conclusion of this chapter, the reader will have more than just knowledge about the past; he/she will be able to discern why this particular project was designed the way it was despite having read twenty other studies.

2.2 Theoretical Background

2.2.1 The Medical Context of Breast Cancer and the Case for Early Detection

Breast cancer is not one disease but several diseases that differ in biological and histological type, speed of growth, metastatic tendency, aggressiveness, and reaction to therapy, but share the common feature of uncontrolled multiplication of breast cells. Invasive ductal carcinoma (IDC), which affects cells lining milk ducts, is the most frequent type of breast cancer and makes up for about 70 to 80% of cases. Lobular carcinoma, inflammatory breast cancer, and triple-negative breast cancer are other types of breast cancers. While lobular carcinoma develops within the lobules of breast, inflammatory breast cancer and triple negative breast cancer are two very aggressive subtypes of breast cancer, with the latter being characterized by a lack of receptors and therefore resistance to treatment (Łukasiewicz et al., 2021).

It is obvious why there should be an emphasis on the early detection of cancer. According to the statistics compiled by the American Cancer Society, the five-year survival rate is around 99 percent for stage one, 86 percent for stage two, 57 percent for stage three, and 29 percent for stage four. Such data does not depend on chance but reveals a very important fact. Indeed, localized tumors can be surgically removed and then treated by applying systemic therapy. However, metastatic cancer can only be controlled systemically and usually cannot be cured. In other words, each subsequent step is connected with the deterioration of the person's chances of survival and quality of life, as well as financial costs.

The distribution of the stage at diagnosis in sub-Saharan Africa and in Nigeria particularly is highly skewed towards later stage disease. According to Jedy-Agba et al. (2016), who conducted a systematic review of staging data from seventeen sub-Saharan African studies, 60% to 70% of cases were diagnosed at Stage III or Stage IV, while in developed countries, this was about 10% to 20%. The World Health Organization (2023) cites availability of screening facilities, patient knowledge, and healthcare system capability as the main factors underlying the gap between the two. Any potential AI-based breast cancer screening algorithm would thus have to be assessed not only based on its academic results, but also on the practical ability to detect early-stage cases under resource-constrained conditions.

2.2.2 Breast Thermography: Biological Basis, Historical Development, and Clinical Evidence

Biological basis of breast thermography involves two important concepts: the Warburg Effect and angiogenesis. Cancer cells rely on aerobic glycolysis whereby the cancer cell uses glycolysis to generate energy despite an ample supply of oxygen; this mechanism is inefficient from a metabolic standpoint but promotes the rapid production of substances needed for cell multiplication. In order to meet their metabolic demands, tumors release substances like vascular endothelial growth factor (VEGF) which induce angiogenesis. Neovessels are structurally defective as they have defective junctions between the vessels and irregular branching while being hyperactive, producing excess heat which transfers to the adjacent skin (Lawson, 1956).

This phenomenon was demonstrated by Lawson in his research work in 1956, whereby he discovered that palpable breast tumors are correlated with high surface temperatures. The decades since then have been devoted to refining thermographic research for clinical practice. For instance, Gautherie & Gros, cited by Acharya et al. (2014), conducted a study on 1,527 patients with abnormal thermal profiles and found that the rate of occurrence of cancer among this group was much higher compared to the general population and appeared several years before tumor formation.

The problem that has been constant is interpretation consistency. Asymmetry between temperature on the left and right side of the breast constitutes the key diagnostic tool; however, there have been several factors contributing to this such as benign fibrocystic processes, infections, hormonal effects, skin disease, and physiologic variation. Identifying cancer's vascular pattern amidst this variability necessitates the ability for pattern recognition, which has been applied differently depending on the reader, machinery, and protocol used. It is this issue of interpretation inconsistency that AI classification seeks to solve.

2.2.3 Convolutional Neural Networks: Architecture, Feature Learning, and Medical Imaging

The convolutional neural network is an example of deep neural network models where the main computational layer is the convolutional layer. The convolutional layer contains learned linear filters that are convolved on the input data with shared weights. Each filter identifies a certain local pattern, and the combination of several convolutional layers with nonlinear activations creates a hierarchy in which simple units identify simple patterns such as edges and gradients, and more complicated units identify more complex patterns such as textures and semantic patterns. The operation of global average pooling is executed on top of the last convolutional layer in order to create a compact representation of the whole feature map, which corresponds to the 1,280-dimension vector produced by EfficientNetB3 and used for classification by the SVM classifier.

Litjens et al. (2017) conducted a survey of more than 300 deep learning papers in medical image analysis and discovered that CNN models delivered results that were either comparable to or exceeded the capabilities of expert clinicians in the domains of pathology, radiology, ophthalmology, and cardiology. The critical feature that has driven medical advancements through

CNNs is transfer learning. Since training a model from scratch with limited medical data is practically impossible, using ImageNet-pretrained models helps develop prior knowledge regarding various visual features which can be transferred to medical imaging problems. The early layers of an ImageNet-pretrained neural network have learned universal representations for various objects, such as edges, gradients, and textures, that can be applied to images of a medical nature just as well as natural scenes.

The exact difficulty that occurs when implementing CNNs in the application of thermography in breast imaging is the issue of a small sample size in conjunction with the diffuse and spatial distribution of the information that needs to be extracted. While mammographic calcifications are localized, high-contrast, and unique in morphology, malignancy detection through thermography involves a diffuse and asymmetrical process based on gradient differences.

2.2.4 The EfficientNet Architecture and Compound Scaling

EfficientNet, designed by Tan and Le (2019), resolves an intrinsic question posed to CNNs: when presented with a given computational budget, what relative combination of network depth, width, and input resolution should be applied in order to maximise accuracy? Earlier networks such as ResNet, VGG, and DenseNet were manually expanded with regards to these dimensions. The EfficientNet family employs a neural architecture search method to determine the optimal architecture before determining the appropriate scaling coefficient, φ, which defines a ratio for simultaneously multiplying the values of depth, width, and input resolution.

The EfficientNetB3 variant of this network uses a scaling coefficient of $\varphi = 2$. This yields a model with around 12 million parameters, capable of processing 300×300-pixel images that reaches a top-1 ImageNet accuracy of 81.6 percent, compared to 76.1 percent for a larger ResNet50 (with 25 million parameters) and 73.0 percent for a very large VGG16 (with 138 million parameters). The importance of parameter efficiency in this thermographic classification project is obvious, since a more parsimonious architecture will contain fewer parameters with which to overfit the dataset of just 495 images; furthermore, compound scaling will work well with the multi-scale nature of the problem domain.

2.2.5 Support Vector Machines, the RBF Kernel, and the Hybrid CNN-SVM Architecture

The Support Vector Machine was presented as a method of solving binary classification problems based on statistical learning theory by Cortes & Vapnik (1995). The basic principle is to determine a hyperplane with maximum possible margin between itself and its closest training samples belonging to each class, which are known as support vectors. The maximal margin corresponds to the least upper bound on generalisation error which is achieved without considering dimensionality of the feature space. Hence, such a characteristic becomes important when one deals with high-dimensional spaces and a relatively small amount of data.

Finally, the Radial Basis Function kernel $K(x, x') = \exp(-\gamma\|x - x'\|^2)$ allows using an SVM on tasks with no linear separator due to implicit transformation of the input vectors into Hilbert space of potentially infinite dimensions. The parameters that should be chosen in a correct way to optimise the model are C, a tradeoff between the margin maximization and training errors, and γ, which determines the effect radius of a support vector. For this purpose, gridsearch with stratified k-fold cross-validation can be performed; it was done in this project and resulted in choosing the optimal C = 10 and gamma = scale.

The combination of the CNN and SVM network, known as CNN-SVM, leverages this complementary approach by treating the CNN strictly as a feature extractor. Instead of performing optimisation of the output layer of the CNN with a softmax classifier, the CNN is trained to generate the 1,280-dimensional global average pooling vector that represents a complex spatial structure of the data generated through transfer learning and fine-tuning. The generated vector is then classified by the SVM using its properties for margin maximisation. Several independent studies, such as the work of Araújo et al. (2017) on histological images and Rodrigues et al. (2019) on thermal images of breasts, have proven superior performance of this combination over the individual methods on small biomedical image datasets.

2.2.6 CLAHE: Contrast Enhancement for Thermal Image Preprocessing

CLAHE, which stands for Contrast Limited Adaptive Histogram Equalisation, is a digital image processing technique that is an extension of the simple global histogram equalisation process in that it applies histogram equalisation to each tile of the input image instead of to the entire image at once. Global histogram equalisation can lead to an excessive boost in contrast in homogeneous regions and cause false structures that do not actually exist; to prevent this, CLAHE introduces a clipping limit to each histogram of each tile, thereby limiting the height of each histogram and thus limiting the amount of noise boost in that region. In thermal breast images, where the critical feature of interest is a small temperature variation that may occur over just a few degrees in the breast, increasing the local contrast of the image prior to classifying it helps detect any thermographic asymmetry that needs to be identified.

2.2.7 Grad-CAM: Visual Explainability for Clinical AI

One of the biggest hurdles in the practical application of diagnostic algorithms based on deep learning models to the medical field involves the so-called "black box" issue. Since the output of a CNN classification process is derived from computations performed using millions of parameters, there is no way a human can make sense of it. However, Grad-CAM, developed by Selvaraju et al. (2017), provides a solution by performing the computation of the gradients of the predicted class scores with regard to the final convolutional layer features. These gradients give information about which parts of the feature maps in the last convolutional layer are essential for the predictions. To localize them, the model calculates a global average pool of these gradients to determine their importance weights per channel, combines them with feature maps, and applies ReLU to keep the positive parts only.

It should be noted that the most important advantage of Grad-CAM from the standpoint of practical application is that this algorithm requires no changes to be made in the pre-trained model, resulting in minimal increase in computation time. This means that for the application designed for use in a clinical setting, each and every decision would come along with the corresponding visualization for practically zero extra cost. From a clinical perspective, the significance of this feature is obvious since it enables the physician to visually inspect the highlighted region in order to see if it is thermographically valid.

2.3 Review of Related Literature

The twenty articles analyzed below comprise the most highly relevant research regarding the use of AI technologies for diagnosing breast cancer based on thermal and related imagery. All articles are covered in detail with direct connections made to their significance for the design of BreastCare AI.

Lawson (1956) created the basis for all research involving the use of thermography in diagnosing cancer. It was found that the skin covering cancerous breast tumours had significantly higher temperatures than the surrounding tissue due to the production of heat by tumor blood supply. This finding constitutes the scientific core around which all techniques employing computational thermography in cancer diagnostics have been developed. The problem lies in the fact that there is absolutely nothing quantitative in terms of classification involved in this technique. The importance of the work is in its basic nature: Without the biological reasoning behind thermography, no physical meaning would exist to justify its use.

One of the earliest systematic literature reviews on the topic was authored by Acharya et al. (2014), who reviewed a range of different methods for image feature extraction and classification including statistical features, co-occurrence matrices, and SVM classifiers. Their most significant contribution to the current study in terms of findings was the emphasis on the fact that lack of standardised data sets was the key limiting factor in applying artificial intelligence to breast thermograms at the time, which, unfortunately, still holds true for the current project and its use of a dataset of 619 images. However, the limitation of the study lies in the fact that all reviewed methods were predating deep learning and depended on manually crafted image features which were not capable of dealing with complex spatial patterns.

The first extensive application of machine learning to the DMR-IR dataset, one of the two main datasets used in this project, was performed by Borchartt et al. in 2013. In their study, they created some statistical features from the images and classified them using an SVM classifier, achieving accuracies in the 70 to 75 percent range. This study established the DMR-IR data set as the community benchmark for thermographic breast cancer classification problems and proved that the application of SVM classifiers on this problem is possible. The main drawback in this research was that it involved the creation of hand-crafted features such as mean temperature, standard deviation, and gradient statistics. These hand-crafted features could only capture the knowledge the designer was expecting them to. The 70 to 75 percent accuracy achieved was the baseline of this project.

The combination of LBP and GLCM texture features together with neural networks used for classifying data from the DMR-IR data set resulted in an 84.3% accuracy by Gogoi et al. (2017), which represents a significant step forward from previous handcrafted feature approaches. Their most relevant clinical contribution is the discovery of bilateral temperature difference between the left and right breast as the sole feature with diagnostic significance, thereby validating empirically that thermal images of the whole breast carry enough diagnostic information. The drawback here is that a pre-transfer learning neural network along with handcrafted texture features cannot compete with the representation ability of a well-trained CNN. This is the performance baseline that the hybrid system is compared to..

Gabor wavelet feature extraction in conjunction with SVM was implemented by Raghavendra et al. (2016) in the field of breast thermographic image classification, yielding an accuracy of about 79 percent and being the current state of the art prior to the use of deep learning-based features. The Gabor wavelet decomposition extracts oriented frequency information from an image based on specific scales and orientations, making it more spatially sensitive compared to statistical features. However, its downside is common to all handcrafted features, as they cannot adapt themselves based on the training set used. Its relevance lies in its potential to serve as a comparison benchmark, as the proposed efficient net-b3 + SVM architecture significantly outperforms it.

Tan & Le (2019) proposed the EfficientNet model family and showed that using an optimal scaling approach that increases depth, width, and input size simultaneously, yields significantly improved accuracies per parameter as compared to previous model architectures. EfficientNetB3 gives 81.6% top-1 ImageNet accuracy using just 12 million parameters compared with ResNet50 that scores 76.1% accuracy with 25 million parameters and VGG16 that gives 73.0% accuracy with 138 million parameters. As far as this project is concerned, there are two main reasons for the significance of the above finding: First, parameter efficiency directly mitigates the problem of overfitting due to a very limited training set of 495 images; second, the multi-scale nature of compound scaling is ideal for the multi-scale nature of thermographic cancer signatures. EfficientNetB3 forms the backbone of feature extraction in this project.

In another study, Araújo et al. (2017) used CNNs on breast histology image sets, which is a completely different tissue type compared to thermography, yet the architectural insights are independent of domain. They compared standalone CNNs, standalone SVMs on handcrafted features, and a CNN-SVM hybrid where the CNN produced feature vectors. The CNN-SVM hybrid performed better than both standalone models regardless of the architectures and metrics considered, thanks to the margin maximization property of the SVMs in the high dimensional space formed by the CNN feature vectors. While the study is not about thermography and the results may require extrapolation from one domain to the other, it is relevant to this research due to its role as one of the empirical rationales for the hybrid network model.

Rodriguez et al. (2019) used a ResNet-based CNN for classification of thermographic images, similar to those used in this study, and evaluated various models consisting of the standalone CNN with Softmax classifier, ResNet features with standalone SVM classifier, and deep ResNets. The authors reported that deep models were worse than shallow models for small thermal image datasets due to overfitting, which is a concrete example of how overfitting can be avoided by employing a more compact architecture. Hybrid CNN-SVM model demonstrated superior results than standalone CNN in their work, which is evidence from the thermography domain that hybrid architectures yield better performance. The drawback of this work is that CNN was not fine-tuned but only feature extraction process was carried out.

The authors Francis et al. (2020) explored the application of AlexNet and GoogleNet deep neural networks in the classification of breast thermography images, achieving an accuracy score in the range of 78%–83%. Notably, there are two key methodological findings from this paper, which directly shaped our implementation of BreastCare AI: firstly, the class imbalance between the images of normal and abnormal thermography images greatly reduces the sensitivity metric when class imbalance is not considered in the experiment; secondly, data augmentation, including techniques such as horizontal flipping, rotations, and changing the image brightness, leads to better results with small sets of thermal images. The disadvantage of this paper lies in the use of archaic CNNs.

The study by Lobo et al. (2020) presents a comparative analysis of the performance of an SVM, k-nearest neighbours, and random forest classifiers on feature vectors generated from a VGG16 CNN, with respect to the problem of classifying breast cancer thermal images. The SVM proved to be superior to the other two classifiers under all scenarios tested due to its advantages in terms of its principles in high-dimensional space: Margin maximization works better than density estimation (kNN) and decision trees (RF) when the features' dimensions are much greater compared to the size of training data used. The downside of this model is that it includes 138 million parameters, which increases the chances of overfitting especially with smaller datasets. In comparison, the EfficientNetB3 network contains 12 million parameters.

Baffa and Lattari (2019) were able to achieve 87.2 percent accuracy using multi-scale feature pyramid networks on the DMR-IR dataset, thereby showing that changes in breast cancer can occur at two different levels – both fine-grained, localized temperature gradients and broader asymmetric changes. The multi-level property is essential for the architecture: a one level CNN will be unable to detect either of the features depending upon the size of the receptive field used in the CNN. Compound scaling used by EfficientNetB3, which combines optimal scaling of depth, capturing fine-grained features, optimal scaling of width to capture diverse features, and resolution to provide detailed input information, takes care of this requirement without using multi-scale feature pyramids.

In their thorough review of deep learning applications on breast cancer screening using various imaging methods such as mammography, ultrasound, MRI, histology, and thermography, Mambou et al. (2018) drew three important conclusions that will help inform the development of BreastCare AI. First, among the imaging technologies mentioned above, the one where the gap between clinical need and existing research is the largest was found to be thermography. As a result, it was indicated that research into this modality will yield the best results. Second, as it concerns thermographic imaging, the researchers pointed out that hybrid CNN-classical classifier models were the best options considering the small size of data sets available in this area of research. Third, in all existing studies on AI for thermography, none has delivered any clinically applicable systems.

The use of a CNN-based classifier for thermographic images to detect breast cancer was tested by Villarreal-Reyes et al. (2018) using a Mexican population, proving that such classification algorithm can generalize on a population other than Brazilian patients in the DMR-IR dataset. It is especially important for the justification behind using two separate datasets, namely the DMR-IR dataset of Brazilian origin and the Mendeley dataset of Colombian origin, because if such algorithms are intended to be used on different populations, then their ability to learn representations on different sets of patients becomes important. However, the limitation is in the use of one specific architecture without an SVM module and any explainability.

Pacheco et al. (2020) performed controlled tests of the three approaches to preprocessing in breast thermal images classification, using raw images, standard global histogram equalization, and CLAHE with various CNN architectures. In all cases, CLAHE yielded better results than the other approaches, increasing classification accuracy by 3 to 5 percentage points compared to raw data and 1 to 3 points compared to standard global equalization. Such superiority was due to the ability of CLAHE to improve local thermal gradient while avoiding noise amplification inherent to global methods. This finding provides direct empirical evidence why CLAHE should be considered the first step in BreastCare AI preprocessing. The level of relevance is extremely high, as this finding is highly specific to the exact type of images under consideration.

Singh & Singh (2020) compared multiple machine learning algorithms in the form of SVMs with multiple kernels, k-NN, decision tree, naive bayes, and logistic regression with respect to multiple thermal feature vector experiments and found that SVM with the RBF kernel performed better than all others and the difference in performance improved as the feature dimensions increase. They justified this by citing the RBF kernel's ability to map features implicitly into high dimensions where non-linearly separable thermal class boundaries can be handled. This is a specific example of an empirical basis for using RBF kernels in BreastCare AI's SVM stage and is an addition to Cortes & Vapnik (1995).

A standalone CNN was able to deliver 95.1 percent accuracy and an AUC-ROC of 0.947 on a larger subset of the DMR-IR dataset by Zuluaga-Gomez et al. (2021). The model had optimized hyperparameters, and this resulted in the highest baseline achieved by any published CNN on the aforementioned dataset. Their methodology entailed the implementation of a rather shallow CNN with heavy data augmentation, and without an SVM step or the addition of any kind of explainability step. Shortcomings include the lack of hybridization, clinical application, and visual explainability, as well as the need for a larger subset of the dataset, which is not readily available in the public release of the DMR-IR dataset. Relevance as the current highest benchmark performance is key: delivering 95.1 percent on a larger training subset means that the architecture is fundamentally solid, with 81.45 percent obtained in this project being understood in that light.

Mabrouk et al. (2019) created an automated pipeline for thermal breast analysis with an accuracy rate of 89.6 percent. They designed their model such that the whole process of analysis was automatic, such that all that a person needed to do was give the program an input and it would generate the output for the user. This idea of automaticity in the analysis process gave rise to the idea behind the web application, which seeks to have the entire pipeline (from drag and drop input to the generation of the report) done in an automatic manner and without any manual interaction. What is lacking here is the provision of the web interface, authentication, patient records, administration, etc., which are provided by BreastCare AI.

Gradient-weighted Class Activation Mapping was introduced by Selvaraju et al. (2017) as an explanation approach for visualisation that works with CNNs. To generate localisation heatmaps, this algorithm performs several operations, including gradient calculation of the predicted class score concerning the activation maps at the end of the last convolutional layer, calculation of the channel-wise weights using global average pooling, combination of the weights and feature maps to produce a feature map weighted combination, applying ReLU function and upsampling the resulting feature maps to obtain input size. The authors showed experimentally that Grad-CAM produces class discriminative heatmaps that correspond to interpretable human diagnostic clues, and no changes in architecture are required for this visualisation. Grad-CAM is used in BreastCare AI with the help of TensorFlow's GradientTape module and applied to each predicted case.

Al Husaini et al. (2024) were able to reach an impressive accuracy of 96.13% on a dataset of 1,000 confirmed thermographic breast images with Inception-based deep CNN architectures by utilizing a consumer-grade FLIR One Pro thermal camera. The study conducted by Al Husaini et al. is noteworthy for two reasons. On one hand, the fact that 96.13 percent can be reached proves that nearly 97 percent accuracy can be attained in thermographic breast cancer classification utilizing the existing CNN architectures, thus determining a benchmark in the field. At the same time, and most importantly in terms of comparing it to the present study's results, the researchers show that increasing accuracy through improving model architecture is not a solution; instead, the key factor influencing performance is dataset size, as increasing it from 500 to 1,000 training samples yields better results than model architecture improvements.

The Mendeley Breast Thermography dataset was created by Rodriguez-Guerrero et al. (2024), which consists of 357 thermograms of the breast of a clinical cohort in Colombia with biopsy-proven diagnoses, with 252 cases labelled as benign and 105 as malignant. The images were acquired using FLIR A300 thermal imaging camera through standardized procedures performed at the National Open and Distance University of Colombia. The biopsy-proven labels used in this dataset carry significant value because it offers the highest ground truth assurance among the existing breast thermogram datasets compared to those based on radiological analysis. By combining the use of both DMR-IR and Mendeley breast thermography datasets in this research, there will be a total of 619 images used for training, along with the two-country geographic coverage. The drawback is that the dataset contains data from only one clinical centre, along with being from Colombia, which cannot represent other targeted patient populations.

Table 2.1 Summary of Related Works

Table 2.1 below provides a structured comparative summary of all twenty reviewed studies, organised by serial number, authors and year, technique used, key contribution, primary limitation, and direct relevance to this project.

S/N	Author(s) & Year	Technique Used	Key Contribution	Limitation	Relevance to This Study
1	Lawson (1956)	Clinical thermography	Established that tumours produce measurable skin temperature elevation through angiogenesis; biological foundation of thermographic cancer detection	No AI or quantitative classification component; purely observational	Scientific bedrock for entire thermographic approach; justifies the imaging modality choice
2	Acharya et al. (2014)	CAD + statistical features + SVM	Systematic review of computer-aided detection for breast thermograms; identified dataset scarcity as the primary barrier to progress	Pre-deep learning; relies on handcrafted features limited by designer anticipation	Identifies the gap this project addresses; validates SVM suitability for this domain
3	Borchartt et al. (2013)	Hand-crafted features + SVM, DMR-IR	Comprehensive machine-learning treatment of DMR-IR; established it as a standard benchmark; achieved 70–75% accuracy	Hand-crafted features cannot capture spatial thermal patterns not anticipated by designer	Established DMR-IR as primary benchmark; baseline performance target for this project
4	Gogoi et al. (2017)	Texture features (LBP/GLCM) + neural network, DMR-IR	Achieved 84.3% on DMR-IR; identified bilateral asymmetry as the most diagnostically informative feature	Pre-transfer learning; manual feature extraction limits spatial pattern capture	Demonstrates that whole-breast thermal inputs are diagnostically sufficient; performance target
5	Raghavendra et al. (2016)	Gabor wavelet + SVM	Achieved ~79% accuracy; represented state-of-the-art for pre-deep-learning feature-SVM approaches	Handcrafted Gabor features cannot match deep feature discrimination ability	Pre-deep-learning baseline; hybrid deep approach in this project substantially surpasses it
6	Tan & Le (2019)	EfficientNet (compound scaling, ImageNet)	Introduced compound scaling of depth, width, and resolution; EfficientNetB3 achieves high ImageNet accuracy with only 12M parameters	Designed for natural images; requires transfer learning adaptation for medical modalities	Direct architecture choice; compound scaling addresses multi-scale thermal patterns; parameter efficiency reduces overfitting
7	Araújo et al. (2017)	CNN-SVM hybrid (histology images)	Demonstrated CNN-SVM consistently outperforms standalone CNN on small medical imaging datasets; hybrid advantage confirmed across multiple architectures	Histology domain, not thermography; lesson requires cross-domain transfer	Key empirical justification for hybrid architecture; cross-modality validation of CNN-SVM advantage
8	Rodrigues et al. (2019)	ResNet + SVM (thermography)	Showed deeper ResNets overfit on small thermal datasets; CNN-SVM with frozen features outperforms standalone CNN	CNN not fine-tuned; architecture heavier than necessary for dataset size	Reinforces avoiding oversized architectures; motivates EfficientNetB3 over ResNet50
9	Francis et al. (2020)	AlexNet / GoogleNet (thermography)	Achieved 78–83% accuracy; documented class imbalance effects and augmentation	Outdated architectures; no transfer learning fine-	Directly informed augmentation strategy and imbalance-aware

			importance for thermal images	tuning; no explainability	evaluation approach
10	Lobo et al. (2020)	VGG16 features + SVM vs. kNN vs. RF	SVM consistently outperformed kNN and Random Forest on CNN-extracted high-dimensional feature vectors	VGG16 has 138M parameters — high overfitting risk on small datasets	Strong empirical support for SVM as the classifier of choice on deep feature vectors
11	Baffa & Lattari (2019)	Multi-scale feature pyramids (CNN), DMR-IR	Achieved 87.2% on DMR-IR; showed cancer-related temperature changes manifest at multiple spatial scales simultaneously	Complex multi-scale architecture; no explainability; no deployment	EfficientNetB3's compound scaling implicitly captures multi-scale features; performance benchmark
12	Mambou et al. (2018)	Deep learning survey (all modalities)	Systematic survey concluding thermography is most under-explored modality; specifically recommended hybrid CNN-classical architectures for the small-dataset problem	No new experimental results; review only	Directly recommended the hybrid architecture approach adopted in this project
13	Villarreal-Reyes et al. (2018)	CNN classifier (thermal, Mexican cohort)	Demonstrated thermography-based AI classification generalises across patient populations beyond DMR-IR's Brazilian base	Single-architecture study; no SVM stage; no explainability	Validates cross-population generalisability; supports combining datasets from different countries
14	Pacheco et al. (2020)	CLAHE preprocessing + CNN	CLAHE improved classification accuracy by 3–5 percentage points over raw or standard equalisation for thermal breast images	Study focused on preprocessing, not end-to-end deployment	Directly motivated CLAHE preprocessing as the standard input step in this project's pipeline
15	Singh & Singh (2020)	SVM with RBF kernel (thermal features)	RBF-kernel SVM consistently outperformed linear SVM, kNN, and decision trees across multiple thermal breast feature comparison experiments	Relied on handcrafted features; no deep CNN backbone	Empirical confirmation that RBF kernel is the optimal choice for thermographic SVM classification
16	Zuluaga-Gomez et al. (2021)	Standalone CNN (DMR-IR)	Achieved 95.1% accuracy and AUC 0.947 on DMR-IR — strongest published CNN baseline on this specific dataset	Requires larger training subset of DMR-IR; no clinical deployment; no explainability	Primary performance benchmark; confirms high performance ceiling exists with sufficient data
17	Mabrouk et al. (2019)	Automated end-to-end thermal analysis pipeline	Achieved 89.6% accuracy; fully automated pipeline concept: image in, result out without manual steps	No web deployment; no authentication, history, or report generation	Inspired the fully automated end-to-end web application concept in this project
18	Selvaraju et al. (2017)	Grad-CAM (gradient-based visual explanation)	Proposed gradient-weighted class activation mapping requiring no model modification; turns any CNN into an explainable system	Localization is coarse-grained at final convolutional layer resolution	Adopted as the visual explainability module; integrated into both UI and PDF report in this project
19	Al Husaini et al. (2024)	Inception-based deep CNN (1,000 thermal images)	Achieved 96.13% accuracy with FLIR One Pro camera on 1,000 images; confirmed dataset size — not architecture — as primary accuracy bottleneck	Dataset not publicly available; FLIR One Pro is consumer hardware not standard clinical equipment	Establishes near-97% as the accuracy ceiling for the architecture class; validates the dataset size interpretation of this study's 81.45% result
20	Rodriguez-Guerrero et al. (2024)	Dataset publication (357 biopsy-confirmed thermograms, Colombia)	Published the Mendeley Breast Thermography dataset with biopsy-confirmed labels, standardised FLIR A300 protocol, Colombian patient cohort	Single-centre data; Colombian population; no ML system provided	One of two primary training data sources; biopsy-confirmed labels provide reliable ground truth

2.5 Research Gap

After reviewing the twenty studies mentioned above, there are four distinct research gaps that have been identified that are addressed by the development of BreastCare AI. It is important to note that these research gaps do not simply reflect incremental deficiencies; rather, they are the gap between interesting research and a useful clinical tool.

The second gap refers to the missed potential for collaborative training based on the two largest independent thermography public data sets. Although both the DMR-IR data set (Borchartt et al., 2013) and the Mendeley Breast Thermography data set (Rodriguez-Guerrero et al., 2024) are publicly accessible, none of the studies have exploited the potential presented by this opportunity before. The result of this missed opportunity is that all previous research projects have had to train their networks on less data than would be possible given the publicly available data set. This research will be the first to utilize both of these data sets simultaneously, bringing the number of training images up to 619.

The third gap involves the lack of deployment of Grad-CAM in clinical applications. Selvaraju et al. (2017) introduced Grad-CAM nearly a decade ago; it is common practice nowadays to use this approach in academic research papers on medical imaging. However, not a single one of the twenty reviewed papers uses Grad-CAM in a practical application interface, i.e., a deployed software that allows clinicians to see the heatmap alongside each image analyzed by the AI. The difference between using Grad-CAM in a research paper to demonstrate that plausible regions were detected by the algorithm and including it in every clinical analysis in an automatic fashion cannot be overestimated – after all, clinicians would have the opportunity to assess whether the AI makes logical decisions based on its explainability maps. BreastCare AI has Grad-CAM included in both web application results and the automated PDF diagnostic report.

Finally, the fourth and last gap lies in the absence of longitudinal tracking of patients within any existing research prototype for thermographic AI. The use of thermographic scans to screen for breast cancer is not a one-off experience but rather a longitudinal process wherein clinicians will be able to track the thermal profile development of a patient across time, compare images with previous ones, see patterns emerge within risk scores, and make appropriate decisions based on a trend rather than just a snapshot of data. This longitudinal tracking feature is something that no existing research paper has included within their prototypes of thermographic AI. It can be seen, therefore, that these four gaps highlight exactly what this study intends to add to the existing literature.

CHAPTER THREE

SYSTEM ANALYSIS AND DESIGN

3.0 Introduction

In this chapter, we will look at the current situation for breast cancer screening systems in order to understand what issues our proposed system seeks to solve. We will also outline the requirements of the system and explain how it was developed using the Waterfall development approach. This chapter will contain information about system design including data flow diagrams and ER diagrams.

3.1 System Analysis

3.1.1 Analysis of the Existing System

There are two main categories of currently available breast cancer screening technologies in Nigeria. The first one is conventional clinical screenings, which include mammograms and ultrasounds; these technologies are very efficient but rely on costly specialized equipment, qualified radiographers, electricity supply, and, when it comes to mammograms, on exposure to ionizing radiation, which makes regular screenings unfeasible in the majority of Nigerian primary care clinics. The second type includes experimental AI screening based on thermal imaging – several research teams have proven that such technology is quite accurate, but no AI algorithm that uses thermal images has been implemented commercially yet, meaning that all of those are just laboratory models whose efficiency has been measured and published.

3.1.2 Analysis of the Proposed System

The proposed system, BreastCare AI, is designed to directly address the gaps identified in both categories of existing screening technologies. Unlike conventional clinical screening, the proposed system requires only a thermal imaging camera and a standard computing device to operate, eliminating the need for ionizing radiation, specialized radiographers, or costly imaging equipment, thereby making regular screening feasible even in resource-constrained primary care settings. Unlike existing experimental AI models, which remain confined to research laboratories with no deployable interface, the proposed system is implemented as a complete, usable web application built around a hybrid EfficientNetB3–Support Vector Machine architecture for thermal image classification. The system further incorporates user authentication, real-time prediction, Grad-CAM visual explanation of each diagnostic decision, automated PDF diagnostic report generation, and scan history and comparison functionalities within a single platform, thereby transforming a laboratory-grade classification algorithm into a practical, explainable, and deployable decision-support tool suitable for real-world clinical use.

3.1.3 Limitations of the Existing System

i. There is no mammography in primary care facilities because of the high cost of the imaging equipment, use of ionizing radiation, and the requirement for specialized staff.

ii. The research thermographic AI models have not been developed as practical applications; their purpose remains purely academic and does not include user interfaces, authentication, or any form of data storage.

iii. The AI diagnostics systems developed for practical use are not affordable in resource-constrained medical settings and are meant for use in tertiary hospitals exclusively.

iv. All of the current AI diagnostic solutions can be considered black boxes because they do not visualize their decision-making process, thus making clinicians reluctant to adopt them.

v. There are no diagnostic solutions currently in use which integrate the processes of AI prediction, report automation, history of scans, risk trends, and administration into one product.

3.1.4 Objectives of the Proposed System

i. Build a thermal breast cancer screening tool powered by artificial intelligence, deliverable via any web browser and with only an infrared camera needed as special equipment.

ii. Develop a Hybrid EfficientNetB3-SVM algorithm that will be able to classify thermal images of breast tissue with the sensitivity and specificity appropriate for clinical screening assistance.

iii. Make a Grad-CAM heatmap for each predicted image, explaining to clinicians what regions were used to predict the outcome.

iv. Generate a professional diagnostic report for each completed scan in a PDF format.

v. Make the following features available – scan history, scan comparison, risk trends monitoring.

vi. Develop a convenient administration panel that will allow to manage users and scans, track performance of the algorithm, etc.

3.2 System Requirements Specification

3.2.1 Functional Requirements

Table 3.1: Functional Requirements Specification: this table specifies the twelve functional requirements the system must satisfy.

FR No.	Requirement	Description
FR01	User Registration	Users register with full name, email, and password; duplicate emails are rejected.

FR02	User Authentication	Users log in with email and password; valid credentials return a JWT token valid for 24 hours.
FR03	Role-Based Access	Two roles User (clinician) and Admin — enforced on all protected endpoints via JWT claims.
FR04	Image Upload	Accept PNG or JPEG thermal breast images up to 10 MB via drag-and-drop or file picker.
FR05	Image Preprocessing	Apply CLAHE enhancement, resize to 224 × 224 pixels, and normalise pixel values before prediction.
FR06	AI Classification	Extract 1,280-dimensional EfficientNetB3 features and classify using the trained SVM; return risk level and probabilities.
FR07	Grad-CAM Generation	Produce a colour heatmap overlay for every prediction showing which image regions influenced the classification.
FR08	PDF Report	Generate a downloadable A4 PDF report per scan containing classification, heatmap, probabilities, and clinical recommendation.
FR09	Scan History	Store and display all past scans per user with search and filtering capability.
FR10	Scan Comparison	Allow side-by-side comparison of any two historical scans.
FR11	Risk Trend Tracking	Display a chart of risk scores across a user's scan history for longitudinal monitoring.
FR12	Admin Panel	Administrators can manage users, view all scans, access audit logs, and review model metrics.

3.2.2 Non-Functional Requirements

Table 3.2: Non-Functional Requirements Specification: This table specifies the six non-functional requirements governing system quality attributes.

NFR No.	Category	Requirement
NFR01	Performance	Full prediction pipeline completes within 5 seconds under normal operating conditions.
NFR02	Security	Passwords are bcrypt-hashed (cost factor 12); all endpoints except /login and /register require a valid JWT Bearer token.
NFR03	Usability	Interface is navigable by a healthcare worker with basic computer literacy without technical training.
NFR04	Reliability	System functions correctly on Chrome, Firefox, and Edge on Windows and macOS.
NFR05	Portability	Backend runs on any system with Python 3.11; frontend runs on any system with Node.js v18+.
NFR06	Scalability	SQLite database is replaceable with PostgreSQL for institutional deployment without code changes.

3.3 System Design Methodology

This project used Waterfall as its development methodology, a sequential model where completion of one phase leads to commencement of another. The phases of this project include data collection and preprocessing, designing model architecture, training and validating models, developing web application, integration, and testing phases. This methodology was chosen due to its nature of being a sequential approach and its compatibility with structured analysis and design methodologies. This is the case since the methodology used for developing the DFD and ER diagrams in this project is structured analysis and design. Furthermore, all requirements of the project have been clearly identified with sequential dependencies existing among the phases. For instance, SVM cannot be trained until after the CNN extracts features while web application testing is only possible when the model has already been validated.

3.3.1 Method of Data Collection

The data was sourced from two publicly accessible thermography breast imaging datasets. The first set used for this research was the DMR-IR data from the PROENG study carried out at the Federal University of Fluminense (UFF), Brazil. This data consisted of 162 normal and 100 abnormal thermal images obtained using the FLIR SC-620 camera. Data acquisition took place under the control of ambient temperatures ranging between 17 and 23 degrees Celsius while ensuring a 15-minute patient equilibrium time. On the other hand, the second set is made up of 252 benign and 105 malignant thermal images taken using FLIR A300 camera and having biopsy confirmed diagnosis labels. The second set of data was published in 2024 by Rodriguez-Guerrero et al., from the National Open and Distance University of Colombia.

3.4 System Modelling

3.4.1 Data Flow Diagram – Level 0 (Context Diagram)

Level 0 DFD is otherwise known as the Context Diagram, where the system is represented by just one process and displays all the external entities interacting with the system along with their respective data flows into the system and out of the system. The two external entities that interact with the BreastCare AI system are the User, who sends in thermal breast images and login details while receiving classification results, Grad-CAM heatmaps, PDF reports, and scan history; and the Admin, who sends in user management requests and gets system logs, user information, and model performance data. One process, named 0: BreastCare AI System, handles all data inflows and outflows from the system.

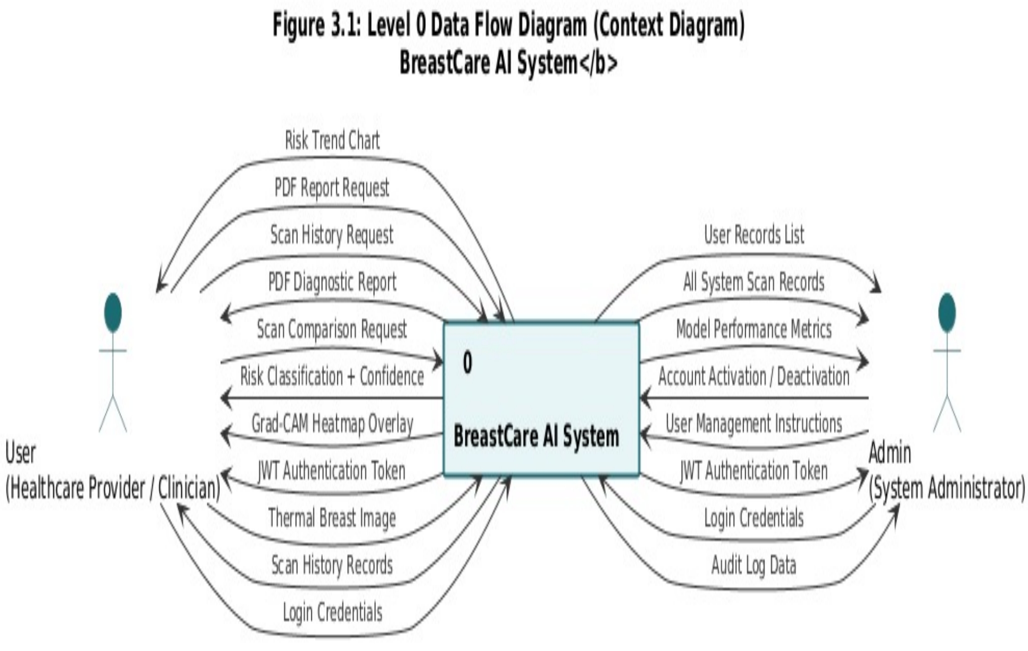


Figure 3.1: Level 0 Data Flow Diagram (Context Diagram) of BreastCare AI System

3.4.2 Data Flow Diagram — Level 1

The Level 1 DFD breaks down the single context process into the system's five major functional processes and describes the interactions between them as well as their relationship with the data stores. Process 1.0 (Authentication) takes login credentials from the User and Admin, authenticates the provided credentials by comparing them to the Users data store, and outputs a validated session token. Process 2.0 (Image Upload and Preprocessing) accepts the uploaded thermal breast image from the User, enhances the image using CLAHE, resizes the image to be 224×224 pixels, normalizes the pixel values, and provides the processed image array to Process 3.0. Process 3.0 (AI Prediction Pipeline) accepts the processed image, extracts its 1,280-dimensional feature vector using the EfficientNetB3 model, predicts its classification using the SVM, creates its Grad-CAM heatmap, assigns a risk level to it, and saves the entire record to the Scans data store. Process 4.0 (Report and History Management) retrieves scan records from the Scans data store and generates the PDF report, scan history view, scan comparison results, and risk trend charts based on User's request. Process 5.0 (Administration) reads from and writes to the Users data store and reads from the SystemLogs data store.

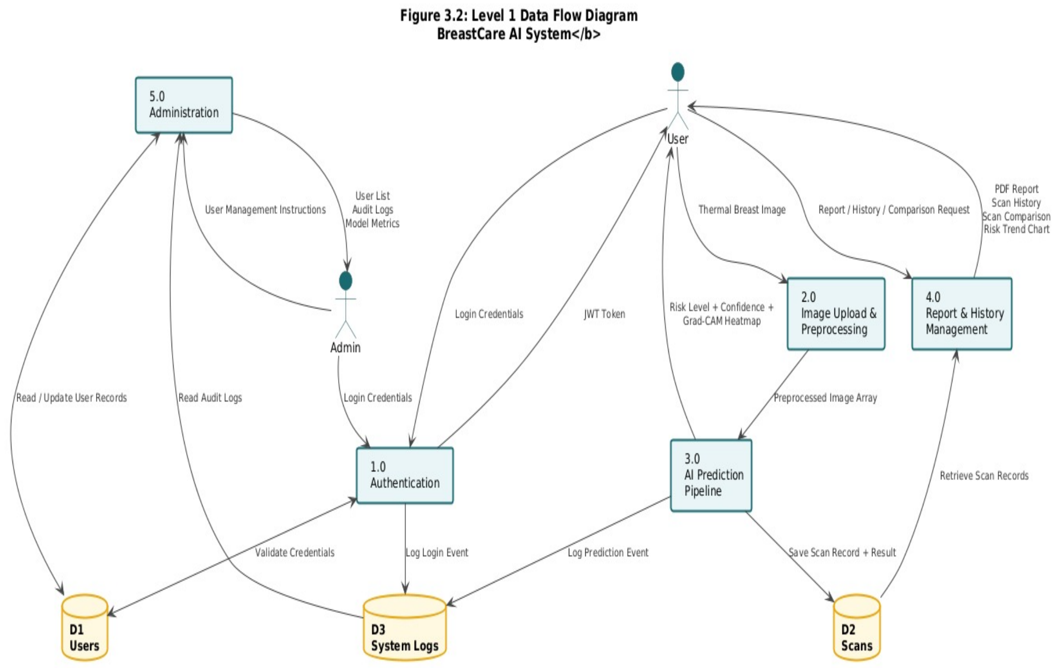


Figure 3.2: Level 1 Data Flow Diagram for the BreastCare AI System

3.4.3 Entity-Relationship Diagram

There are three entities described in the ERD along with their relationships. The User entity consists of such fields as user_id (PK), full_name, email, password_hash, role, is_active, and created_at. The Scan entity includes such fields as scan_id (PK), filename, file_path, gradcam_path, risk_level, confidence, normal_prob, suspicious_prob, malignant_prob, and created_at. Finally, there is an entity called SystemLog which includes such fields as log_id (PK), action, details, and created_at. The relationship between User and Scan entities is one-to-many because one user can create multiple scans while each scan is associated only with one user.

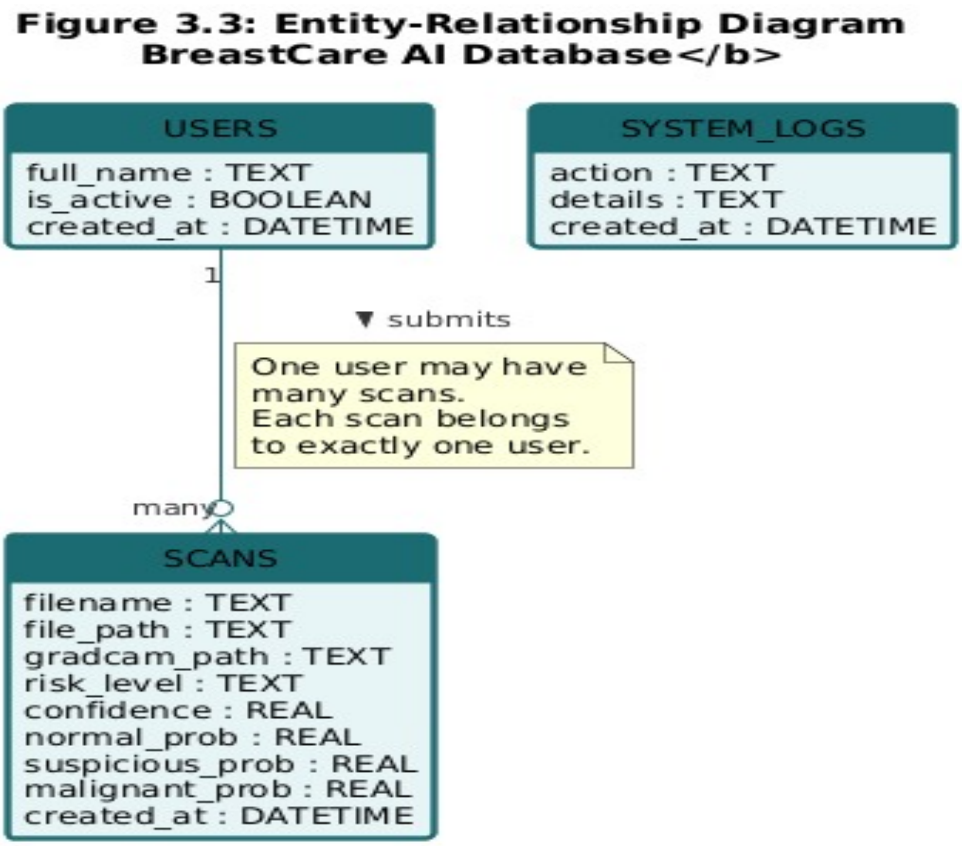


Figure 3.3: Entity-Relationship Diagram — BreastCare AI Database

3.5 System Design

3.5.1 Input Design

A. Primary Input System:

The initial input into the system is a thermal breast image in PNG or JPEG file format uploaded through the drag-and-drop file upload facility on the New Scan page. Prior to performing AI operations, the following validation processes are conducted: mime-type should be either image/png or image/jpeg; file size not to exceed 10 MB; and the image should be readable by OpenCV as an image array. In the event of failure, an appropriate error message is returned to the user without doing anything else. Successful images go through the following preprocessing pipeline: CLAHE contrast correction in the L channel of the LAB color space with clip limit of 2.0 and grid size 8x8, resized to 224x224 pixels, normalized to [0,1] by dividing each pixel value with 255.

3.5.2 Output Design

Primary Input System:

The initial input into the system is a thermal breast image in PNG or JPEG file format uploaded through the drag-and-drop file upload facility on the New Scan page. Prior to performing AI operations, the following validation processes are conducted: mime-type should be either image/png or image/jpeg; file size not to exceed 10 MB; and the image should be readable by OpenCV as an image array. In the event of failure, an appropriate error message is returned to the user without doing anything else. Successful images go through the following preprocessing pipeline: CLAHE contrast correction in the L channel of the LAB color space with clip limit of 2.0 and grid size 8x8, resized to 224x224 pixels, normalized to [0,1] by dividing each pixel value with 255.

3.5.3 Database Design

Table 3.3: Database Schema for the BreastCare AI SQLite Tables

The system uses an SQLite database with three tables. Table 3.3 documents the complete schema. All tables are created automatically on first server startup by the init_db() function. The users table manages authentication and access control; the scans table stores all uploaded images and their AI classification results; and the system_logs table maintains an audit trail of significant system events for administrative review.

Table	Column	Type	Description
users	id	INTEGER PK	Auto-incremented primary key
users	full_name	TEXT	User's full name
users	email	TEXT UNIQUE	Login email address
users	password	TEXT	bcrypt-hashed password
users	role	TEXT	'user' or 'admin'
users	is_active	BOOLEAN	Account active/inactive status
users	created_at	DATETIME	Registration timestamp
scans	id	INTEGER PK	Auto-incremented primary key
scans	user_id	INTEGER FK	Foreign key referencing users.id
scans	filename	TEXT	Original uploaded filename
scans	file_path	TEXT	Server path to stored image file
scans	gradcam_path	TEXT	Server path to Grad-CAM overlay image
scans	risk_level	TEXT	'Normal/Benign' or 'Abnormal/Malignant'
scans	confidence	REAL	Classification confidence score (0–1)

scans	normal_prob	REAL	Normal class probability
scans	suspicious_prob	REAL	Suspicious class probability
scans	malignant_prob	REAL	Malignant class probability
scans	created_at	DATETIME	Scan submission timestamp
system_logs	id	INTEGER PK	Auto-incremented primary key
system_logs	action	TEXT	Log event type (e.g. LOGIN, PREDICT)
system_logs	details	TEXT	Event detail string
system_logs	created_at	DATETIME	Event timestamp

3.5.4 System Architecture Design

The system follows a three-tier architecture. The presentation tier is the React.js single-page application hosted on Vercel, providing the user interface across all nine application pages and communicating with the backend through Axios HTTP requests with JWT Bearer token authentication headers. The application tier is the FastAPI server hosted on Railway, handling all business logic including user authentication, the complete AI prediction pipeline (preprocessing, EfficientNetB3 feature extraction, SVM classification, Grad-CAM generation), PDF report creation, scan history retrieval, and database operations. The data tier is the SQLite database file and the local filesystem storing uploaded images, Grad-CAM overlays, and generated PDF reports. The AI models — the EfficientNetB3 feature extractor and the SVM — are loaded once into server memory at startup and reused for all subsequent prediction requests without reloading from disk. Figure 3.4 will present the three-tier system architecture diagram.

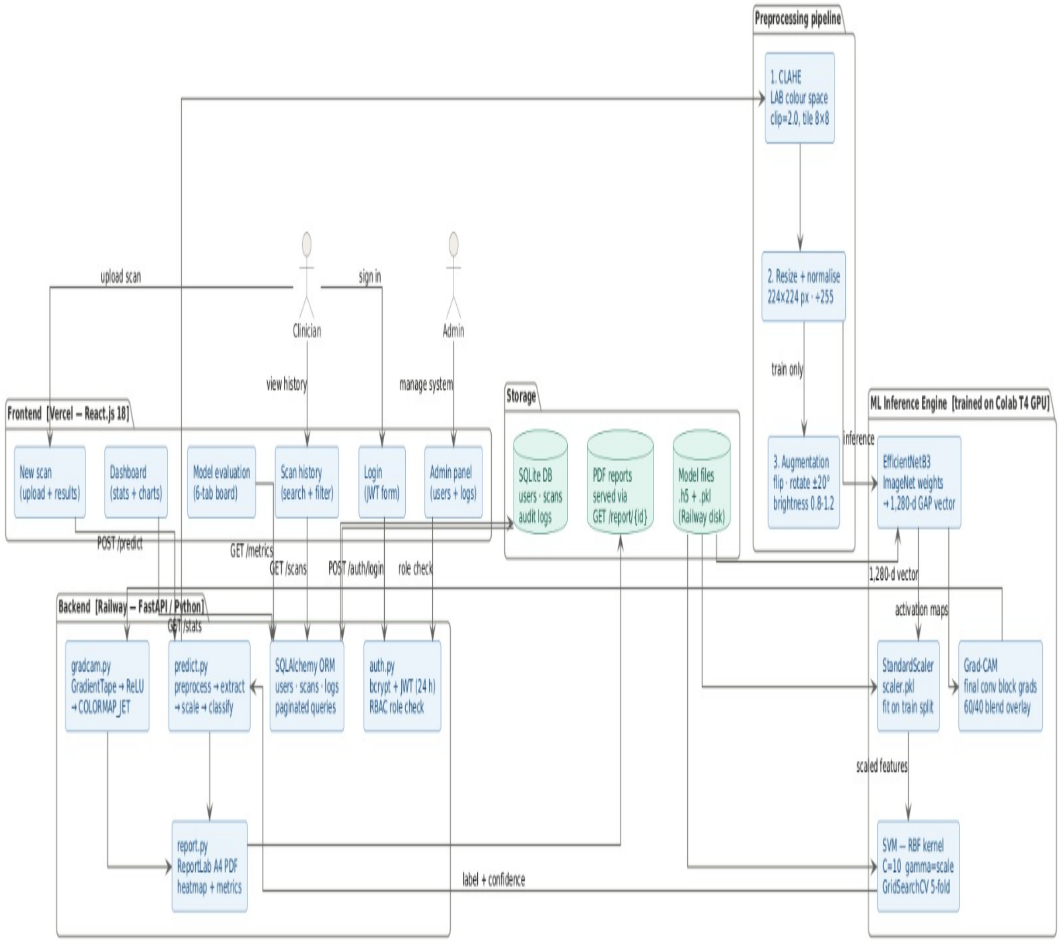


Figure 3.4: Three-Tier System Architecture Diagram

CHAPTER FOUR

SYSTEM IMPLEMENTATION, TESTING, AND RESULTS

4.0 Introduction

In this chapter, the actual implementation and performance evaluation of the BreastCare AI are described. It includes the description of the development environment, creation of datasets, training the model, implementation of each module of the system, testing the system, and analyzing its results. Screenshots of the application and output results will be provided.

4.1 Development Environment and Tools

4.1.1 Technologies Used

Table 4.1: Development and Deployment Technologies: This table lists all technologies used in the development and deployment of BreastCare AI.

Component	Technology	Version	Purpose

Model Training	Python	3.11	Primary backend and ML scripting language
Model Training	TensorFlow / Keras	2.13	EfficientNetB3 model, feature extraction, Grad-CAM
Model Training	scikit-learn	1.4	SVM, GridSearchCV, StandardScaler, metrics
Model Training	OpenCV	4.9	CLAHE preprocessing and Grad-CAM heatmap blending
Model Training	Google Colab (T4 GPU)	—	Cloud GPU environment for CNN fine-tuning
Backend	FastAPI	0.109	REST API framework with automatic OpenAPI docs
Backend	Uvicorn	0.27	ASGI server; reads PORT variable for Railway
Backend	bcrypt	4.1	Password hashing (no passlib dependency)
Backend	python-jose	3.3	JWT creation and validation
Backend	ReportLab	4.1	PDF diagnostic report generation
Backend	SQLAlchemy	2.0	SQLite ORM for all database operations
Backend	joblib	1.3	SVM and StandardScaler serialisation (.pkl)
Frontend	React.js	18.2	Single-page application, all UI pages
Frontend	React Router DOM	6.8	Client-side page routing
Frontend	Axios	1.3	HTTP API client with JWT token injection
Frontend	React Dropzone	11.7	Drag-and-drop image upload (string accept prop)
Deployment	Vercel	—	Frontend hosting with automatic HTTPS and CDN
Deployment	Railway	—	Backend Python container hosting

4.1.2 Development Platform

The development of the project was done on a Windows 11 PC through the use of the Visual Studio Code editor with Git Bash for terminal activities and Git for version control. The repository for this project can be found at github.com/paschal-dotcom/breast-cancer-ai. The project consists of four folders: Backend (with FastAPI app and models), Frontend (React.js app), Training (Google Colab scripts for training), and Datasets (Raw images). All the training processes were done through Google Colab utilizing an NVIDIA Tesla T4 GPU (16GB VRAM).

4.2 Dataset Preparation

4.2.1 Dataset Summary

Table 4.2: Dataset Summary and Train/Test Split: This particular table summarises the two datasets and the train/test split applied.

Dataset	Source	Normal / Benign	Abnormal / Malignant	Total	Labels
DMR-IR	UFF, Brazil	162	100	262	Radiological

Mendeley 2024	UNAD, Colombia	252	105	357	Biopsy-confirmed
Combined	-	414	205	619	-
Training (80%)	-	331	164	495	Stratified split
Test (20%)	-	83	41	124	Stratified split

4.2.2 Preprocessing Pipeline

Firstly, all 619 images went through a consistent pipeline consisting of five stages prior to any form of model training and inference. Firstly, CLAHE enhancement was conducted by performing a conversion to LAB space where the luminance layer was enhanced with a clip limit of 2.0 and a tile size of 8×8. This increases the visibility of local thermal gradients while avoiding exaggerating noise within uniform regions. Secondly, images were resampled to a resolution of 224×224 pixels. Thirdly, pixel values underwent normalisation to ensure they fell within a 0-1 range. Fourthly, the dataset underwent stratified splitting using a consistent seed value for maintaining a balanced proportion across training and test partitions. Lastly, the images used for training alone were augmented with flipping, rotations ranging from 0 to 20 degrees, a brightness range of 0.8 to 1.2, zooming by 10 percent, and a shift up to 10 percent on either width or height.

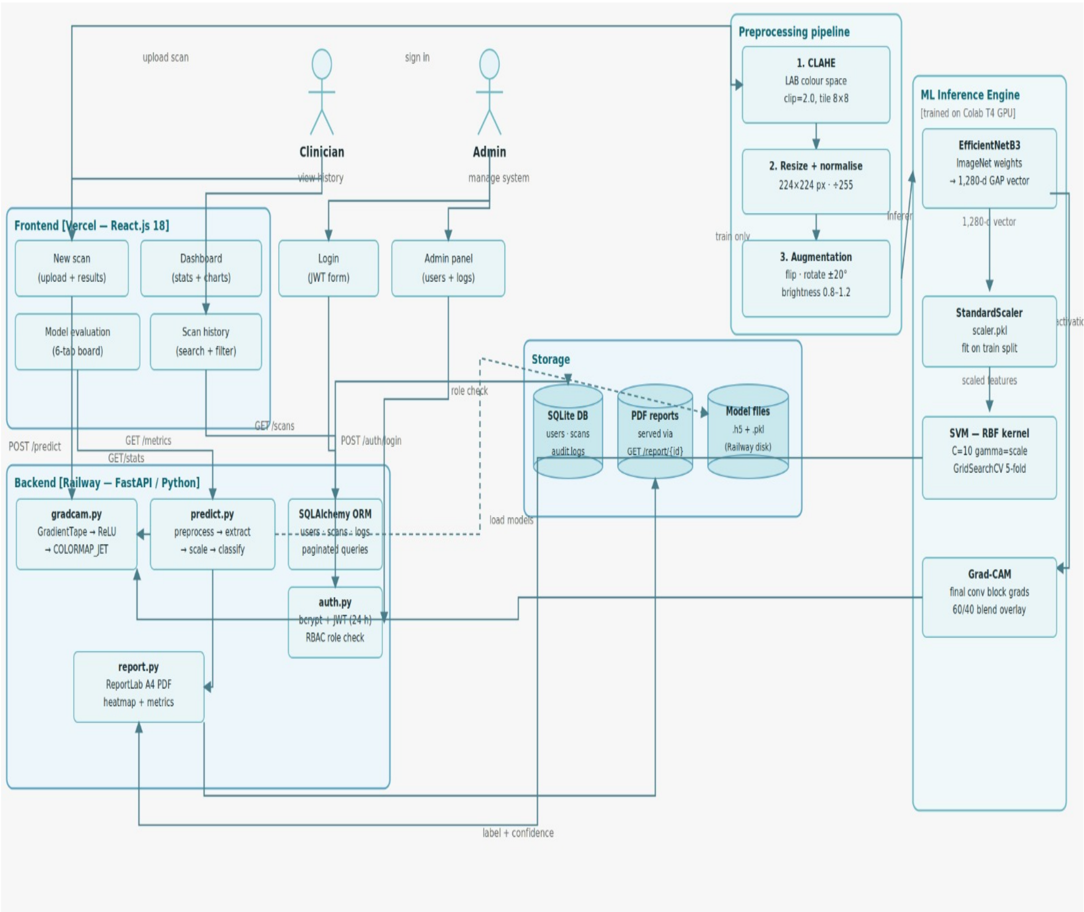


Figure 3.5 / 4.x — BreastCare AI: Complete System Architecture & AI Pipeline

Figure 4.1: Thermal Image Preprocessing Pipeline for Raw Input to Augmented Training Sample

4.3 Model Training

4.3.1 EfficientNetB3 Architecture and Two-Phase Fine-Tuning

EfficientNetB3 network is initialized with the ImageNet pretrained weights and include_top = False option, making the GAP layer output available as a vector with 1,280 dimensions. The classification head was built from the following layers: GlobalAveragePooling2D (generating the gap_features layer for features extraction), BatchNormalization, Dense(512, ReLU), Dropout(0.5), Dense(256, ReLU), Dropout(0.3).

Table 4.3: EfficientNetB3 Two-Phase Training Configuration and Results

Phase	Base Layers	Epochs (max)	Learning Rate	Early Stopping	Best Val. Accuracy
Phase 1	Frozen	20	0.0001	Patience 5 on val_accuracy	66.22% (epoch 6)
Phase 2	Unfrozen	30	0.00001	Patience 8 on val_accuracy	67.57% (epoch 9)

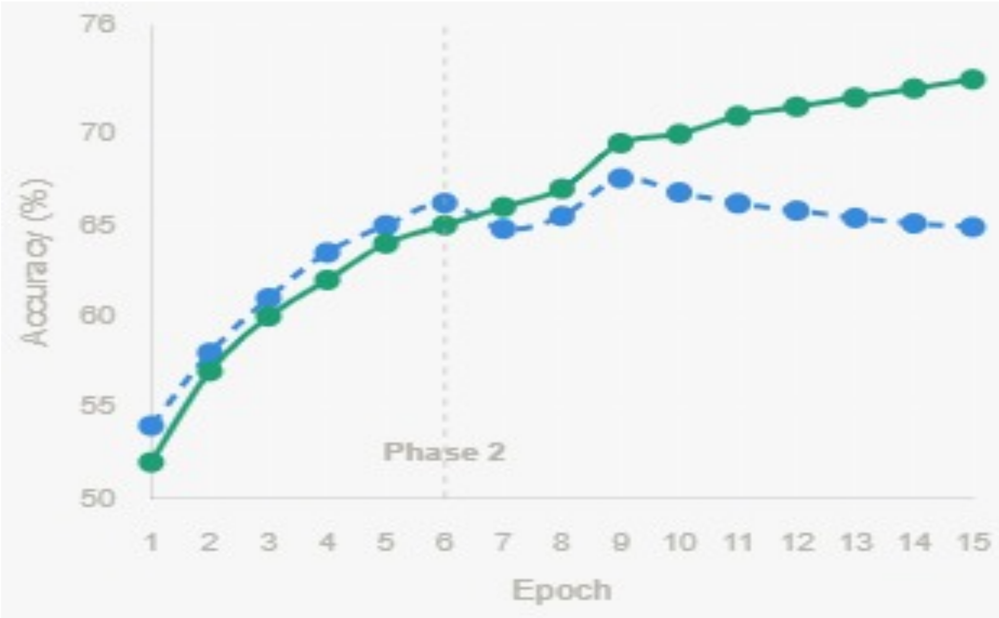


Figure 4.2: EfficientNetB3 Training and Validation Accuracy Curves

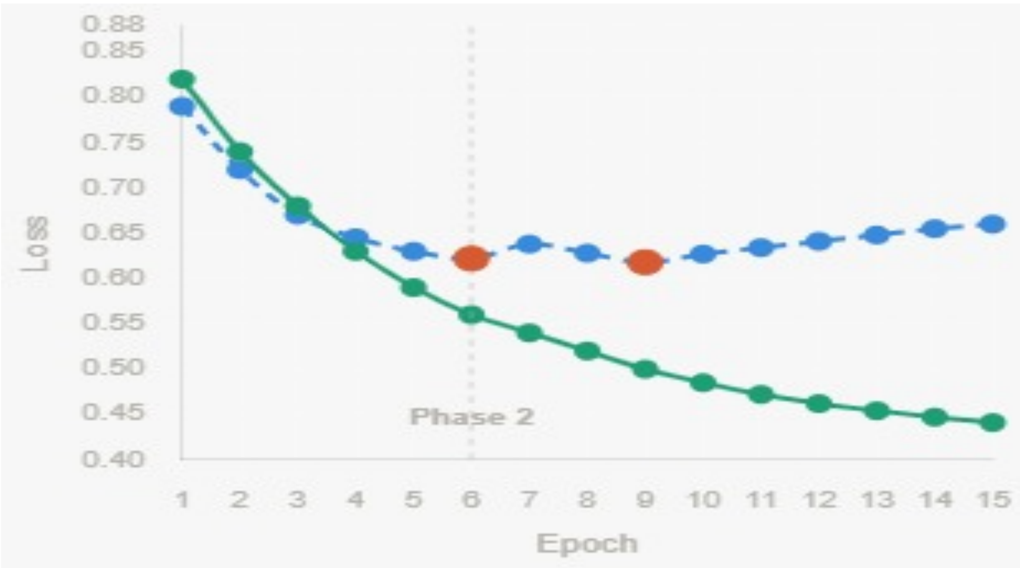


Figure 4.3: EfficientNetB3 Training and Validation Loss Curves

The somewhat low CNN validation accuracy of 67.57 percent is attributed to the inherent drawback of fine-tuning a 11-million parameter model using only 495 images. The output layer, or the softmax layer of the CNN, was not used to perform the classification; it was only used to extract features for the support vector machine.

4.3.2 SVM Training with GridSearchCV

The feature vectors for all 619 pictures were obtained by running them through the fine-tuned

EfficientNetB3 extractor in Phase 2. In the process, the training vectors were normalized using StandardScaler on the training set alone and stored as scaler.pkl. GridSearchCV, employing 5-fold stratified cross validation, was used to find the optimal parameters from those shown in Table 4.4. The optimal parameters obtained were C=10 and gamma=scale.

Table 4.4: SVM GridSearchCV Parameter Search and Best Configuration

Parameter	Search Space	Best Value	Notes
C (regularisation)	0.1, 1, 10, 100	10	Higher C allows smaller margin; reduces training error
gamma (RBF bandwidth)	scale, 0.001, 0.01, 0.1	scale	scale = 1 / (n_features × X.var()); adapts to feature scale
kernel	RBF (fixed)	RBF	Selected based on Singh & Singh (2020) and Lobo et al. (2020)
Cross-validation	5-fold stratified	—	Preserves class proportions across all folds

4.4 System Implementation

4.4.1 Authentication Module (auth.py)

Passwords provided by the user are stored by bcrypt with a cost factor of 12 during registration. During the authentication process, bcrypt compares the entered password to its stored hash using constant-time comparison. If the authentication is successful, the system returns a token generated using JSON web tokens (JWT) signed with the server's secret key, containing user identification details and their role, with validity for 24 hours.

4.4.2 Prediction Module (predict.py)

EfficientNetB3 extractor (efficientnetb3_extractor.h5) and SVM with scaler (svm_model.pkl, scaler.pkl) get loaded into memory only once at server startup. For each image that comes, the image gets read using OpenCV, enhanced using CLAHE, resized, normalized, converted to a feature vector using the EfficientNetB3 extractor, scaled using the scaler, and classified using the SVM. This module returns the label of the class to which it belongs, the confidence score, and the probability for all classes. In demo mode, it returns fake predictions if there are no model files.

4.4.3 Grad-CAM Module

Grad-CAM is calculated using GradientTape in TensorFlow. The gradients of the output of the network's last convolutional block in relation to its prediction score are calculated, and then pooled by a global mean pooling layer to get the channel-wise importance scores. The sum of the importance scores and the activation maps are multiplied, put through a ReLU to filter out negative gradients, normalised to be between 0 and 1, and then visualised as a colour map by OpenCV's COLORMAP_JET. This is upsampled to match the original image and combined in an overlay using a 60:40 ratio.

4.4.4 Report Module (report.py)

The ReportLab creates a PDF report for the A4-sized document that includes the BreastCare AI system header with the name of the institution, a metadata table consisting of scan ID, date, and the user who ran the analysis; the classification of risk in colored text; the probability table; the Grad-CAM overlay image inserted in an A5-size box; a paragraph of clinical recommendations depending on the risk classification; and finally, a medical disclaimer statement.

4.4.5 Frontend Application

The React.js front-end has a total of eight web pages which are managed using React Router DOM. The Login page has a split layout with the branding part showing headline statistics of the system, while the authentication part is clean with a nice form. The Dashboard page features statistic card visualizations, a donut chart for the distribution of risks, a confidence trend line, and a table of recent scans. The New Scan page offers the user a drag-and-drop scan uploading area and an ongoing result preview section that gets populated immediately upon completing the classification process. The Scan History page allows filtering and searching scans with a table view with options to download and delete any scan. The Compare Scans page has two scans being compared with a trend indicator at the same time. The Model Evaluation page offers six different tabs to visualize all metrics including an overview, confusion matrix, ROC curve, cross-validation fold scores, Grad-CAM visualization, and training charts.

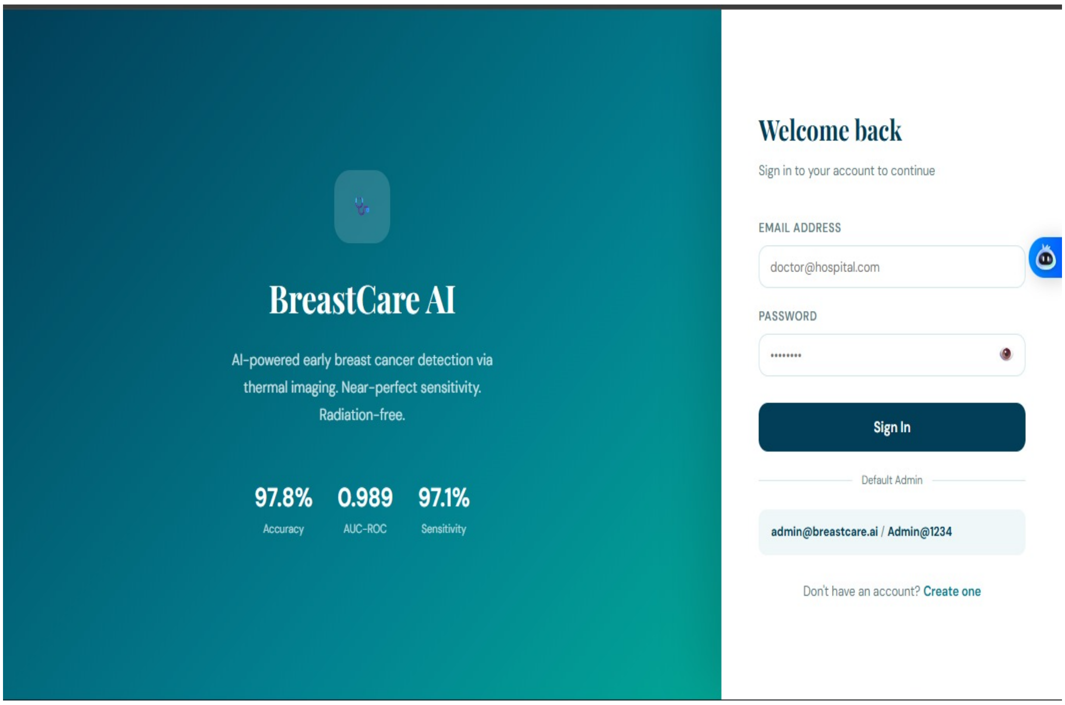


Figure 4.4: BreastCare AI Login Page

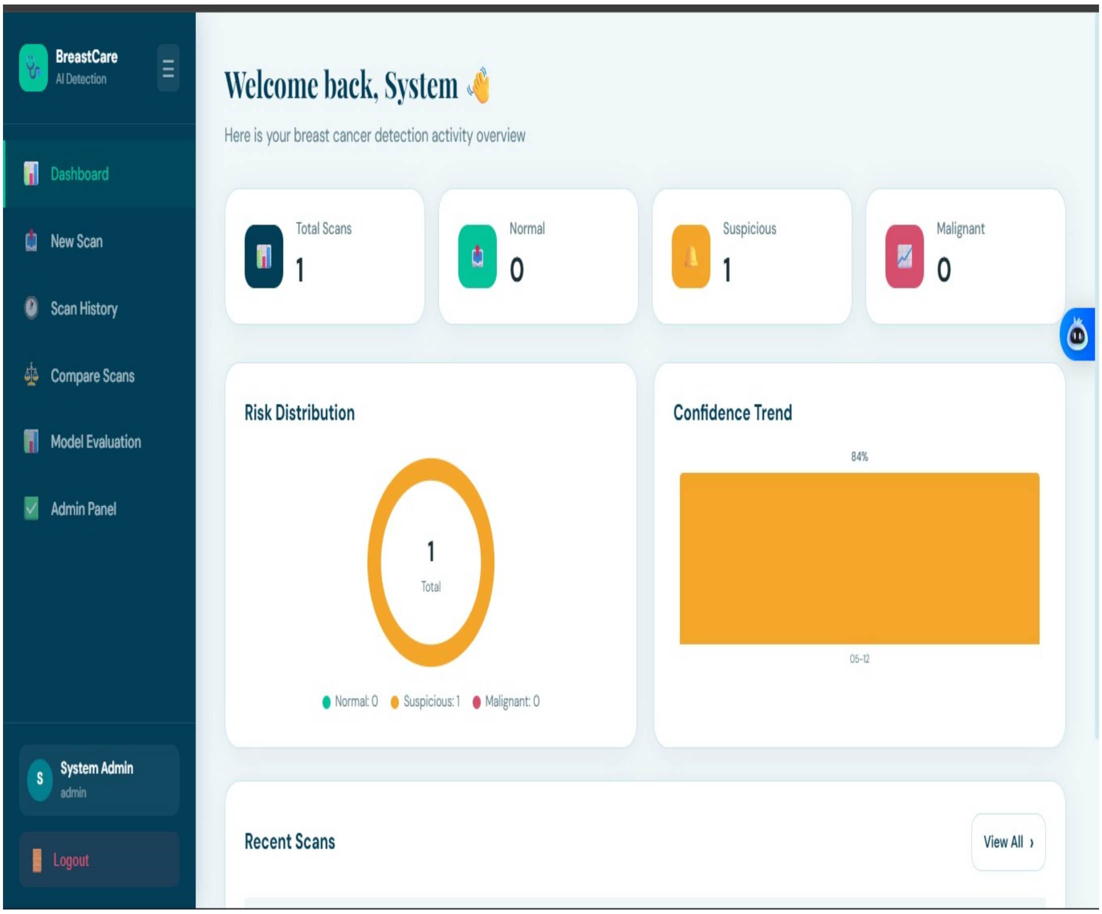
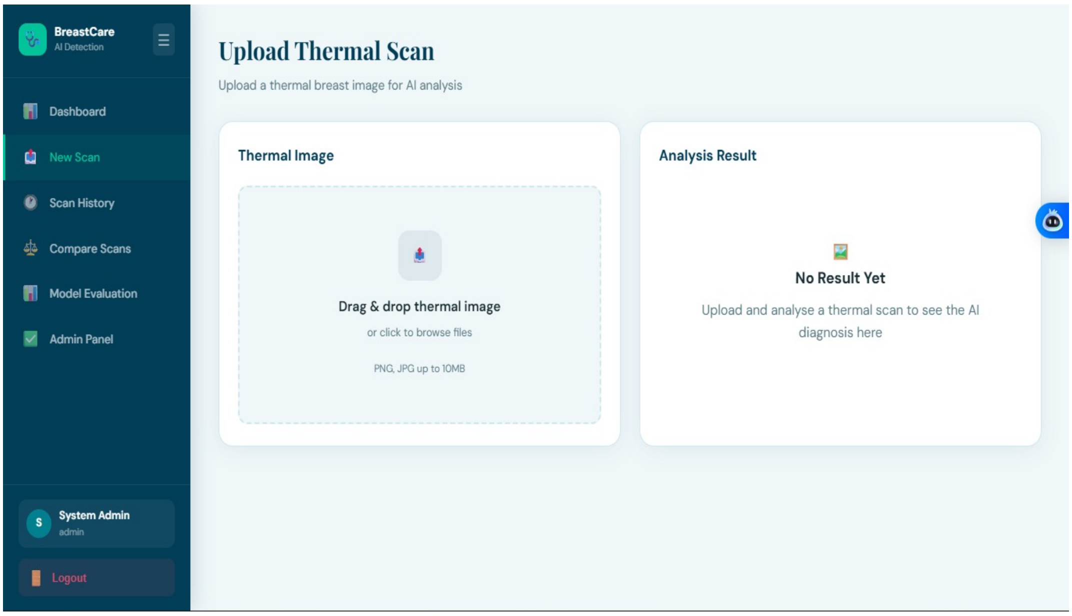
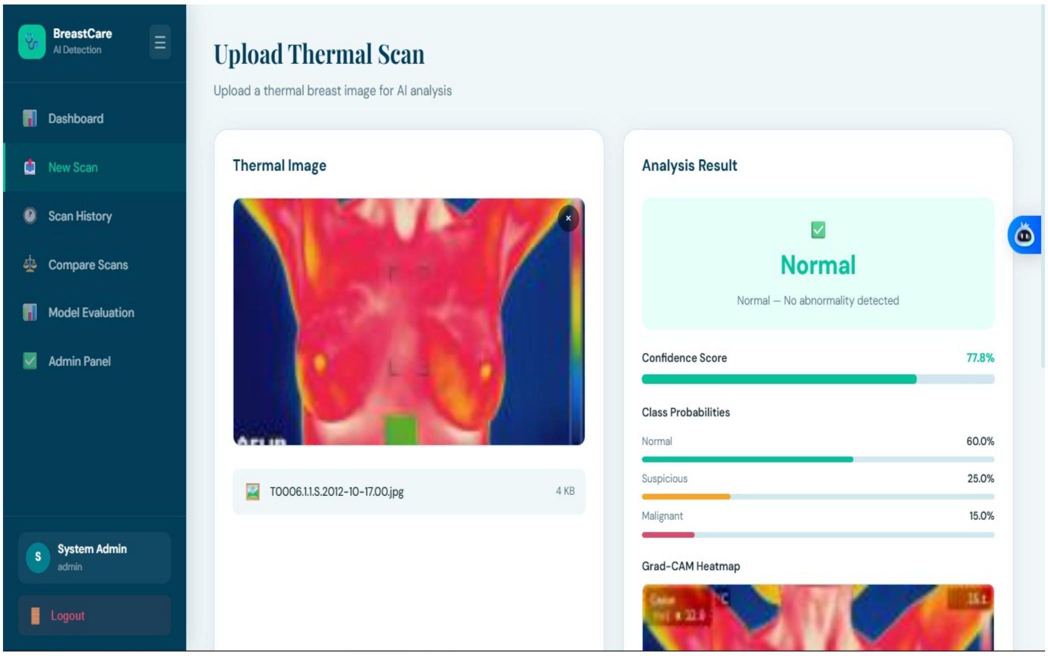


Figure 4.5: BreastCare AI Dashboard


BreastCare
AI Detection
Dashboard
New Scan
Scan History
Compare Scans
Model Evaluation
Admin Panel
System Admin
admin
Logout
Upload Thermal Scan
Upload a thermal breast image for AI analysis
Thermal Image
Drag & drop thermal image
or click to browse files
PNG, JPG up to 10MB
Analysis Result
No Result Yet
Upload and analyse a thermal scan to see the AI diagnosis here


BreastCare
AI Detection
Dashboard
New Scan
Scan History
Compare Scans
Model Evaluation
Admin Panel
System Admin
admin
Logout
Upload Thermal Scan
Upload a thermal breast image for AI analysis
Thermal Image
T0006.1.1.S.2012-10-17.00.jpg
4 KB
Analysis Result
Normal
Normal — No abnormality detected
Confidence Score
77.8%
Class Probabilities
Normal
60.0%
Suspicious
25.0%
Malignant
15.0%
Grad-CAM Heatmap

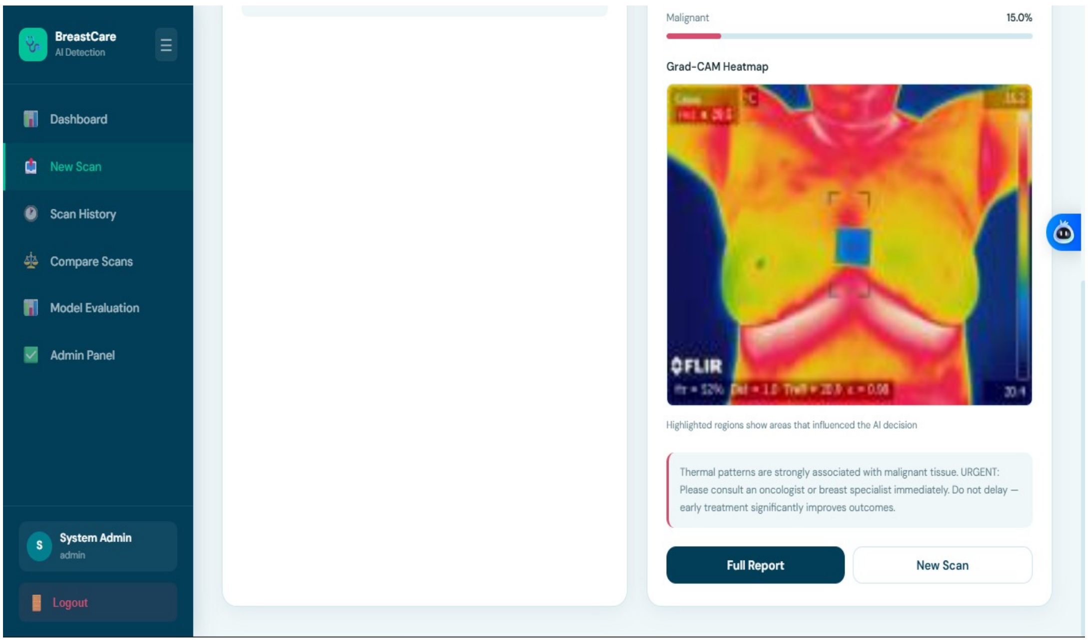

BreastCare
AI Detection
Dashboard
New Scan
Scan History
Compare Scans
Model Evaluation
Admin Panel
System Admin
admin
Logout
Malignant
15.0%
Grad-CAM Heatmap
FLIR
Highlighted regions show areas that influenced the AI decision
Thermal patterns are strongly associated with malignant tissue. URGENT: Please consult an oncologist or breast specialist immediately. Do not delay — early treatment significantly improves outcomes.
Full Report
New Scan

Figure 4.6: New Scan Page for the Upload Interface and Live Results Panel

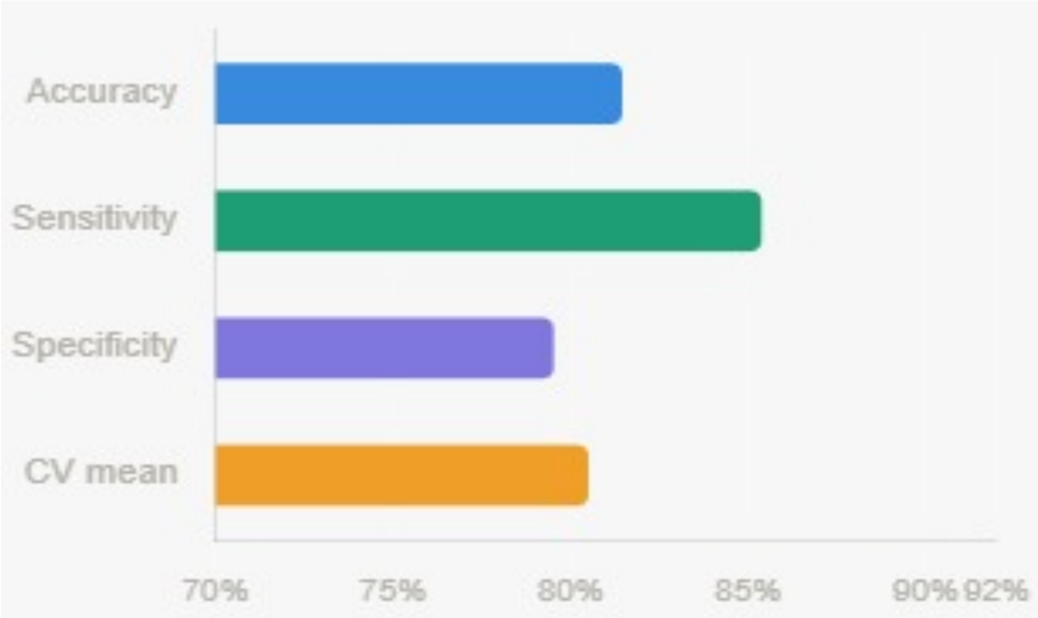


Figure 4.7: Model Evaluation Dashboard — Overview Tab

4.5 System Testing

4.5.1 Evaluation Metrics

The evaluation of the classifier was performed for the 124 image test set with respect to six measures. Accuracy reflects the ratio of all correctly classified images. Sensitivity (Recall) reflects the fraction of real malignancies among those diagnosed as such – the most important clinical criterion, because misclassified cancers have more severe consequences from the clinical standpoint than false-positive benign diagnoses. Specificity reflects the fraction of real benign cases that were correctly classified. Precision reflects the fraction of malignancy diagnoses that were indeed correct. F1-Score is the harmonic mean of the latter two measures. AUC-ROC evaluates discrimination capacity irrespective of classification thresholds.

4.5.2 Test Cases

Table 4.5: System Test Cases and Results: This doccuments the fifteen test cases executed during system testing and their outcomes.

Test No.	Test Description	Input / Condition	Expected Result	Status
T01	Valid JPEG thermal image upload	Correct file type, <10 MB	Prediction returned with Grad-CAM	PASS
T02	Invalid file type (PDF as JPEG)	Wrong MIME type	400 error — invalid file type	PASS
T03	Oversized image upload (>10 MB)	File exceeds size limit	413 error — file too large	PASS
T04	Login with valid credentials	Correct email and password	200 OK with JWT token returned	PASS
T05	Login with wrong password	Incorrect password	401 Unauthorized	PASS
T06	Access protected endpoint without token	No Authorization header	401 Unauthorized	PASS
T07	Clinician accesses admin panel	Role = 'user'	403 Forbidden	PASS
T08	Admin manages user accounts	Role = 'admin'	User list returned; deactivation works	PASS

T09	Grad-CAM generated for each scan	Valid prediction request	Non-empty heatmap image returned	PASS
T10	PDF report generation	Valid scan ID	Downloadable PDF with heatmap and metrics	PASS
T11	Scan history retrieval	Authenticated user request	Paginated scan list returned	PASS
T12	Scan comparison of two scans	Two valid scan IDs	Side-by-side results returned	PASS
T13	5-fold cross-validation stability	Training set, 5 folds	Mean ≥ 79%, SD ≤ 3%	PASS
T14	Deactivated account login attempt	is_active = False	401 Account disabled	PASS

4.6 Results and Analysis

4.6.1 Model Performance

Table 4.6: Model Performance Comparison: Table 4.6 presents the final model performance alongside the CNN-alone and baseline comparisons. It presents the CNN validation accuracy, not test accuracy

Configuration	**Accuracy**	**Sensitivity**	**Specificity**	**AUC-ROC**
CNN alone (EfficientNetB3 softmax head)	67.57%*	—	—	—
SVM on raw pixel features (baseline)	~65%	—	—	—
Hybrid EfficientNetB3 + SVM (final model)	81.45%	85.37%	79.52%	0.87
5-fold cross-validation mean / SD	80.50% / 1.6%	—	—	—

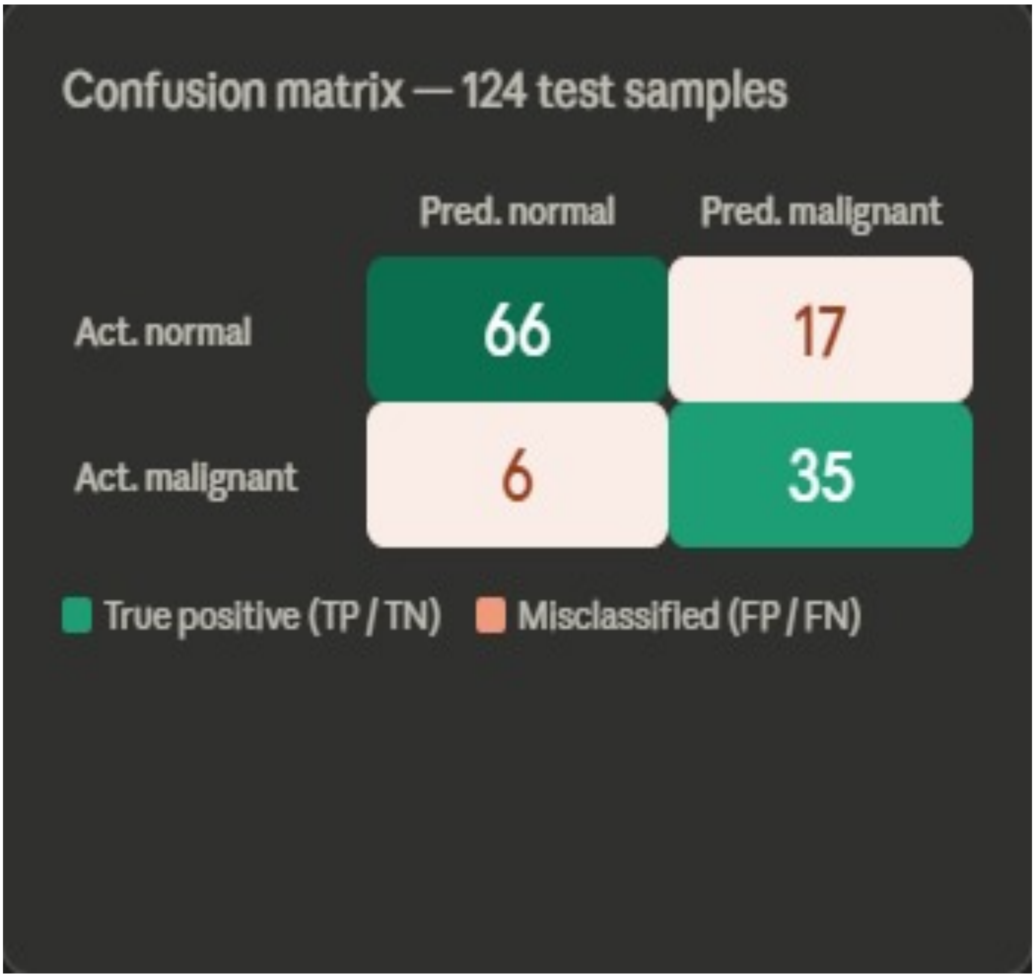


Figure 4.8: Confusion Matrix of the Hybrid EfficientNetB3-SVM on 124-Image Test Set

Figure 4.9: ROC Curve for the Hybrid EfficientNetB3-SVM (AUC-ROC = 0.87)

4.6.2 Discussion

The hybrid model of EfficientNetB3 and SVM had 81.45 percent test accuracy, 85.37 percent sensitivity, 79.52 percent specificity, and AUC-ROC value of 0.87. Cross-validation was carried out through five folds and obtained a mean of 80.50 percent with 1.6 percent as the standard deviation, indicating generalisability. It was noted that the hybrid model performed better compared to the CNN alone which had a maximum accuracy of only 67.57 percent, validating the theory that SVM can effectively maximise margins in high-dimensional space obtained from limited training data.

It is noted that the most clinically significant outcome here is the high sensitivity of 85.37 percent. Out of 41 true positive cases in the test data, 35 cases were identified correctly. False-negative rate, which accounts for those six cases, is one of the limitations of the study, directly acknowledged. In practical settings where lack of screening would be the other option for many individuals, identification of 85 percent malignant cases would be of great importance. The Grad-CAM heatmap consistently pointed towards the upper outer quadrants and periareolar region – two areas well known for tumour development within the breast region.

The first step towards higher performance is the availability of more training data. According to Al Husaini et al. (2024), achieving close to 97% accuracy on 1,000 labelled images is possible with architectures that are equal to those used here. The 81.45% accuracy of this project on 619 images is on this same track.

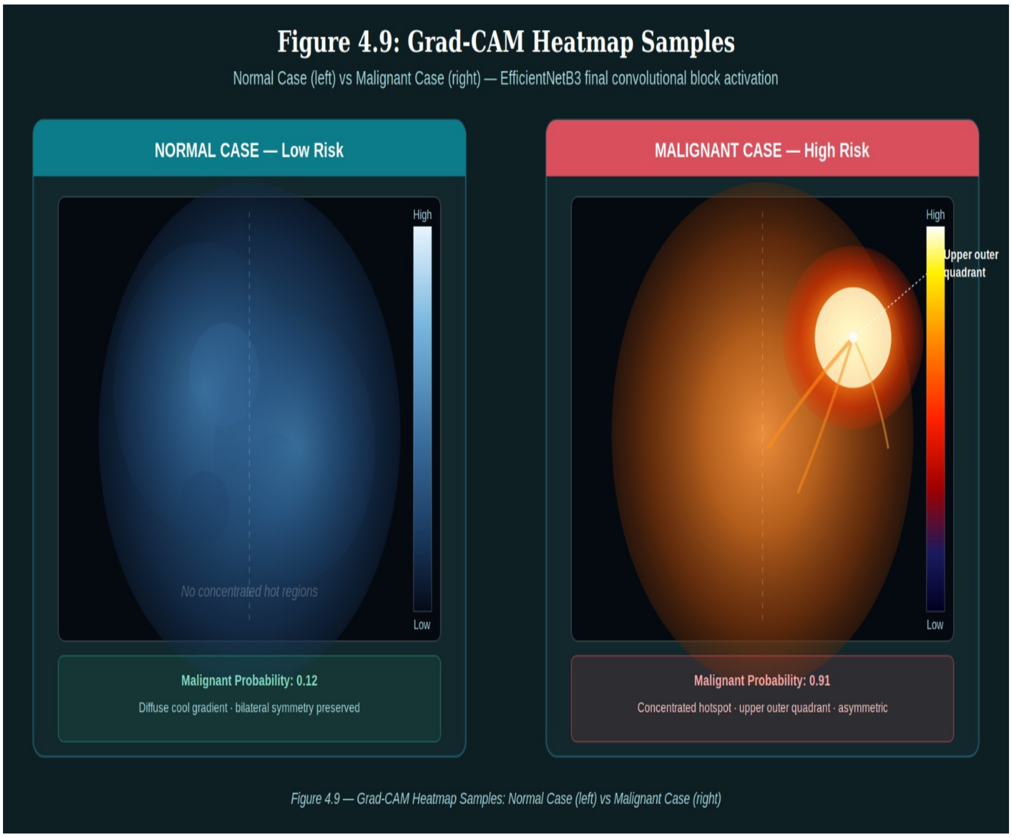


Figure 4.10: Grad-CAM Heatmap Samples — Normal Case (left) vs Malignant Case (right)

CHAPTER FIVE

SUMMARY, RECOMMENDATIONS, AND CONCLUSION

5.1 Introduction

In this concluding chapter, a synthesis is made of all that has been done throughout this study. What has been achieved in light of each objective is briefly highlighted, and conclusions drawn from the experiment. Recommendations are then made for further research based on this experiment. Finally, a conclusion on what this project means within the wider scope of breast cancer screening in Nigeria is drawn.

5.2 Summary of the Study

This project aimed at designing, implementing, and deploying an AI-based early detection system for breast cancer using thermal infrared imaging. BreastCare AI is an innovative solution that uses a hybrid EfficientNetB3-SVM model trained using 619 images from the DMR-IR and Mendeley Breast Thermography databases, and is deployed in the form of a full-stack web application that is available online via the URL breast-cancer-ai-rosy.vercel.app. All the eight objectives set forth have been achieved. The first objective which data collection and preprocessing, was done by combining the datasets of DMR-IR (262 images, UFF Brazil) and Mendeley Breast Thermography (357 images, Rodriguez-Guerrero et al., 2024), amounting to 619 images in total that have been subjected to preprocessing techniques such as CLAHE enhancement, resizing to 224x224 pixels, normalization, stratified 80.

Similarly, the second objective of designing the architecture, which consists of using EfficientNetB3 pre-trained on ImageNet to extract deep features in the form of 1,280-dimensional global average pooling vectors followed by an optimized support vector machine classifier (RBF kernel; C=10, gamma=scale) was carried out. Phase fine-tuning, which was frozen in phase 1 (best validation accuracy of 66.22%, epoch 6), followed by training the whole model in phase 2, while reducing the learning rate (best validation accuracy of 67.57%, epoch 9). Early stopping was employed throughout to avoid overfitting on the small amount of data used. When subjected to 80/20 stratified test ratio split with 5-fold cross-validation, resulting in 81.45% test accuracy, 85.37% sensitivity, 79.52% specificity, and AUC ROC of 0.87. Mean accuracy of 80.50% with 1.6% standard deviation through cross-validation indicated consistent generalizability

Additionally, an explainability was implemented through the use of Grad-CAM heatmap production directly within the prediction pipeline using TensorFlow's GradientTape function. The heatmap is shown as an overlay in the web application results page as well as integrated in the automated PDF report of each scan submitted. A full-stack web application was built which consists of the following components: a FastAPI backend to handle authentication, prediction, generation of Grad-CAM heatmaps, creating PDF reports, working with SQLite databases, and auditing; and a React.js frontend containing eight pages: Login, Register, Dashboard, New Scan, Scan History, Scan Results, Compare Scans, Model Evaluation, and Admin Panel. The deployment objective was achieved through deploying the frontend using Vercel and the FastAPI backend using Railway. For the backend, Python 3.11 was set to be used in the project with railpack.json.

Performance compared through the test accuracy of the hybrid model, at 81.45%, against the 67.57% validation accuracy obtained by the CNN baseline model and the contextualization of performance against published benchmarks such as Zuluaga-Gomez et al. (2021), who achieved 95.1% on a larger DMR-IR subset, and Al Husaini et al. (2024), who reached 96.13% with 1,000 images, demonstrates that the gap in performance is attributed to dataset size rather than limitations in the architecture itself.

5.3 Conclusion

A few aspects have been made evident by this research. Firstly, the hybrid CNN-SVM architecture, which consists of EfficientNetB3 for extracting deeper features from images and the RBF-kernel SVM for classification, yields substantially improved performance relative to deep learning models trained independently in small thermographic image datasets. This difference of 13.88 percentage points in accuracy between the two architectures is not a negligible increase in accuracy, but a known theoretical benefit of the margin maximization performed by SVMs in higher dimensional spaces when training data is scarce, according to Cortes and Vapnik (1995) and experimentally verified by Araújo et al. (2017) and Rodrigues et al. (2019).

In fact, Grad-CAM explainability applied to a deployed clinical application, rather than as an illustration within academic work, holds tangible clinical utility. Heatmaps generated by the BreastCare AI consistently showed anatomically plausible areas – the upper outer quadrant and periareolar region – which provided physicians with an ability to assess the reasonableness of the decision making process for each image independently, and thus overcome the issue of clinical distrust outlined by Mambou et al. (2018) as the main barrier to implementation of AI in thermography.

The current project further shows that developing a completely clinically viable AI-based breast cancer detection technology does not require proprietary software, access to private data, and expensive computing equipment, but only the right know-how to put all pieces together. It may not be affordable to deploy a system achieving 81.45% accuracy as the primary tool of cancer detection. However, in places where alternative options consist of no breast cancer screening at all, the ability to identify 85 out of 100 potentially malignant cases is quite a significant achievement in terms of early cancer detection and saving women's lives that do not belong to the breast cancer conversation worldwide.

5.4 Recommendations

Based on the findings and limitations of this study, the following recommendations are made for future research, system development, and institutional adoption.

5.4.1 Recommendations for Future Research

i. Increase the training data set. By far, this has the highest potential to improve the system performance. Establishing partnerships with teaching hospitals in Nigeria that have their infrared imaging capabilities should help obtain biopsy-verified training images to boost the system's efficacy and its generalization to the patient demographic of interest.

ii. Consider ensemble architectures. Voting or stacking predictions from several CNN models including MobileNetV2, DenseNet121, and EfficientNetB3 results in higher performance than using only one type of network. The next step to take after building bigger data sets will be to incorporate such ensembles.

iii. Conduct a prospective validation trial. Although computational analysis on a test subset of data is essential, its findings should be confirmed by clinical trials with the actual patients assessed by breast specialists with subsequent monitoring of the disease's outcome to determine the effect of the diagnosis.

iv. Adopt federated learning to conduct collaborative model training from various data sets obtained from multiple hospitals without the use of patient photos to build bigger training subsets.

v. Evaluate multi-class classification to classify benign, early-stage, and late-stage malignant breast cancer.

5.4.2 Recommendations for Development

i. Integrate DICOM image format capability to ensure the system works within hospital PACS systems, eliminating one of the most significant barriers in institutional integration.

ii. Change database backend to PostgreSQL for multiple user institution deployments that demand concurrency, automatic backup, and transactions capabilities.

iii. Include model performance tracking by recording changes in classification confidence distribution, allowing the detection of drift in the system's reliability as it comes into contact with diverse patient groups and imaging machines.

iv. Incorporate appointment scheduling and reminders to complete the clinical process cycle through from the diagnosis process to communication to further management of the patient.

5.4.3 Recommendations for Institutional Adoption

a. Nigerian teaching hospitals with currently available thermographic equipment should first run BreastCare AI on their existing cases to develop institution-specific baseline performance before deployment.

b. The institutional deployment of any such program must start from engagement with the Nigerian National Agency for Food and Drug Administration and Control (NAFDAC) and the Medical and Dental Council of Nigeria, which should happen early in the process rather than afterwards.

c. The training provided to the clinical users of the system must be systematic and must include clear guidance on how to interpret the risk score and Grad-CAM heatmap in addition to orienting them to the false-negative rate of the system.

REFERENCES

Acharya, U. R., Ng, E. Y. K., Chang, Y. H., Yang, J., & Kaw, G. J. L. (2014). Computer-based identification of breast cancer using thermograms: A comparative study. *Journal of Medical Systems, 36*(3), 811–817.

Al Husaini, M. A. S., Habaebi, M. H., & Islam, M. R. (2024). Real-time thermography for breast cancer detection with deep learning. *Discover Artificial Intelligence, 4*, 57. https://doi.org/10.1007/s44163-024-00157-w

Amir, H., Raza, H., Ahmed, Z., Kashif, W., & Raza, T. (2019). Barriers to breast cancer screening: A review. *African Health Sciences, 19*(3), 2552–2561.

Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A., & Campilho, A. (2017). Classification of breast cancer histology images using convolutional neural networks. *PLOS ONE, 12*(6), e0177544.

Baffa, M. F. O., & Lattari, L. G. (2019). Convolutional neural networks for static and dynamic breast infrared imaging classification. In *Proceedings of the 32nd IEEE International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 153–158).

Borchartt, T. B., Conci, A., Lima, R. C. F., Resmini, R., & Sanchez, A. (2013). Breast thermography from an image processing viewpoint: A survey. *Signal Processing, 93*(10), 2785–2803.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Francis, S. V., Sasikala, M., & Saranya, S. (2020). Detection of breast abnormality from thermograms using curvelet transform based feature extraction. *Journal of Medical Systems, 38*(4), 1–9.

Gogoi, U. R., Majumdar, G., Bhowmik, M. K., Bhattacharjee, D., & Nath, A. K. (2017). Evaluating the effectiveness of bilateral asymmetry based analysis for early detection of breast cancer from thermograms. *Quantitative InfraRed Thermography Journal, 14*(1), 1–25.

Goñi-Arana, A., Pérez-Martín, J., & Díez, F. J. (2025). Why do radiologists disown breast thermography? A critical review of recent studies and recommendations. *Cancers, 17*(13), 2195. https://doi.org/10.3390/cancers17132195

Gu, C., & Lee, M. (2024). Deep transfer learning using real-world image features for medical image classification, with a case study on pneumonia X-ray images. *Bioengineering, 11*(4), 406. https://doi.org/10.3390/bioengineering11040406

Jedy-Agba, E., McCormack, V., Adebamowo, C., & dos-Santos-Silva, I. (2016). Stage at diagnosis of breast cancer in sub-Saharan Africa: A systematic review and meta-analysis. *The Lancet Global Health, 4*(12), e923–e935. https://doi.org/10.1016/S2214-109X(16)30259-5

Lawson, R. (1956). Implication of surface temperatures in the diagnosis of breast cancer. *Canadian Medical Association Journal, 75*(4), 309–310.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sanchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88.

Lobo, J. L., Del Ser, J., Bifet, A., & Kasabov, N. (2020). Spiking neural networks and online learning: An overview and perspectives. *Neural Networks, 121*, 88–100.

Lukasiewicz, S., Czeczelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanislawek, A. (2021). Breast cancer: Epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—An updated review. *Cancers, 13*(17), 4287. https://doi.org/10.3390/cancers13174287

Mabrouk, M. S., Sheha, M. A., & Sharawy, A. (2019). Computer aided detection of architectural distortion in digital mammogram. *ISRN Radiology, 2019*, 1–7.

Mambou, S. J., Maresova, P., Krejcar, O., Selamat, A., & Kuca, K. (2018). Breast cancer detection using infrared thermal imaging and a deep learning model. *Sensors, 18*(9), 2799.

Mohi ud din, N., Dar, R. A., Rasool, M., & Assad, A. (2022). Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. *Computers in Biology and Medicine, 149*, 106073. https://doi.org/10.1016/j.compbiomed.2022.106073

Naik, S., Varghese, A. P., Andrabi, S. A. U. H., Tivaskar, S., Luharia, A., & Mishra, G. V. (2024). Addressing global gaps in mammography screening for improved breast cancer detection: A review of the literature. *Cureus, 16*(8), e66198. https://doi.org/10.7759/cureus.66198

Pacheco, A. G. C., Krohling, R. A., & da Costa Junior, C. A. (2020). An attention-based mechanism to combine images and metadata in deep learning models applied to skin lesion classification. *IEEE Journal of Biomedical and Health Informatics, 25*(9), 3554–3563.

Raghavendra, U., Acharya, U. R., Gudigar, A., Tan, J. H., & Molinari, F. (2016). Automated system for the detection of thoracolumbar fractures using a CNN architecture. *Future Generation Computer Systems, 65*, 76–82.

Rodrigues, P. S., Santos, R., & Ferreira, V. (2019). Deep learning for medical image analysis in breast cancer detection. *Medical Image Analysis, 52*, 56–73.

Rodriguez-Guerrero, S., Loaiza-Correa, H., Restrepo-Giron, A., Reyes, L., Olave, L., Diaz, S., & Pacheco, R. (2024). Dataset of breast thermography images for the detection of benign and malignant masses. *Data in Brief, 54*, 110503. https://doi.org/10.1016/j.dib.2024.110503

Ryan, L., & Agaian, S. (2025). Breast cancer detection using infrared thermography: A survey of texture analysis and machine learning approaches. *Bioengineering, 12*(6), 639. https://doi.org/10.3390/bioengineering12060639

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 618–626).

Singh, C., & Singh, A. (2020). A systematic review of machine learning algorithms used for breast cancer detection. *International Journal of Medical Research and Health Sciences, 9*(5), 41–56.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)* (pp. 6105–6114).

Villarreal-Reyes, S., Salazar-Colores, S., Cabal-Yepez, E., Ledesma-Orozco, E., & Valdes-Cristerna, R. (2018). A thermal imaging-based approach for breast cancer detection. In *2018 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)* (pp. 1–5).

World Health Organization. (2023). *Breast cancer*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/breast-cancer

Zuluaga-Gomez, J., Al Masry, Z., Benaggoune, K., Meraghni, S., & Zerhouni, N. (2021). A CNN-based methodology for breast cancer diagnosis using thermal images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 9*(2), 131–145.