J. Adv. Anal. Healthc. Manag.
Vol. 06, No 03, pp. 1, 11.
Published by TensorGate
OPEN ACCESS
Journal of Advanced Analytics
in Healthcare Management

Development and Evaluation of Deep Learning-Based Diagnostic Framework for Accurate Differentiation Between Benign and Malignant Breast Tumors Using Histopathological Imaging Data

Weng Chan Hackensack Meridian School of Medicine

06, July, 2023

Recieved: 5, June, 2023. | Revised: August 2023. | Published: September 2023

Abstract

Breast cancer is the most prevalent form of cancer affecting women, characterized by abnormal cell division in breast tissue. Not all tumors pose a significant threat to life, as they can be either benign or malignant. Determining the nature of a tumor—whether benign or malignant—is necessary for guiding appropriate treatment strategies and ensuring patient well-being. In this scenario, medical imaging is becoming a key application of artificial intelligence in healthcare to improve diagnostic accuracy. This study presents the development and evaluation of a deep learning-based tool designed to distinguish between benign and malignant breast tumors using histopathological images. The research used the Breast Cancer Histopathological Image Classification (BreakHis) dataset, which contains 7,909 images from 82 patients across four magnification levels (40X, 100X, 200X, and 400X). The goal is to enhance the accuracy and scalability of breast cancer diagnosis through the application of computer vision and deep learning models. Data preprocessing in this study involved resizing images to a uniform size and applying data augmentation techniques, including random brightness adjustments, flips, and rotations. These methods were employed to improve the models' ability to handle variations in image orientation and lighting. The study evaluated several deep learning models, including a Convolutional Neural Network (CNN) and 4 transfer learning models including MobileNetV3, EfficientNetB1, VGG16, and ResNet50V2. The findings showed that EfficientNetB1 achieved the highest performance, with a ROC-AUC score of 0.8767, demonstrating strong potential for distinguishing between benign and malignant cases. However, the model also produced a relatively high number of false positives, which is a concern for clinical application. The CNN, although simpler, achieved the highest accuracy, suggesting its potential for use in resource-limited settings. The findings indicate that deep learning models can be applied in breast cancer diagnosis. Further refinement is, however, needed to reduce false positives and ensure the models' reliability in clinical practice.

Keywords: *artificial intelligence, breast cancer, computer vision, deep learning, medical imaging, tumor classification*

1 Introduction

Breast cancer begins in the epithelial tissue of the mammary gland and is a serious health concern because it can proliferate uncontrollably and metastasize. This type of cancer typically starts in the cells lining the milk ducts (ductal carcinoma) or the lobules (lobular carcinoma) of the breast. The initial transformation of normal cells into cancerous ones results in a malignant tumor, which, in its early stages, may remain localized (in situ). Ductal carcinoma in situ (DCIS) is the most common form of non-invasive breast cancer, confined to the ductal system and not yet capable of spreading to other tissues. Lobular carcinoma in situ (LCIS), although rarer, similarly remains within the lobules and, while not directly life-threatening, increases the future risk of invasive breast cancer (Elmore et al., 2005) (Carey et al., 2006). In 2022, the global incidence of breast cancer among

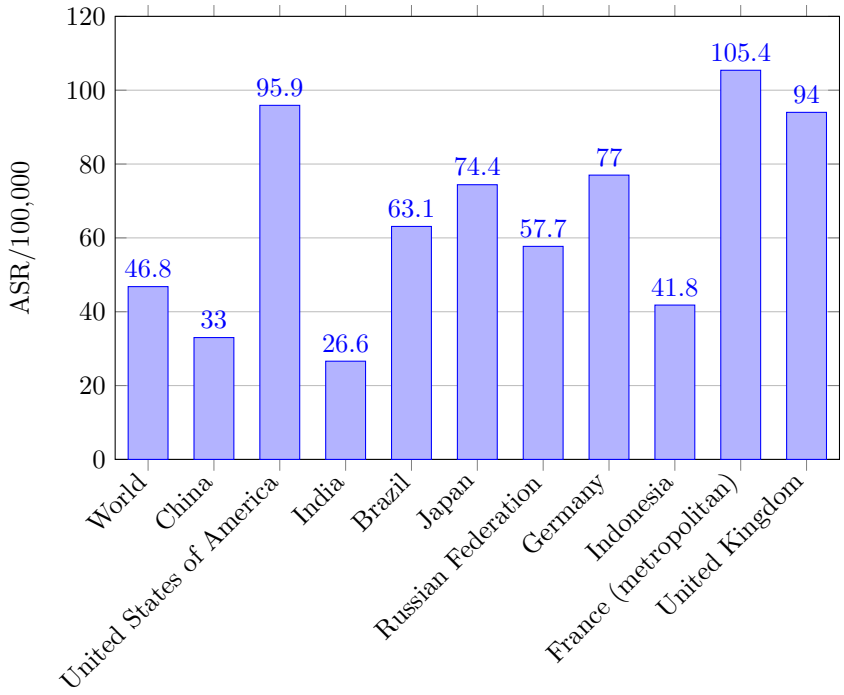

Figure 1: Breast Cancer Rates (ASR/100,000) by Country in 2022.
Source: *World Cancer Research Fund International* (World Cancer Research Fund International, 2024)

women reached a staggering 2,296,840 cases, with an age-standardized rate (ASR) of 46.8 per 100,000 women. France recorded the highest incidence rate, with an ASR of 105.4 per 100,000 women, indicating a significant public health challenge (World Cancer Research Fund International, 2024). Close behind was Cyprus, highlighting regional disparities in breast cancer rates. The United States also reported a notably high incidence rate at 95.9, despite having a smaller population compared to countries like China and India. China, while having the highest number of cases at 357,161, exhibited a lower incidence rate of 33.0 per 100,000, likely due to its vast population. These variations in breast cancer incidence rates across countries underscore the influence of factors such as healthcare access, screening practices, lifestyle, and genetic predispositions.

The global burden of breast cancer mortality in 2022 was also significant, with 666,103 deaths reported worldwide. Fiji and Jamaica had the highest mortality

Risk Factor	Description	Notes
Getting older	The risk for breast cancer increases with age.	Most breast cancers are diagnosed after age 50.
Genetic mutations	Women who have inherited changes (mutations) to certain genes, such as BRCA1 and BRCA2, are at higher risk of breast and ovarian cancer.	Inherited gene changes significantly increase risk.
Reproductive history	Starting menstrual periods before age 12 and starting menopause after age 55 expose women to hormones longer, raising their risk of getting breast cancer.	Longer exposure to hormones increases risk.
Having dense breasts	Dense breasts have more connective tissue than fatty tissue, which can sometimes make it hard to see tumors on a mammogram.	Women with dense breasts are more likely to get breast cancer.
Personal history of breast cancer or certain noncancerous breast diseases	Women who have had breast cancer are more likely to get breast cancer a second time.	Some noncancerous breast diseases are associated with higher risk.
Family history of breast or ovarian cancer	A woman's risk for breast cancer is higher if she has a first-degree relative or multiple family members with breast or ovarian cancer.	Having a first-degree male relative with breast cancer also raises risk.
Previous treatment using radiation therapy	Women who had radiation therapy to the chest or breasts before age 30 have a higher risk of getting breast cancer later in life.	Early exposure to radiation therapy increases risk.
Exposure to the drug diethylstilbestrol (DES)	DES was given to some pregnant women in the United States between 1940 and 1971 to prevent miscarriage.	Women who took DES or whose mothers took DES have a higher risk of getting breast cancer.

Table 1: Overview of key risk factors for breast cancer, including age, genetic mutations, reproductive history, breast density, personal and family medical history, previous radiation therapy, and exposure to diethylstilbestrol (DES).

rates, emphasizing the critical need for improved healthcare interventions in these regions. Among the larger nations, India recorded the highest number of deaths at 98,337, with an ASR of 13.7 per 100,000, reflecting both the high incidence and the challenges in effective treatment. Indonesia and Nigeria, despite having lower incidence rates, showed relatively high mortality rates at 14.4 and 26.8 per 100,000, respectively, pointing to potential issues in early detection and treatment access. The disparity between the incidence and mortality rates in countries like China and the United States suggests differences in healthcare infrastructure and the effectiveness of treatment protocols, with the U.S. having a more advanced healthcare system that likely contributes to better survival rates despite high incidence.

When these malignancies progress, they can become invasive, breaking through the basement membrane and infiltrating surrounding breast tissues. Invasive breast cancers are classified primarily into invasive ductal carcinoma (IDC), which constitutes 70-80% of cases, and invasive lobular carcinoma (ILC), which accounts for 10-15%. IDC begins in the milk ducts and spreads to the surrounding tissue, often forming a palpable mass. ILC, originating in the lobules, typically spreads in a linear fashion, which may not form a distinct lump, making it less detectable through standard imaging techniques. The prognosis for these invasive cancers is heavily dependent on the stage at diagnosis, molecular characteristics, and the presence or absence of metastasis.

Breast cancer's ability to metastasize is what makes it dangerous. Metastasis involves cancer cells breaking away from the original tumor site and traveling through the bloodstream or lymphatic system to other parts of the body. The

Type of Breast Cancer	Description
Non-invasive breast cancers	Non-invasive breast cancers, also known as stage 0 or carcinomas in situ, are abnormal cells that remain in the area of the breast where they first formed. They are not generally life-threatening but can become invasive if left untreated. Examples include Ductal Carcinoma in Situ (DCIS) and Lobular Carcinoma in Situ (LCIS).
Invasive breast cancers	Invasive breast cancers do not stay in their original sites; they invade nearby breast tissue, lymph nodes, and distant organs. Examples include Invasive Ductal Carcinoma (IDC) and Invasive Lobular Carcinoma (ILC).

Table 2: Comparison of non-invasive and invasive breast cancers, highlighting key differences in their progression and impact.

first site of spread is often the axillary lymph nodes, located under the arm. The involvement of these lymph nodes is a critical factor in staging breast cancer and determining the prognosis. When breast cancer spreads to distant organs—commonly the lungs, liver, bones, or brain—it is classified as metastatic or stage IV. At this stage, the disease is generally considered incurable, and treatment focuses on extending life and improving the quality of life rather than cure.

The risk of developing breast cancer is influenced by several factors, some of which are non-modifiable (Sun et al., 2017). Age is one of the most significant risk factors, with the incidence of breast cancer increasing with advancing age (Hortobagyi, 1998). The majority of breast cancer cases are diagnosed in women over the age of 50. Genetic factors also play a crucial role; mutations in the BRCA1 and BRCA2 genes significantly increase the risk of both breast and ovarian cancers. Women with these mutations may have up to an 85% lifetime risk of developing breast cancer (Houghton and Hankinson, 2021). Other genes, such as TP53 and PTEN, are also associated with hereditary breast cancer syndromes, albeit to a lesser extent.

Type	Description
Ductal Carcinoma in Situ (DCIS)	The most common type of non-invasive breast cancer, DCIS starts in the cells lining the milk ducts of the breast. It is generally considered an early form of breast cancer.
Lobular Carcinoma in Situ (LCIS)	A rarer non-invasive breast cancer that starts in the cells lining the breast lobules. LCIS is not generally thought to progress to invasive cancer but increases the risk of breast cancer in the future.

Table 3: Overview of common types of non-invasive breast cancers, including DCIS and LCIS.

Family history is another important risk factor. Women with a first-degree relative (mother, sister, or daughter) who has been diagnosed with breast cancer have a higher risk of developing the disease themselves. The risk increases further if multiple relatives are affected or if the cancer occurred at a young age. Moreover, women with a personal history of breast cancer are at an increased risk of developing a second primary breast cancer, either in the same breast or the opposite one. Certain benign breast conditions, such as atypical ductal hyperplasia or LCIS, are also associated with an elevated risk of developing invasive breast cancer (Key et al., 2001) (MacMahon et al., 1973).

Hormonal factors contribute to breast cancer risk as well. Women who experience early menarche (before age 12) or late menopause (after age 55) have a longer lifetime exposure to estrogen, which is associated with an increased risk of breast cancer. Similarly, nulliparity (having no children) or having the first child

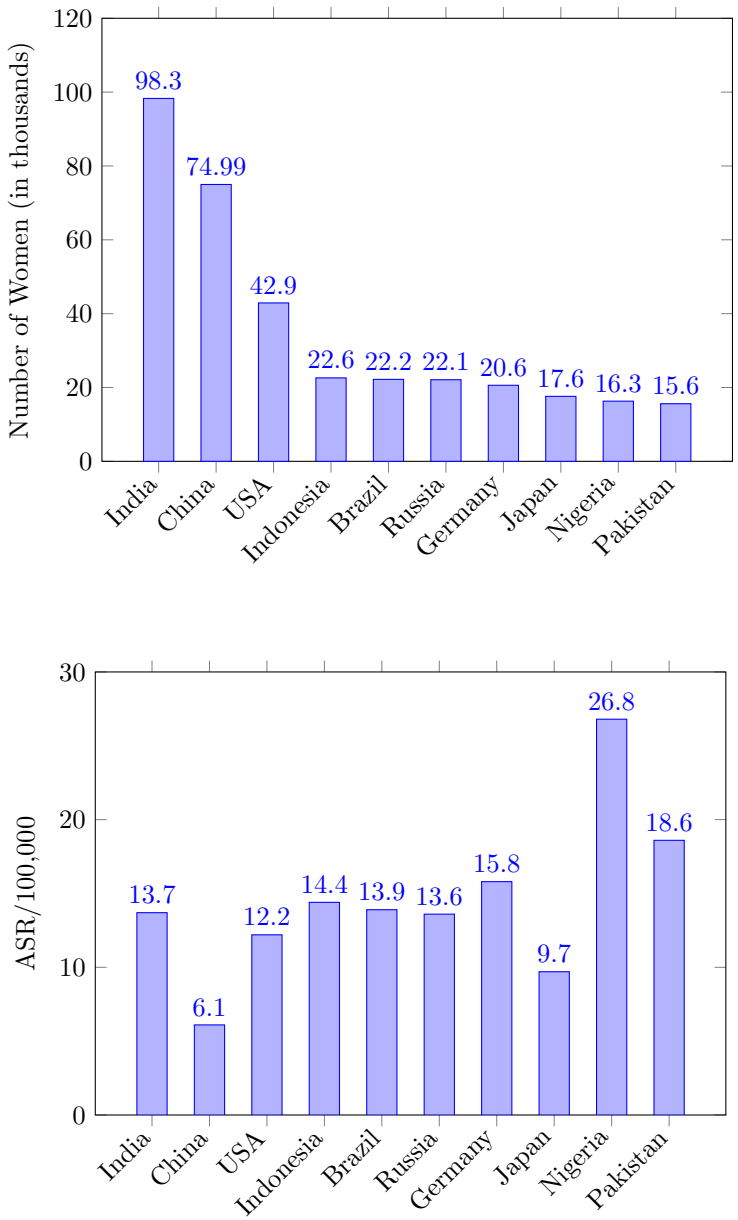

Figure 2: Number of women (in thousands) and ASR/100,000 for the top 10 countries.
Source: *World Cancer Research Fund International* (World Cancer Research Fund International, 2024)

after age 30 are risk factors, likely due to prolonged exposure to estrogen without the interruption of pregnancy. The use of hormone replacement therapy (HRT) during menopause combined estrogen-progestin therapy, has also been linked to an increased risk of breast cancer, though this risk declines after discontinuation of HRT (Momenimovahed and Salehiniya, 2019).

Environmental and lifestyle factors, while less significant than genetic and hormonal factors, still play a role in breast cancer risk. Exposure to ionizing radiation during adolescence and early adulthood, increases the risk of developing breast cancer later in life. This is relevant for women who received radiation therapy to the chest for conditions such as Hodgkin's lymphoma. Lifestyle factors such as alcohol consumption, obesity, and lack of physical activity have also been associated with an increased risk of breast cancer. Alcohol, even in moderate amounts, has been shown to increase breast cancer risk, possibly due to its effects on estrogen metabolism. Obesity after menopause, increases the risk due to the higher levels of estrogen produced by adipose tissue. Conversely, regular physical

activity is associated with a reduced risk of breast cancer, likely due to its effects on hormone levels, body weight, and immune function (Prat and Perou, 2011).

Breast cancer can be categorized into several subtypes based on the molecular characteristics of the tumor the presence or absence of hormone receptors (estrogen and progesterone receptors) and the HER2 protein. These subtypes are critical in determining the most appropriate treatment strategy. Hormone receptor-positive (HR-positive) breast cancers, which express estrogen or progesterone receptors, are the most common subtype. These cancers typically have a better prognosis and respond well to hormone therapies, such as tamoxifen or aromatase inhibitors, which either block the hormone receptors or decrease estrogen production in the body (Stephens et al., 2012).

Type of Invasive Cancer	**Description**
Invasive Ductal Carcinoma (IDC)	The most common type of invasive breast cancer, originating in the milk ducts. IDC typically forms a hard mass in the breast and can spread to other tissues.
Invasive Lobular Carcinoma (ILC)	The second most common type, originating in the breast lobules. ILC often presents differently from IDC, with tumors growing in lines rather than lumps.
Hormone Receptor (HR)–Positive	Breast cancer classified as HR-positive is fueled by estrogen and/or progesterone and has a good prognosis when treated with hormone therapies.
HER2-Positive	Breast cancer that overexpresses the HER2 protein, leading to aggressive tumor growth. HER2-positive breast cancer is treated with targeted therapies like trastuzumab.
Triple-Negative Breast Cancer (TNBC)	TNBC lacks estrogen, progesterone, and HER2 receptors, making it more challenging to treat. TNBC is often more aggressive and diagnosed at later stages.

Table 4: Common types of invasive breast cancers, including their molecular characteristics and implications for treatment.

HER2-positive breast cancers, characterized by an overexpression of the HER2 protein, tend to be more aggressive but respond to targeted therapies such as trastuzumab (Herceptin). The development of trastuzumab in the late 1990s was a significant breakthrough in breast cancer treatment, dramatically improving the prognosis for patients with HER2-positive breast cancer. More recently, the recognition of HER2-low breast cancers has expanded the potential for targeted therapies in a broader range of patients.

Triple-negative breast cancer (TNBC), which lacks estrogen, progesterone, and HER2 receptors, is a challenging subtype to treat due to the absence of targeted therapies. TNBC is more common in younger women, African American women, and those with BRCA1 mutations. It tends to be more aggressive and has a higher likelihood of recurrence compared to other breast cancer subtypes. Treatment for TNBC primarily involves surgery, chemotherapy, and radiation, but recent research is exploring new avenues such as immunotherapy and PARP inhibitors, which have shown promise in clinical trials (Waks and Winer, 2019).

The treatment of breast cancer is highly individualized, depending on the stage of the disease, the molecular characteristics of the tumor, and the patient's overall health. Early-stage breast cancer is typically treated with surgery, which may involve a lumpectomy (removal of the tumor with a margin of healthy tissue) or mastectomy (removal of the entire breast). Radiation therapy is often used in conjunction with surgery to reduce the risk of local recurrence. For patients with HR-positive breast cancer, hormone therapy is a cornerstone of treatment in postmenopausal women, where aromatase inhibitors are often used to reduce estrogen levels.

Chemotherapy is a common treatment for more advanced breast cancers or

those with a high risk of recurrence. It involves the use of cytotoxic drugs that target rapidly dividing cells, including cancer cells. Chemotherapy can be administered before surgery (neoadjuvant chemotherapy) to shrink the tumor, making it more amenable to surgical removal, or after surgery (adjuvant chemotherapy) to eliminate any remaining cancer cells. While effective, chemotherapy is associated with significant side effects, including fatigue, nausea, hair loss, and an increased risk of infections due to its impact on the immune system.

Despite significant advances in treatment, metastatic breast cancer remains a major challenge. Once breast cancer has spread to distant organs, it is generally considered incurable, with treatment focusing on prolonging survival and maintaining quality of life. The management of metastatic breast cancer involves a combination of systemic therapies, including hormone therapy, chemotherapy, targeted therapy, and immunotherapy, tailored to the specific characteristics of the tumor. Palliative care, aimed at managing symptoms and improving the patient's comfort, is also a crucial component of care for patients with metastatic disease.

The prognosis for breast cancer patients has improved significantly over the past few decades, largely due to advances in early detection, improved surgical techniques, and the development of more effective therapies. However, breast cancer remains a leading cause of cancer-related death among women worldwide, underscoring the need for continued research and public health efforts. Regular screening with mammography is the most effective method for early detection of breast cancer, allowing for diagnosis at a stage when the disease is most treatable. Public awareness campaigns and education about the importance of screening, risk factors, and the signs and symptoms of breast cancer are vital in reducing the burden of this disease (Waks and Winer, 2019).

Histopathological imaging involves the microscopic examination of tissue samples to identify abnormalities at the cellular level in the context of disease diagnosis such as cancer. In breast cancer diagnosis, histopathological images are critical as they provide detailed visual information about the tissue structure, cell morphology, and the presence of cancerous cells. These images are typically obtained from biopsy samples, which are stained to highlight different cellular components, making it easier to identify key features that indicate malignancy. The interpretation of these images requires a high level of expertise, as subtle differences in cell structure and tissue organization can distinguish between benign and malignant tumors.

The data derived from histopathological imaging is rich in detail, capturing the complexity of tissue architecture and cellular variations. This data can be extensive in large-scale studies or clinical settings where multiple samples are analyzed across different magnification levels. For example, the BreakHis dataset, which is commonly used in research, contains thousands of histopathological images of breast tissue at various magnifications, offering a comprehensive view of both benign and malignant cases. Such datasets are used for developing and training machine learning models that aim to assist in the diagnosis process by automating the identification and classification of cancerous cells.

Analyzing histopathological imaging data through computational methods, such as deep learning, has the potential to significantly enhance diagnostic accuracy and efficiency. Machine learning models can be trained on these images to learn patterns and features associated with malignancy, enabling them to classify new samples with high precision. The use of such models can reduce the variability and subjectivity associated with human interpretation, providing more consistent and reliable diagnoses. As the availability of large, annotated histopathological datasets continues to grow, the integration of artificial intelligence in histopathology is likely to become a standard practice in clinical diagnostics.

2 Dataset

The Breast Cancer Histopathological Image Classification (BreakHis) dataset constitutes a significant resource in the domain of digital pathology in the classification and analysis of breast cancer histopathological images (Adeshina et al.,

2018). Comprising a total of 9,109 microscopic images, the dataset was meticulously compiled from 82 patients, capturing the intricate histological variations inherent in breast tumor tissues. The images are organized across four distinct magnification factors—40X, 100X, 200X, and 400X—each magnification providing a different level of detail, critical for nuanced analysis. This variation in magnification is used for understanding the morphological features at different scales for aiding in the development of more robust classification models. The images are stored in a standardized format—700x460 pixels, 3-channel RGB, 8-bit depth per channel, and in PNG format—ensuring consistency in data quality and facilitating reproducibility in research (Bayramoglu et al., 2016).

A key aspect of the BreakHis dataset is its categorization of breast tumor tissues into two primary classes: benign and malignant. Within the benign category, the dataset includes four histologically distinct subtypes: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA). Each of these subtypes presents unique histopathological characteristics, making the benign class diverse in its representation. For instance, adenosis is characterized by the proliferation of glandular tissue, often mimicking carcinoma for presenting a challenge in differential diagnosis. Fibroadenoma, the most common benign breast tumor, exhibits a mixture of stromal and epithelial elements, which can vary in cellularity and morphology. Phyllodes tumors, although classified as benign, have the potential for malignant transformation, exhibiting a leaf-like stromal growth pattern. Tubular adenomas are rare and are distinguished by their well-defined tubular structures (Spanhol et al., 2015).

The malignant category, representing the more clinically significant aspect of the dataset, encompasses four distinct subtypes: ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC). Ductal carcinoma, specially the invasive type (IDC), is the most prevalent form of breast cancer, characterized by its ability to breach the basement membrane and invade surrounding tissues. Lobular carcinoma, often more subtle in its histological presentation, exhibits a discohesive growth pattern due to the loss of E-cadherin, a key cell adhesion molecule. Mucinous carcinoma, also known as colloid carcinoma, is less common and is identified by the presence of extracellular mucin pools, which can significantly alter the tumor microenvironment. Papillary carcinoma, although rare, is distinguished by its papillary structures lined by epithelial cells, often surrounded by a fibrovascular core.

The dataset's structure allows for comprehensive analysis across different magnification levels, providing insights into how histopathological features vary with scale. For instance, at lower magnifications (40X), the overall architecture and organization of the tissue can be observed, which is crucial for identifying patterns such as ductal or lobular arrangements. At higher magnifications (400X), cellular details such as nuclear morphology, mitotic figures, and cellular atypia become more prominent, aiding in the differentiation between benign and malignant tissues. This multi-scale approach is essential for the development of machine learning models that can accurately classify breast tumors based on their histopathological features.

In addition to the classification of tumors, the dataset also includes metadata embedded within the filenames of the images. This metadata provides crucial information about the biopsy method, tumor class, tumor type, patient identification, and magnification factor. For example, the filename "SOB_B_TA-14-4659-40-001.png" reveals that the image represents a benign tubular adenoma, collected using the SOB method from patient sample 14-4659 at 40X magnification. Such detailed metadata not only facilitates the organization and retrieval of images but also allows for more nuanced analyses, such as patient-specific studies or comparisons across different biopsy methods (Zhu et al., 2019) (Yamlome et al., 2020).

Each subtype within the benign and malignant categories presents unique challenges for classification, both for pathologists and for automated systems. For instance, the overlap in features between certain benign and malignant subtypes, such as between adenosis and low-grade ductal carcinoma, necessitates the development of sophisticated algorithms capable of capturing subtle differences in histological features. Moreover, the inclusion of rare subtypes, such as mucinous

carcinoma and phyllodes tumor, provides an opportunity to explore the boundaries of current classification systems and to potentially identify new biomarkers or features that can improve diagnostic accuracy.

The dataset's composition, with 2,480 benign and 5,429 malignant samples, reflects the clinical reality where malignant cases are more prevalent, thereby aligning the dataset's distribution with real-world scenarios. This distribution is crucial for training machine learning models, as it ensures that the models are exposed to a representative sample of both benign and malignant cases for improving their generalizability and performance in clinical settings. Additionally, the large sample size of the dataset supports robust statistical analyses, enabling researchers to validate their models across a diverse set of images and conditions (Wei et al., 2017).

3 Data preprocessing

The dataset includes images taken at four different magnifications: 40X, 100X, 200X, and 400X. It has been observed that training models on images with a magnification of at least 200X often leads to better results, as these higher magnifications provide more detail, which is useful for distinguishing between benign and malignant tissues.

The dataset presents a challenge due to its class imbalance, with a greater number of malignant images compared to benign ones. This imbalance can make accuracy a misleading performance metric, as a model that simply predicts the majority class (malignant) would achieve an accuracy of 65.94% without correctly identifying any benign cases. However, such a model would have an ROC-AUC score of only 50%, indicating it performs no better than random guessing. This illustrates the importance of using more reliable metrics like ROC-AUC, which measure the model's ability to differentiate between classes.

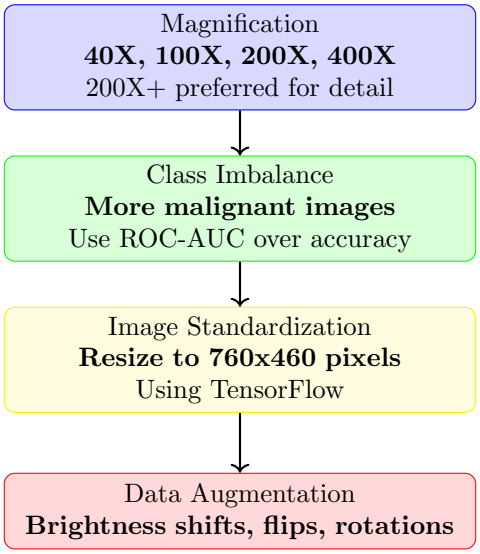

Figure 3: Preprocessing Stages

The preprocessing of the images involves converting them into a format that machine learning models can work with. Using the TensorFlow image module, the PNG images are loaded, decoded into 3D tensors representing pixel values in RGB channels, and resized to a uniform size (typically 760x460 pixels). Standardizing the image size ensures that all images are consistent, which is important for model training.

Data augmentation techniques are applied to increase the diversity of the training data and improve model robustness. These techniques include random brightness shifts to simulate different lighting conditions, as well as random horizontal and vertical flips and rotations to help the model recognize features regardless of orientation. These steps help in making the model more generalizable and less prone to overfitting.

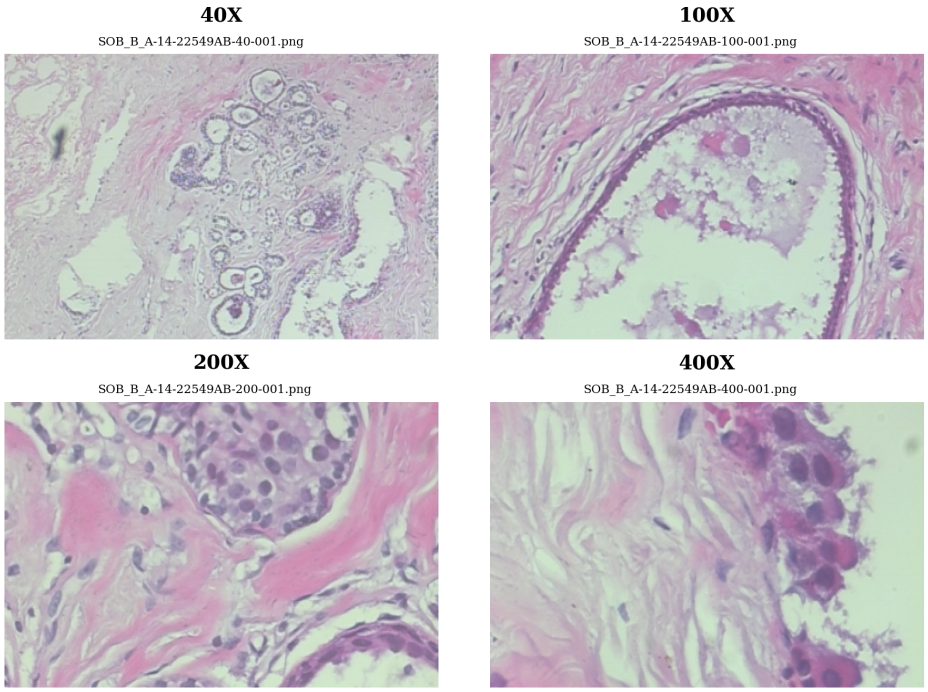

Figure 4: Sample images at different magnification factors

The preprocessing of the BreakHis dataset involves selecting appropriate magnification levels, addressing class imbalance, standardizing image sizes, and applying data augmentation. These processes are essential for preparing the data for effective use in machine learning models that can distinguish between benign and malignant breast cancer tissues.

4 Deep Learning Methods

MobileNetV3, EfficientNetB1, VGG16, and ResNet50V2 represent some of the most influential convolutional neural network (CNN) architectures in the field of deep learning for image classification tasks. Each of these models brings specific advantages and trade-offs in terms of accuracy, computational efficiency, and suitability for deployment in different environments, including resource-constrained settings. Understanding their architectural differences, mathematical formulations, and practical implications is essential for their effective application in various computer vision tasks, such as medical image analysis (Aceves-Fernandez, 2020).

4.1 MobileNetV3

MobileNetV3 is designed with a focus on efficiency, making it suitable for deployment on mobile devices and other environments with limited computational resources. It builds on the foundation of previous MobileNet versions, incorporating several advanced techniques to balance accuracy and latency. The core of MobileNetV3 is based on depthwise separable convolutions, a factorization of a standard convolution into a depthwise convolution and a pointwise convolution. This reduces the computational cost from $O(k^2 \cdot D_i \cdot D_o)$ to $O(k^2 \cdot D_i + D_i \cdot D_o)$, where k is the kernel size, D_i is the number of input channels, and D_o is the number of output channels.

Mathematically, a depthwise separable convolution can be expressed as:

$$y_{i,j,o} = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} x_{i+m,j+n,c} \cdot w_{m,n,c,o} + b_o$$

where $y_{i,j,o}$ is the output, $x_{i+m,j+n,c}$ is the input, $w_{m,n,c,o}$ represents the filter weights, and b_o is the bias term.

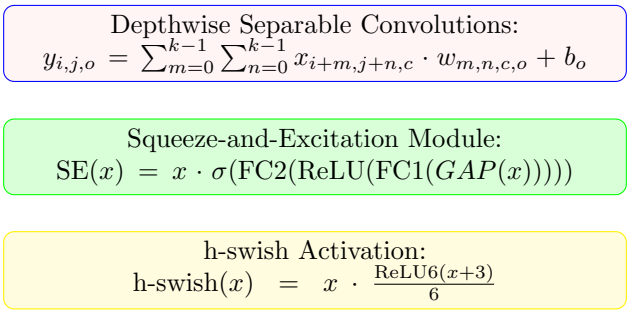

Figure 5: MobileNetV3 Key Components

MobileNetV3 also introduces novel elements such as the use of squeeze-and-excitation (SE) modules, which dynamically recalibrate feature maps, and the h-swish activation function, a computationally efficient approximation of the swish activation function. The SE module can be expressed as:

$$\text{SE}(x) = x \cdot \sigma(\text{FC2}(\text{ReLU}(\text{FC1}(GAP(x)))))$$

where GAP stands for global average pooling, and σ is the sigmoid function. The h-swish activation is given by:

$$\text{h-swish}(x) = x \cdot \frac{\text{ReLU6}(x+3)}{6}$$

These improvements enable MobileNetV3 to achieve higher accuracy than its predecessors while maintaining low latency, making it ideal for edge applications such as mobile devices.

4.2 EfficientNetB1

EfficientNetB1 is part of the EfficientNet family, which is designed to achieve state-of-the-art accuracy with a relatively low number of parameters and FLOPs (floating point operations per second). The key innovation behind EfficientNet is the compound scaling method, which uniformly scales the network's depth, width, and resolution using a set of predetermined scaling coefficients. The scaling can be mathematically formulated as:

$$\text{depth} = \alpha^d, \quad \text{width} = \beta^d, \quad \text{resolution} = \gamma^d$$

where d is the scaling factor, and α, β, and γ are constants determined through a grid search. This method enables EfficientNetB1 to strike a balance between computational efficiency and accuracy.

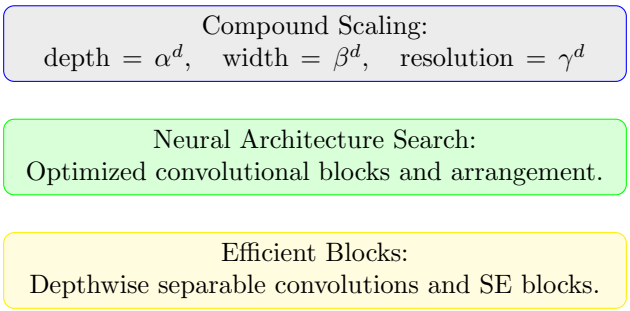

Figure 6: EfficientNetB1 Key Components

EfficientNetB1 employs a backbone architecture similar to that of MobileNetV3, including depthwise separable convolutions and SE blocks, but it uses a different network scaling strategy. The baseline EfficientNet architecture (EfficientNet-B0) was optimized using a neural architecture search (NAS), which automatically discovered the optimal convolutional blocks and their arrangement. EfficientNetB1

is a scaled-up version of this baseline, offering higher accuracy at the cost of increased computational requirements. The architecture's design makes it well-suited for tasks that require a good trade-off between accuracy and efficiency, such as medical image analysis where both factors are critical.

4.3 VGG16

VGG16 is one of the earlier deep CNN architectures that achieved significant success in image classification tasks in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014. The architecture of VGG16 is characterized by its simplicity and uniformity, consisting of 16 layers, including 13 convolutional layers and 3 fully connected layers. Unlike more recent architectures, VGG16 exclusively uses 3x3 convolutional kernels throughout the network, with the depth of the network providing its representational power.

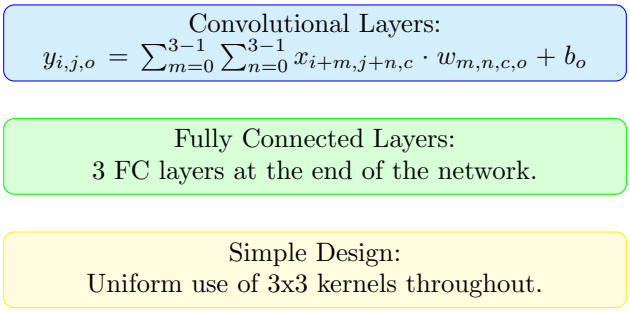

Figure 7: VGG16 Key Components

The convolutional operation in VGG16 can be expressed as:

$$y_{i,j,o} = \sum_{m=0}^{3-1} \sum_{n=0}^{3-1} x_{i+m,j+n,c} \cdot w_{m,n,c,o} + b_o$$

where the 3x3 filter size is applied across the depth of the input volume. Despite its relatively simple design, VGG16 has a large number of parameters, approximately 138 million, which makes it computationally expensive and memory-intensive. However, its straightforward design and high accuracy have made it a popular choice for transfer learning, where the pre-trained weights of VGG16 are used as a starting point for various downstream tasks.

One of the limitations of VGG16 is its large memory footprint and computational cost, which limits its deployment in environments with limited resources. However, its high accuracy and the ease of fine-tuning make it a tool for applications where computational resources are not a primary concern.

4.4 ResNet50V2

ResNet50V2 is a variant of the ResNet (Residual Networks) architecture, which introduced the concept of residual learning to address the degradation problem in deep networks. The degradation problem refers to the phenomenon where increasing the depth of a network leads to higher training error, contrary to the expectation that a deeper network should perform at least as well as a shallower one. ResNet solves this problem by introducing shortcut connections, or residual connections, which bypass one or more layers.

The basic building block of ResNet can be expressed as:

$$y = F(x, \{W_i\}) + x$$

where x is the input, y is the output, and $F(x, \{W_i\})$ represents the residual mapping to be learned. In ResNet50V2, each residual block consists of three layers of 1x1, 3x3, and 1x1 convolutions, with batch normalization (BN) and ReLU activation applied after each convolution. The shortcut connections enable

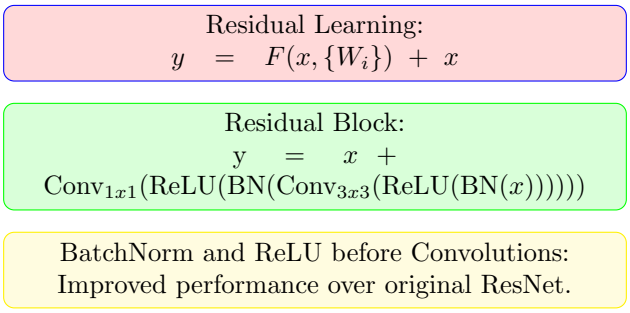

Figure 8: ResNet50V2 Key Components

the training of very deep networks by mitigating the vanishing gradient problem, as gradients can flow directly through the shortcut connections.

ResNet50V2 introduces a few modifications compared to the original ResNet architecture (ResNet50). It reorders the batch normalization and ReLU layers before the convolution layers in the residual blocks, which has been shown to improve performance. The architecture can be described mathematically as:

$$y = x + \text{Conv}_{1x1}(\text{ReLU}(\text{BN}(\text{Conv}_{3x3}(\text{ReLU}(\text{BN}(x))))))$$

This architecture is highly effective for a wide range of image classification tasks in scenarios where high accuracy is required without excessively increasing the number of parameters. ResNet50V2, with its 50 layers, strikes a balance between depth and computational efficiency, making it suitable for both research and practical applications.

Comparison and Practical Considerations

When choosing between MobileNetV3, EfficientNetB1, VGG16, and ResNet50V2, several factors must be considered, including the computational resources available, the required accuracy, and the specific application domain.

MobileNetV3 is advantageous in resource-constrained environments due to its lightweight design and low latency. It is well-suited for deployment on mobile devices or embedded systems where computational power and energy efficiency are critical. The use of depthwise separable convolutions and efficient activation functions ensures that MobileNetV3 can perform well even with limited resources.

EfficientNetB1, on the other hand, offers a good balance between efficiency and accuracy. The compound scaling method allows the architecture to be adapted to different resource constraints by scaling the depth, width, and resolution uniformly. This flexibility makes EfficientNetB1 a strong candidate for applications that require high accuracy but still need to consider computational cost, such as in medical image analysis where diagnostic accuracy is paramount.

VGG16, despite its simplicity and large number of parameters, remains a popular choice for transfer learning. Its high accuracy and straightforward architecture make it easy to fine-tune for specific tasks, although its large memory and computational requirements limit its use in environments with limited resources. VGG16 is often used as a benchmark model or in scenarios where pre-trained models are needed for tasks with ample computational resources.

ResNet50V2, with its residual connections and deep architecture, is ideal for tasks that require high accuracy and can afford the computational cost. The introduction of residual connections allows ResNet50V2 to train deeper networks without suffering from the degradation problem, making it suitable for complex tasks where a deep model is necessary. ResNet50V2 is often used in research and industry for tasks that require robust performance across a wide range of image classification problems.

5 Results

The results obtained from the series of experiments conducted with various convolutional neural network (CNN) architectures highlight the effectiveness and challenges associated with using different models for the classification of breast cancer

histopathological images. These experiments involved both CNNs and pre-trained models utilizing transfer learning techniques. The evaluation of these models was based on key performance metrics such as accuracy, loss, and the area under the receiver operating characteristic curve (ROC-AUC), with a special focus on the ability of the models to minimize false positives and false negatives, which are critical in medical diagnostics.

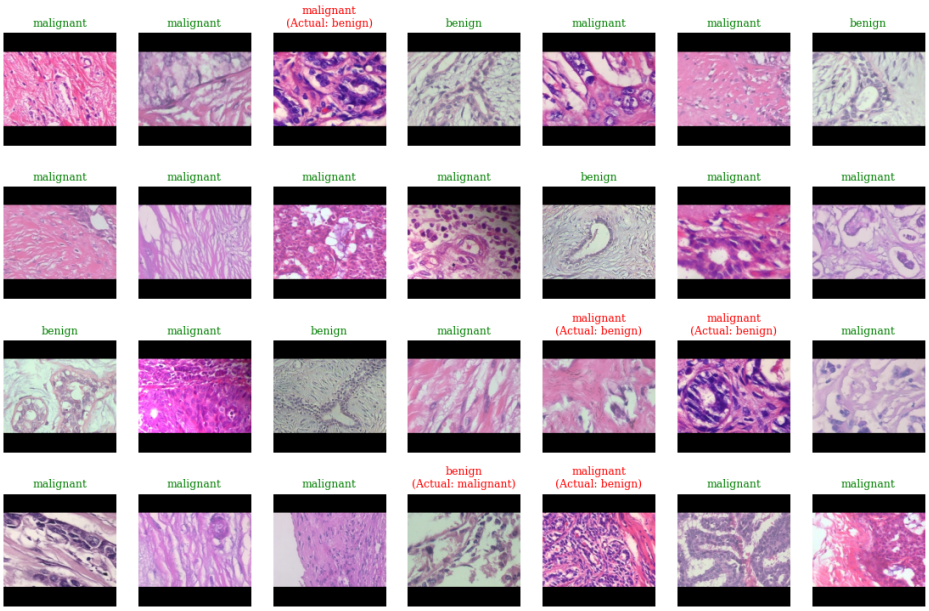

Figure 9: CNN prediction

Starting with the CNN model, it was designed with a relatively simple architecture consisting of multiple convolutional layers followed by max-pooling, global average pooling, and fully connected layers. The network was augmented with dropout layers to mitigate overfitting, a common issue when training on limited datasets. The architecture was composed of 134,541 parameters, nearly all of which were trainable. Despite its simplicity, the model achieved a respectable accuracy of 85.47% and an ROC-AUC score of 83.18%, with a moderate loss of 0.5351. The model produced 86 false positives and 81 false negatives, indicating a relatively balanced trade-off between the two types of errors. This performance is notable given the lightweight nature of the model and the limited data on which it was trained. However, the presence of false negatives remains a significant concern, as these could lead to missed diagnoses in a clinical setting.

In contrast, when employing transfer learning techniques, more complex pre-trained models were evaluated. The first of these was MobileNetV3, a model optimized for mobile and edge devices due to its efficiency in computation and power consumption. This model was configured with a total of 3,250,625 parameters, with 254,273 trainable. Despite its computational efficiency, MobileNetV3 underperformed compared to other models in terms of key metrics, achieving an accuracy of 75.98% and an ROC-AUC score of 77.86%, with a loss of 0.4964. Notably, the model generated a substantial number of false positives (190), which is problematic as it could result in unnecessary alarms for healthy individuals. The model's performance suggests that while it is lightweight and efficient, it may not be as suitable for tasks requiring high precision, such as cancer detection, where the cost of false positives is high.

EfficientNetB1, another model evaluated using transfer learning, demonstrated the best overall performance among the tested models. This architecture, known for its balance between efficiency and accuracy, was composed of 7,267,317 parameters, with 336,193 trainable. The model achieved the highest ROC-AUC score of 87.67% and an accuracy of 83.99%, with the lowest loss recorded at 0.3861. Despite these impressive metrics, the model still produced 120 false positives, which, although fewer than those from MobileNetV3, remain a significant concern. The success of EfficientNetB1 in this task is likely due to its advanced architecture,

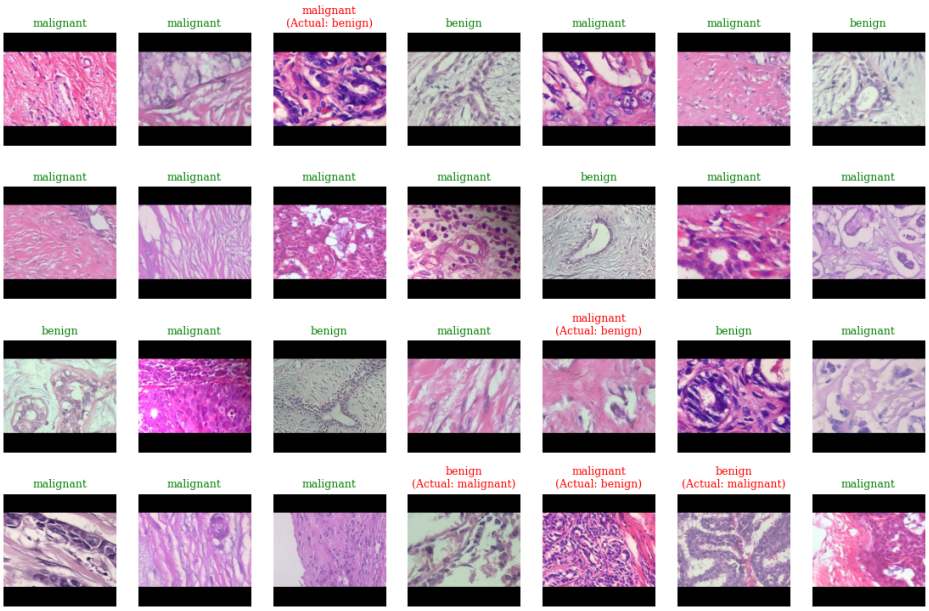

Figure 10: EfficientNetB1 prediction

which scales efficiently with both depth and width of the network, allowing it to capture a broad range of features across different scales in the images.

The VGG16 model, known for its simplicity and depth, was also tested. With a total of 14,936,385 parameters, of which 221,697 were trainable, VGG16 achieved an accuracy of 82.94% and an ROC-AUC score of 87.34%, with a loss of 0.4317. However, the model produced a notably high number of false negatives (107), which is a critical shortfall in medical diagnostics as it could lead to undetected cases of cancer. This issue suggests that while VGG16 can capture detailed features due to its deep architecture, it may struggle with generalizing well enough to minimize both false negatives and false positives effectively.

Lastly, the ResNet50V2 model, a deep residual network with 24,097,601 parameters (532,801 trainable), was evaluated. This model achieved an accuracy of 79.63% and an ROC-AUC score of 84.41%, with a loss of 0.4753. ResNet50V2, however, produced the highest number of false negatives (140), which is a significant drawback, indicating that the model might be overly conservative in its predictions, potentially leading to missed diagnoses. The residual connections in ResNet, which help in training very deep networks, are effective in preventing vanishing gradients but may have contributed to the model's cautious nature, resulting in a higher rate of missed cancer cases.

When comparing the models based on key metrics, EfficientNetB1 stands out as the most promising, with the highest ROC-AUC score and the lowest loss, indicating its superior ability to distinguish between benign and malignant cases. However, despite its high performance, the presence of false positives still highlights the need for further refinement. The CNN model, while achieving the highest accuracy, still faces challenges in minimizing both false positives and negatives effectively.

EfficientNetB1 shows the most potential due to its high ROC-AUC score and low loss. The CNN model, despite being lightweight, achieved commendable results, while pre-trained models like VGG16 and ResNet50V2 highlighted the trade-offs between model complexity and performance. The results suggest that for tasks involving critical medical diagnostics, such as cancer detection, it is essential to balance accuracy, false positive rates, and false negative rates to ensure reliable and trustworthy predictions. Further fine-tuning and possibly combining multiple models could lead to improved performance and more robust outcomes in clinical settings.

6 Conclusion

The research presented in this study highlights the critical role that deep learning models can play in the accurate differentiation between benign and malignant breast tumors using histopathological imaging data. The investigation utilized a diverse array of convolutional neural network (CNN) architectures, including MobileNetV3, EfficientNetB1, VGG16, and ResNet50V2, to assess their efficacy in classifying breast cancer images from the Breast Cancer Histopathological Image Classification (BreakHis) dataset. This comprehensive evaluation not only provides insights into the relative performance of these models but also underscores the broader implications for the application of deep learning in medical diagnostics.

One of the key conclusions of this research is the recognition that deep learning models, when appropriately designed and trained, have the potential to significantly improve diagnostic accuracy in breast cancer detection. Among the models evaluated, EfficientNetB1 emerged as a promising architecture, achieving the highest ROC-AUC score of 0.8767. This result indicates that EfficientNetB1 is highly effective at distinguishing between benign and malignant tumors. The compound scaling method employed by EfficientNetB1, which optimizes the balance between network depth, width, and resolution, proved to be instrumental in achieving this high level of performance. This scaling approach allows the model to efficiently handle the complexities of histopathological images, which often exhibit significant variations in texture, structure, and color.

Despite its strong performance, the study also identified a critical limitation of the EfficientNetB1 model: a relatively high rate of false positives. In the context of breast cancer diagnosis, false positives can lead to unnecessary anxiety for patients, additional testing, and potentially harmful interventions. Therefore, while EfficientNetB1 demonstrates strong potential, it is clear that further refinement is needed to reduce the occurrence of false positives and enhance the model's reliability in clinical practice. This could involve the integration of additional data, the development of more sophisticated preprocessing techniques, or the refinement of the model's architecture to better distinguish between subtle variations in benign and malignant tissue.

The CNN model, despite being simpler than the other architectures evaluated, achieved the highest accuracy in this study. This finding is noteworthy as it suggests that, in resource-limited settings, simpler models like CNNs could offer a viable solution for breast cancer diagnosis. The ability of a basic CNN to deliver high accuracy with fewer computational requirements positions it as a practical option for deployment in regions with limited access to advanced medical infrastructure. This highlights the importance of considering not only the raw performance metrics of a model but also its suitability for the specific context in which it will be used.

The evaluation of VGG16 and ResNet50V2 also provided insights into the trade-offs between model complexity, accuracy, and computational efficiency. VGG16, while achieving high accuracy, was found to be computationally intensive and less suitable for real-time clinical applications due to its large number of parameters and high memory requirements. However, its architecture remains useful for transfer learning, where pre-trained weights can be fine-tuned for specific tasks with limited data. ResNet50V2, on the other hand, demonstrated the advantages of residual learning in enabling deeper networks without suffering from the degradation problem. This capability makes ResNet50V2 a strong candidate for applications where high accuracy is required, and computational resources are available.

Another important conclusion drawn from this research is the value of data augmentation and preprocessing techniques in enhancing the robustness and generalizability of deep learning models. The study employed a range of augmentation methods, including random brightness adjustments, flips, and rotations, which helped the models better handle variations in image orientation and lighting. These techniques are important in medical image analysis, where variability in image acquisition can introduce noise and artifacts that complicate the diagnostic process. By improving the models' ability to generalize across different

imaging conditions, data augmentation contributes to more reliable and consistent performance in real-world clinical scenarios.

The findings of this research also shows the broader potential of artificial intelligence (AI) and deep learning in transforming medical diagnostics. As the volume of medical imaging data continues to grow, the demand for automated tools that can assist clinicians in making accurate and timely diagnoses is increasing. Deep learning models, with their ability to learn complex patterns and relationships within large datasets, are uniquely positioned to meet this demand. The success of models like EfficientNetB1 in this study demonstrates that AI can achieve performance levels that are competitive with, or even surpass, those of human experts in certain tasks. However, the deployment of these models in clinical practice requires careful consideration of their limitations, including the need to minimize false positives and ensure interpretability and transparency in decision-making processes.

Several avenues for future research and development emerge from this study. One promising direction is the exploration of ensemble methods, where multiple models are combined to leverage their complementary strengths. By integrating the outputs of different architectures, it may be possible to reduce the incidence of false positives and improve overall diagnostic accuracy. Additionally, future research could investigate the incorporation of multimodal data, such as combining histopathological images with genomic or clinical data, to provide a more comprehensive assessment of tumor characteristics. This could lead to the development of more personalized diagnostic tools that take into account the unique biological and clinical context of each patient.

In medical settings, it is crucial that AI-driven decisions can be understood and validated by clinicians (Sisodia et al., 2020) (Reddy, 2018). Techniques such as attention mechanisms, saliency maps, and model visualization can help in making the decision-making process of deep learning models more transparent. This would not only increase clinician trust in AI tools but also provide insights into the features and patterns that are most indicative of malignancy, potentially leading to new discoveries in cancer pathology.

References

Aceves-Fernandez, M. A. (2020). Advances and applications in deep learning.

Adeshina, S. A., Adedigba, A. P., Adeniyi, A. A., and Aibinu, A. M. (2018). Breast cancer histopathology image classification with deep convolutional neural networks. In *2018 14th international conference on electronics computer and computation (ICECCO)*, pages 206–212. Ieee.

Bayramoglu, N., Kannala, J., and Heikkilä, J. (2016). Deep learning for magnification independent breast cancer histopathology image classification. In *2016 23rd International conference on pattern recognition (ICPR)*, pages 2440–2445. IEEE.

Carey, L. A., Perou, C. M., Livasy, C. A., Dressler, L. G., Cowan, D., Conway, K., Karaca, G., Troester, M. A., Tse, C. K., Edmiston, S., et al. (2006). Race, breast cancer subtypes, and survival in the carolina breast cancer study. *Jama*, 295(21):2492–2502.

Elmore, J. G., Armstrong, K., Lehman, C. D., and Fletcher, S. W. (2005). Screening for breast cancer. *Jama*, 293(10):1245–1256.

Hortobagyi, G. N. (1998). Treatment of breast cancer. *New England Journal of Medicine*, 339(14):974–984.

Houghton, S. C. and Hankinson, S. E. (2021). Cancer progress and priorities: breast cancer. *Cancer epidemiology, biomarkers & prevention*, 30(5):822–844.

Key, T. J., Verkasalo, P. K., and Banks, E. (2001). Epidemiology of breast cancer. *The lancet oncology*, 2(3):133–140.

MacMahon, B., Cole, P., and Brown, J. (1973). Etiology of human breast cancer: a review. *Journal of the National Cancer Institute*, 50(1):21–42.

Momenimovahed, Z. and Salehiniya, H. (2019). Epidemiological characteristics of and risk factors for breast cancer in the world. *Breast Cancer: Targets and Therapy*, pages 151–164.

Prat, A. and Perou, C. M. (2011). Deconstructing the molecular portraits of breast cancer. *Molecular oncology*, 5(1):5–23.

Reddy, S. (2018). Use of artificial intelligence in healthcare delivery.

Sisodia, D. S., Pachori, R. B., and Garg, L. (2020). *Handbook of research on advancements of artificial intelligence in healthcare engineering*. IGI Global.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. (2015). A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, 63(7):1455–1462.

Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G. R., et al. (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature*, 486(7403):400–404.

Sun, Y.-S., Zhao, Z., Yang, Z.-N., Xu, F., Lu, H.-J., Zhu, Z.-Y., Shi, W., Jiang, J., Yao, P.-P., and Zhu, H.-P. (2017). Risk factors and preventions of breast cancer. *International journal of biological sciences*, 13(11):1387.

Waks, A. G. and Winer, E. P. (2019). Breast cancer treatment: a review. *Jama*, 321(3):288–300.

Wei, B., Han, Z., He, X., and Yin, Y. (2017). Deep learning model based breast cancer histopathological image classification. In *2017 IEEE 2nd international conference on cloud computing and big data analysis (ICCCBDA)*, pages 348–353. IEEE.

World Cancer Research Fund International (2024). Breast cancer statistics. https://www.wcrf.org/cancer-trends/breast-cancer-statistics/. Accessed: 2022-08-24.

Yamlome, P., Akwaboah, A. D., Marz, A., and Deo, M. (2020). Convolutional neural network based breast cancer histopathology image classification. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1144–1147. IEEE.

Zhu, C., Song, F., Wang, Y., Dong, H., Guo, Y., and Liu, J. (2019). Breast cancer histopathology image classification through assembling multiple compact cnns. *BMC medical informatics and decision making*, 19:1–17.

Affiliation of Weng Chan Hackensack Meridian School of Medicine : Hackensack Meridian School of Medicine