CHAPTER I

INTRODUCTION

1.1 Background

Breast cancer remains one of the most significant global health challenges and continues to be the leading cause of cancer-related mortality among women worldwide (Halid et al., 2024). According to the World Health Organization (2023), breast cancer accounts for approximately 2.3 million new cases and 685,000 deaths annually, making it the most frequently diagnosed cancer in women across 157 countries. The Global Cancer Observatory (GLOBOCAN) 2020 statistics further emphasize this alarming trend, reporting that breast cancer represents 11.7% of all cancer cases globally, surpassing lung cancer as the most commonly diagnosed cancer (Sung et al., 2021). These staggering figures underscore the urgent need for improved detection strategies and accessible screening programs, particularly in regions where healthcare resources are limited. The disproportionate impact of breast cancer on low- and middle-income countries (LMICs) is especially concerning, as these nations account for approximately 60% of breast cancer deaths despite having only 45% of global incidence (International Agency for Research on Cancer, 2020). This disparity highlights the critical gap between disease burden and healthcare capacity that must be addressed through innovative approaches.

Early detection of breast cancer has been proven to significantly improve treatment outcomes and patient survival rates(Nyoman et al., 2026). The American Cancer Society (2023) emphasizes that when breast cancer is detected early at a localized stage, the 5-

year relative survival rate reaches 99%. However, this rate drops dramatically to 27% when the cancer has metastasized to distant organs. This dramatic difference underscores why early detection is not merely beneficial but absolutely critical for patient survival. Mammography has established itself as the gold standard screening modality for early breast cancer detection, with numerous large-scale randomized controlled trials demonstrating its effectiveness in reducing breast cancer mortality by 20-40% (Tabár et al., 2019). The landmark studies by Tabár and Duffy (2012) provided compelling evidence that regular mammographic screening could substantially reduce breast cancer deaths, leading to widespread adoption of screening programs in developed nations. The Swedish Two-County Trial, one of the largest and longest-running mammography studies, demonstrated a 31% reduction in breast cancer mortality among women invited to screening, providing foundational evidence for screening programs worldwide (Tabár et al., 2019).

Despite its proven benefits, the interpretation of mammograms presents considerable challenges that can compromise diagnostic accuracy. Mammographic images are inherently complex due to the superimposition of breast tissue, variations in breast density, and the subtle nature of early cancerous changes (Birdwell, 2009). The presence of dense breast tissue, which affects approximately 40-50% of women undergoing screening, can mask underlying lesions and significantly reduce the sensitivity of mammography (Boyd et al., 2007). Women with extremely dense breasts have 4-6 times higher breast cancer risk and substantially reduced mammographic sensitivity, creating a double burden of increased risk and decreased detection capability. Furthermore, the overlapping of normal fibroglandular tissue can create patterns that mimic or obscure malignancies, leading to both false-positive and false-negative interpretations

(Wittenberg et al., 2020). False-positive findings cause unnecessary anxiety, additional imaging, and invasive procedures such as biopsies, while false-negative findings delay diagnosis and treatment, potentially allowing cancers to progress to advanced stages. These interpretive challenges are compounded by significant inter-reader variability, with studies showing that radiologists' experience levels and individual interpretive styles can substantially affect diagnostic accuracy (Mazurowski et al., 2012). Sickles and Dershaw (2010) reported that even among expert breast imagers, there exists considerable variation in the interpretation of the same mammographic studies, highlighting the subjective nature of conventional mammogram reading. This variability is particularly pronounced for subtle or ambiguous findings and in less-experienced readers, creating challenges for healthcare systems with limited access to specialized breast imaging expertise(Hadju et al., 2024).

The advent of artificial intelligence (AI) (Marti et al., 2020) and deep learning technologies has opened new avenues for addressing these diagnostic challenges. Deep learning, particularly Convolutional Neural Networks (CNNs), has demonstrated remarkable capabilities in medical image analysis, often achieving performance comparable to or exceeding that of human experts in specific tasks (LeCun, Bengio, & Hinton, 2015). Unlike traditional computer-aided detection (CAD) systems that rely on hand-crafted features defined by human experts, deep learning models automatically discover relevant features directly from data, potentially identifying patterns too subtle for human specification. In the domain of mammography, deep learning models have shown promising results in detecting malignant lesions, reducing false positives, and improving workflow efficiency (Esteva et al., 2017). The ability of CNNs to automatically learn hierarchical features from raw image data makes them particularly well-suited for capturing the complex patterns and subtle abnormalities present in mammograms (Rahman et al., 2023). Studies by López-Cabrera et al. (2020) and Mahmood et al. (2021) have demonstrated that deep learning models can achieve high accuracy in breast cancer classification, with some architectures reaching performance levels comparable to experienced radiologists.

However, the successful deployment of deep learning models in medical imaging faces a fundamental challenge: the requirement for large, well-annotated datasets for training. Medical image datasets are notoriously difficult and expensive to acquire due to privacy concerns, the need for expert annotation, and the relative rarity of pathological cases (Susilo & Sugiharti, 2021). Creating a high-quality medical imaging dataset requires multiple steps: obtaining ethical approval, securing patient consent, collecting images across diverse populations and disease presentations, and having expert radiologists provide accurate annotations. Each of these steps presents significant logistical and financial barriers. This data scarcity problem is particularly acute in low- and middle-income countries (LMICs)(Nugraha Gautama et al., 2024), where healthcare infrastructure limitations and competing priorities often preclude the development of large-scale digital imaging repositories (International Agency for Research on Cancer, 2020). Transfer learning has emerged as a powerful technique to mitigate this challenge by leveraging knowledge learned from large, general-purpose image datasets (such as ImageNet) and adapting it to specific medical imaging tasks (Yosinski et al., 2014). Models pre-trained on ImageNet, which contains over 14 million natural images across 1000 categories, learn a rich hierarchy of features that can be transferred to medical imaging tasks. This approach enables the development of effective diagnostic models even with limited domain-specific data, making it particularly attractive for applications in resource-constrained settings.

The situation in Southeast Asia, particularly in Indonesia and Timor-Leste, exemplifies the challenges and opportunities in implementing AI-based breast cancer detection. Indonesia, as the fourth most populous country in the world, faces a substantial and growing burden of breast cancer. According to the Global Cancer Observatory, Indonesia reported approximately 65,858 new breast cancer cases and 22,430 deaths in 2020, with age-standardized incidence rates of 40.3 per 100,000 women (Sung et al., 2021). The Indonesian Ministry of Health (Kementerian Kesehatan Republik Indonesia, 2021) has identified late-stage diagnosis as a critical problem, reporting that over 70% of breast cancer patients are diagnosed at advanced stages (III and IV), where 5-year survival drops to less than 40%. This late-stage diagnosis is particularly prevalent in rural and underserved areas where access to mammography screening is severely limited. The country's vast geographical spread across more than 17,000 islands creates immense logistical challenges for healthcare delivery, with advanced diagnostic services concentrated primarily in major urban centers on Java island. According to the Indonesian Radiology Society, Indonesia has approximately 1,200 radiologists serving a population of 270 million people, resulting in a ratio of only 0.44 radiologists per 100,000 population far below the World Health Organization recommendation of 8-12 radiologists per 100,000 population. Furthermore, the distribution of mammography units is highly unequal, with the majority located in private healthcare facilities in major cities, leaving public hospitals in rural areas with limited or no mammography capabilities. The Indonesian Ministry of Health reports that there are approximately 500 mammography units nationwide, but many are non-functional due to maintenance issues or lack of trained operators.

Timor-Leste, one of Southeast Asia's youngest nations, faces even more profound challenges in healthcare delivery. The World Bank (2023) reports that Timor-Leste's healthcare system remains in the early stages of development, with limited infrastructure, a severe shortage of trained healthcare professionals, and minimal access to advanced diagnostic technologies. According to the Timor-Leste Ministry of Health, the country has only three radiologists serving a population of approximately 1.3 million people, and mammography services are available only at the national hospital in the capital city, Dili, with a single functional mammography unit. There are no radiation oncology services available within the country, meaning that women diagnosed with breast cancer must seek treatment abroad or forgo it entirely. The absence of population-based cancer registries makes accurate epidemiological data difficult to obtain, but hospital-based studies suggest that over 90% of breast cancer cases are diagnosed at advanced stages when curative treatment is no longer possible. Women presenting with symptomatic breast cancer in Timor-Leste often have large, palpable tumors that have already spread to axillary lymph nodes, dramatically reducing treatment options and survival probabilities. The combination of limited infrastructure, scarce human resources, and absent screening programs creates an urgent need for innovative approaches that can leverage available resources more effectively.

The convergence of these factors the global burden of breast cancer, the interpretive challenges of mammography, the potential of deep learning to improve diagnostic accuracy, and the specific healthcare constraints in Indonesia and Timor-Leste provides the motivation for this research. This study aims to develop and validate a deep learning-based system for breast cancer detection using mammograms, with a specific focus on the populations of Timor-Leste and Indonesia. The research employs a comparative evaluation framework using two datasets: a primary dataset of 110 mammogram images acquired from RSUD Buleleng in Indonesia and DMC Dili in Timor-Leste, and the public CBIS-DDSM dataset comprising approximately 7,100 images. This represents the first documented effort to develop AI-based breast cancer detection tools specifically for these populations. Local validation is crucial because deep learning models trained on Western populations may not generalize well to Southeast Asian populations due to differences in breast density distributions, body habitus, and potentially tumor characteristics (Jaamour et al., 2023). Studies have shown that Asian women tend to have denser breasts and different tumor biology (higher prevalence of triple-negative and HER2-positive subtypes) compared to Western populations, potentially affecting mammographic appearance and model performance (Maskarinec et al., 2006; Lin et al., 2019). By applying identical methodologies to both datasets, this study enables direct comparison of model performance across different data scales and populations, providing insights into the relationship between dataset size and achievable accuracy while establishing essential baseline data for future investigations in these underserved populations.

1.2 Problem Identification

Based on the comprehensive background presented above, three critical and interconnected problems can be identified that collectively hinder effective breast cancer detection and diagnosis in Timor-Leste and Indonesia. These problems span healthcare infrastructure limitations, diagnostic accuracy challenges, and data scarcity issues that together create significant barriers to improving breast cancer outcomes in these populations. Understanding these problems in their full complexity is essential for developing appropriate and effective solutions.

First, access to screening services is severely limited. In Indonesia, healthcare resources are concentrated in urban areas, leaving rural populations with minimal access to mammography equipment and trained radiologists. The country has only 1,200 radiologists serving 270 million people (0.44 per 100,000 population), far below the WHO recommendation of 8-12 per 100,000. In Timor-Leste, the situation is worse, with only three radiologists and a single functional mammography unit located in the capital Dili, making screening practically inaccessible for most rural women. Consequently, over 70% of patients in Indonesia and an estimated 90% in Timor-Leste are diagnosed at advanced stages when prognosis is poor.

Second, diagnostic inaccuracies persist in mammogram interpretation. Higher breast density in Asian populations can obscure lesions and reduce mammography sensitivity. Overlapping tissue can mimic malignancies while subtle cancer signs are easily overlooked. Significant inter-reader variability among radiologists further affects accuracy, and with few experienced breast imagers in both countries, general radiologists with limited training often interpret mammograms, increasing error risk.

Third, annotated mammography data from target populations is severely scarce. Deep learning models require thousands of examples for robust performance, yet the primary dataset comprises only 110 images with just 23 malignant cases. This scarcity causes overfitting, class imbalance bias toward benign predictions, and insufficient exposure to diverse malignant presentations. Consequently, no validated AI tools exist specifically for breast cancer detection in Timor-Leste and Indonesia.

These three problems form a self-reinforcing cycle. Limited screening access leads to late-stage diagnosis, reducing early-stage cancer cases for dataset development. Diagnostic inaccuracies undermine confidence in mammography. Data scarcity prevents locally validated AI tools that could address both access and accuracy problems. Breaking this cycle requires systematic data acquisition, model development, and validation the focus of this research.

1.3 Problem Scope

This study is scoped to address the identified problems through a focused investigation of deep learning-based breast cancer classification, with specific attention to the constraints and opportunities presented by the target populations in Timor-Leste and Indonesia. The scope is carefully defined to ensure that the research objectives are achievable while maintaining scientific rigor and maximizing the potential impact of the findings.

The primary focus is the development and evaluation of deep learning models based on the ResNet-50 architecture with transfer learning for breast cancer detection in mammograms. ResNet-50 was specifically chosen for several compelling reasons. First, its residual learning framework enables the training of deeper networks without suffering from the vanishing gradient problem, allowing the model to learn more complex and discriminative features from mammographic images (He et al., 2016). Second, ResNet-50 has demonstrated state-of-the-art performance in numerous medical imaging applications, including breast cancer detection, making it a well-established and reliable choice for this task (Rahman et al., 2023). Third, the availability of pretrained weights on ImageNet facilitates effective transfer learning, which is essential given the limited size of the primary dataset.

The study employs a comparative evaluation framework using two datasets with identical methodology. The primary dataset consists of 110 mammogram images acquired from RSUD Buleleng in Indonesia and DMC Dili in Timor-Leste, used to develop a binary classification model (Benign/Malignant) and establish the first documented baseline for AI-based breast cancer detection in these populations. The secondary dataset utilizes the public CBIS-DDSM dataset, one of the largest and most widely used mammography resources. After preprocessing and filtering, approximately 7,100 images with binary labels (benign/malignant) are used. This dataset provides sufficient scale for robust model training and enables assessment of model generalization on larger, more diverse data from a different population. By applying identical techniques to both datasets, this framework enables direct comparison of model behavior across different data scales and provides insights into the relationship between dataset size and achievable performance.

The technical scope encompasses three transfer learning strategies (frozen backbone, partial fine-tuning, and full fine-tuning), super aggressive augmentation techniques applied specifically to malignant cases, weighted random sampling with a 10x multiplier for malignant cases, and focal loss with gamma tuning up to 5.0. Model evaluation focuses on accuracy, precision, recall, specificity, F1-score, and AUC-ROC, with particular emphasis on malignant recall as the primary metric of clinical utility. The scope also includes development of a functional web application using Gradio that enables real-time inference and generates Grad-CAM visualizations to enhance model explainability.

1.4 Research Questions

This research systematically and ensure that all aspects of the identified problems are addressed, the following research questions are formulated. These questions are designed to progressively build understanding from establishing baseline performance to investigating specific techniques and finally to deriving quantitative guidance for future work:

1. What extent can a ResNet-50 model with transfer learning, trained using super aggressive augmentation, weighted random sampling, and focal loss, effectively classify mammograms from Timor-Leste and Indonesia as benign or malignant, given the limited dataset size (110 images) and severe class imbalance (87 benign versus 23 malignant cases)?
2. How do super aggressive augmentation (nine techniques), weighted random sampling (10x multiplier for malignant cases), and focal loss (gamma tuning up to 5.0) impact model performance, particularly in improving malignant recall on the small and imbalanced primary dataset?
3. What is the minimum number of malignant samples required to achieve clinically acceptable malignant recall exceeding 80% using the specified techniques?

1.5 Research Objectives

In alignment with the research questions formulated above, this study aims to achieve the following objectives, each carefully designed to contribute meaningfully to the understanding and advancement of AI-based breast cancer detection in Timor-Leste and Indonesia.

1. The first objective is to develop and validate a ResNet-50-based deep learning model employing transfer learning, super aggressive augmentation, weighted random sampling, and focal loss to classify mammograms from Timor-Leste and Indonesia as benign or malignant. This objective directly addresses the critical gap in locally validated AI tools for these populations and establishes the first documented baseline for AI-based breast cancer detection in Timor-Leste and Indonesia.
2. The second objective is to evaluate the effectiveness of super aggressive augmentation, weighted random sampling (10x multiplier for malignant cases), and focal loss (gamma tuning from 2.5 to 5.0) in improving malignant recall on the small and imbalanced primary dataset. By systematically applying these techniques and observing their effects on model performance, this objective aims to provide evidence-based guidance for researchers facing similar challenges of class imbalance and limited data in medical imaging applications.
3. The third objective is to determine the minimum number of malignant samples required in the primary dataset to achieve clinically acceptable malignant recall exceeding 80%. By analyzing the relationship between dataset size and model performance, this objective aims to provide concrete, quantitative recommendations for future data collection efforts in similar contexts.

1.6 Research Benefits

This research is expected to yield significant benefits across theoretical and practical dimensions, providing contributions to academic knowledge and tangible applications for healthcare practice in Timor-Leste and Indonesia.

Theoretical Benefits

This research offers two main theoretical contributions :

a) First, it provides methodological insights into handling small, imbalanced medical datasets by demonstrating the effectiveness of combining super aggressive augmentation, weighted random sampling, and focal loss. These insights extend beyond breast cancer detection to other medical imaging applications where pathological cases are rare.
b) Second, it establishes quantitative data requirements by showing the relationship between malignant sample size and achievable recall, advancing understanding of dataset size requirements for deep learning in resource-constrained settings.

Practical Benefits

This research offers two practical benefits :

a) First, it enables potential improvement in diagnostic accuracy through locally validated AI tools that can assist healthcare providers in interpreting mammograms. In settings where access to expert radiologists is severely limited, such tools could help prioritize urgent cases and reduce diagnostic errors.
b) Second, it facilitates capacity building and technology transfer by demonstrating the feasibility of conducting AI research in resource-constrained settings,

establishing collaborative partnerships between academic and healthcare institutions, and providing a functional web application with Grad-CAM visualization for future clinical validation studies.