Decoding Cancer Tissues: A Comparative Deep Learning View of Breast Histopathology

Maanasvee Khetan, Sanya Malik, Reya Oberoi, Dr. Hima Deepthi Vankayalapati
Department of Artificial Intelligence
SVKM NMIMS Mukesh Patel School of Technology Management and Engineering, Mumbai, India

***Abstract*—Worldwide, breast cancer continues to be one of the riskiest killers of women, so early and correct diagnoses will be pivotal to treatment. As the field progresses in deep learning, detection of histopathological images has revolutionized the development of automated, reliable and scalable diagnostics. In this study, we present a thorough comparative analysis of six deep learning image classification architectures, including, a Convolutional Neural Network (CNN), ResNet50, DenseNet, EfficientNetV2-B0, MobileNetV1, and InceptionV3 in classifying breast cancer histopathological images from the BreaKHis dataset. We report each models performance based upon various key metrics of accuracy, precision, recall, and F1-score in order to evaluate e Diagnoses Capability and computational efficiency of models. Our study identified significant differences between framework performances, with the transfer models out performing CNN models in accuracy and generalization ability. We provide an insight into the strengths of utilizing deep learning framework architectures for medical imaging and provide new directions for more powerful AI-assisted diagnostic processes.**

***Index Terms*—Breast Cancer, Histopathology Images, Deep Learning, CNN, Transfer Learning, EfficientNet, Comparative Analysis**

I. Introduction

A. Background

Breast cancer is still one of the top causes of death for women around the world. Early detection and accurate diagnosis is vital to the prognosis and survivorship of an individual because it allows timing clinical intervention. Among the range of diagnostic modalities for breast cancer, the histopathological image analysis is considered the gold standard for confirming the presence and type of breast cancer. Microscopically viewed, these tissue images can provide insights into the cellular characteristics and architecture of the tissue, which are necessary factors in distinguishing between benign and malignant tumors. The drawback is that the initial examination of histopathological slides by pathologists is often long, tedious, and subjective due to variations in experience and fatigue [2]. This predicament has created the need for automated, precise, and scalable computational methods to assist pathologists for their diagnostic scenarios.

B. Deep Learning in Medical Imaging

Recently, Deep Learning and in particular Convolutional Neural Networks (CNNs) is altering the practice of diagnosis in image-based medicine. CNNs can automatically learn deep, complex hierarchical representations from raw images without the need for feature extraction by the user. CNNs have been demonstrated to perform at a high level across various imaging domains, including radiology, pathology, and dermatology [3]. The addition, of transfer learning techniques to medical imaging has allowed researchers to leverage current pretrained architectures (e.g. ResNet, DenseNet, EfficientNet, Inception, and MobileNet) trained on large-scale data sets (e.g., ImageNet) by modifying and adapting them with specialized low sample medical [3]; all of which improves speed of convergence, generalizability and limited computative cost.

C. Research Motivation Gap

To date, there has been extensive deep learning research focused on breast cancer classification, although the vast majority of studies have looked at a single model, or adapted to a particular dataset. In comparison, much less research has conducted a rigorous evaluation of multiple state-of-the-art architectures directly against each other in controlled conditions. This ultimately produces a gap in the literature that indicates the need for a systematic evaluation of today's state-of-the-art CNN architectures through similar dataset and preprocessing and evaluation protocols.

D. Objectives

The primary objectives of this study are as follows:

- To perform a comprehensive comparative analysis of six deep learning architectures **ResNet50**, **custom CNN**, **DenseNet**, **EfficientNetV2-B0**, **MobileNetV1**, and **InceptionV3** for breast cancer histopathological image classification using the **BreakHis** dataset.
- To evaluate the performance of each architecture using standard evaluation metrics, including **accuracy**, **precision**, **recall**, and **F1-score**, under consistent experimental conditions.
- To identify the most efficient and accurate architecture that can potentially be deployed in real-world **breast cancer diagnostic systems**, assisting pathologists in rapid and reliable diagnosis.

II. Literature Review

Deep learning has widely transformed digital histopathology in the last few years by allowing end-to-end learning of features directly from tissue images, attempting to overcome many of the shortcomings of the traditional manual microscopy and handcrafted features approaches. In this section we highlight a number of important previous work; then

we structure them by theme and make connections to the architectures you will evaluate.

A. Advancements in Histopathology Analysis

Most early approaches to analysis of histopathology images relied on manually engineered image features texture, morphology, shape of nuclei, glandular architecture and classical machine learning classifiers (SVM, random forests). While these studies demonstrated a reasonable degree of accuracy, they were often limited in generalizability and required an understanding of domain to implement. With the development of convolutional neural networks (CNN), the extraction of image features moved to the learning domain, improving classification performance dramatically, and making large-scale automated analyses possible.

As an example of the review of breast-cancer pathology images, Zhao et al. (2022) reviewed over 200 articles in the detection, segmentation, and classification tasks; they noted that DL model classification accuracy exceeded that of expert pathologists in some instances [3].

B. Deep Learning Architectures & Transfer Learning

The transition to deeper models in the research community was sped up through transfer learning: pretrained networks (e.g., from ImageNet) were fine-tuned in histopathology tasks. Some examples of these architectures include:

- **Classic CNNs** (e.g., VGG, AlexNet)
- **Skip-connection networks** (e.g., ResNet50): allow deeper networks to be trained by resolving the vanishing gradient problem.
- **Dense connectivity models** (e.g., DenseNet): allows for feature-reuse and compact representations.
- **Lightweight models** (e.g., MobileNetV1): designed for optimized inference and resource-constrained deployments.
- **Compound-scaling architectures** (e.g., EfficientNetV2-B0): tuned depth, width, and resolution for a better accuracy/efficiency trade-off.
- **Multi-scale feature extractors** (e.g., InceptionV3): are able to learn features at multiple receptive-field sizes.

These architectures have produced large gains for histopathological classification and segmentation tasks across datasets [3].

C. Deep Learning Applications in Breast Cancer Histopathology

There are several reviews recently conducted examining breast cancer specifically in the context of pathology:

- **Jiang et al. (2024)** reviewed DL applications in breast cancer (pathology), segmentation and classification via large scale datasets (e.g. TCGA, multi-centre cohorts). They report that DL models improve diagnostic accuracy [2], detection of metastasis, and clinical prognosis prediction. They highlight that there is need for robust annotated datasets and challenges of interpretability and regulatory compliance.
- **Unger et al. (2024)** considered DL in cancer genomics and histopathology, taking note of multimodal models combining image and genomic data in their survey [1]. They note challenges regarding reproducibility, dataset bias, and generalisation across institutions.
- **Obeid et al. (2024)** published a review "Advancing Histopathology with Deep Learning Under Data Scarcity: A Decade in Review [2]", which touches especially on the challenge of limited labelled datasets in digital pathology, and categorises other approaches including weak supervision, self-supervised learning, multiple-instance learning, and active learning.

These references identify important trends: moving from patch classification to whole-slide imaging analysis, mainstreaming explainability and visualisation tools, and a shift away from research prototypes to clinically graded systems.

1) Data Scarcity & Generalisability: Obeid et al. argue that the many billions of parameters of complexity of their model have increased, but the amount of annotated data available in histopathology is limited and poses a major barrier to generalisable systems. Self-supervised pre-training, domain adaptation, and weak supervision are all examples of methods that could be employed.

2) Dataset Diversity & Magnification Variation: In histopathology and digital pathology, heterogeneity in data exists as a result of variation in image magnification, staining protocols, and slide preparation. Some studies reported systems trained on annotated data at one magnification did not generalise well to other magnification levels. The need for datasets that contained multiple magnifications in the same dataset and magnification invariant explicit modelling are also highlighted.

3) Interpretability & Clinical Deployment: Faust et al. (2018) developed methods for visualising feature embeddings of CNN outputs to evaluate how models make predictions in histopathology.

More recent work (e.g., Dolezal et al., 2024) [1] has provided some frameworks, such as *Slideflow*, which encompass data processing, explainability, uncertainty quantification and real-life implementation of models in pathology settings.

However, the leap between research proofs of concept and robust clinical validation across institutions remains to be filled.

4) Evaluative Metrics & Benchmarking: While many studies report comparably high accuracy, precision and recall metrics, they seldom compare models under controlled conditions or compare approaches across model architectures. The work of Obeid et al. calls for systematic benchmark tables and standardised protocols within histopathology tasks.

In fact, in digital histopathology, augmentation strategies are also non-trivial: Ameen et al. (2023) specifically examined when and how augmentation will influence performance and in doing so, concluded that augmenting the training data after the test-set is demarcated will yield a more truthful evaluation of performance.

D. Justification for Comparative Study of Architectures

While DL is developing fast in breast histopathology, the majority of related projects are published as either one-architecture case studies, or as simply comparing two or three architectures. There remains a need for a direct, structured comparison of a broader set of architectures using the same dataset conditions (preprocessing, dataset splits, level of magnification) hence the rationale for the current study to compare six of the major architectures (CNN, ResNet50, DenseNet, EfficientNetV2-B0, MobileNetV1, InceptionV3) on the same dataset (BreakHis) using the same metrics for evaluation. The comparative findings can potentially help inform researchers and future clinical consumers about the considerations of accuracy, efficiency and deployability of the architectures.

III. Dataset Description

The current research utilizes the **Breast Cancer Histopathological Image Classification (BreakHis)** dataset, an openly available benchmark dataset of microscopic breast-tumor biopsy images. This dataset comprises **7,909 RGB images**, separated into benign and malignant tumor categories. All samples originated from authentic clinical investigations and were digitized via a microscopic camera connected to a standard optical microscope. This database has been built in collaboration with the **P&D Laboratory – Pathological Anatomy and Cytopathology, Parana, Brazil**.

A. Data Composition and Class Distribution

The dataset is composed of two primary diagnostic categories: **benign** and **malignant** tumors. Specifically, it contains **2,480 benign** and **5,429 malignant** tumor images. Each class includes multiple histologic subtypes, introducing natural variations in tissue morphology, cellular structure, and staining intensity.

TABLE I
Magnification and Effective Pixel Size

Visual magnification	Objective lens	Effective pixel size (μm)
40×	4×	0.49
100×	10×	0.20
200×	20×	0.10
400×	40×	0.05

The preparation method involves several steps such as fixation, dehydration, clearing, infiltration, embedding, and trimming. Sections of 3 μm are cut from tissue samples using a microtome and mounted on slides. These sections are then stained and covered with a glass coverslip. Expert anatomopathologists visually analyze the tissue slides under a microscope to determine the tumoral regions. The final diagnosis for each case is established by experienced pathologists, often confirmed by complementary tests, particularly immunohistochemistry analysis.

The breast tissue slides were captured using an *Olympus BX-50 system microscope* with a *3.3× relay lens* coupled to a *Samsung digital color camera SCC-131AN*. This camera employs a *1/3" Sony Super-HAD* (Hole-Accumulation Diode) interline transfer charge-coupled device (CCD) with a pixel size of 6.5 μm × 6.25 μm and a total resolution of 752 × 582 pixels. The images were acquired in three-channel red–green–blue (RGB) *TrueColor* format (24-bit color depth, 8 bits per channel) at magnification levels of 40×, 100×, 200×, and 400×, corresponding respectively to objective lenses of 4×, 10×, 20×, and 40×.

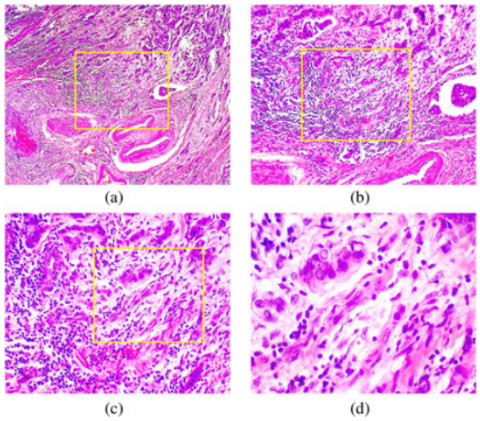

Fig. 1. Slide of breast malignant tumor (stained with HE) seen in different magnification factors: (a) 40×, (b) 100×, (c) 200×, and (d) 400×. Highlighted rectangle (manually added for illustrative purposes only) is the area of interest selected by pathologist to be detailed in the next higher magnification factor.

B. Image Format and Resolution

Each image is stored in **RGB JPEG** format with a native resolution of approximately **700 × 460 pixels**. The dataset exhibits heterogeneous illumination conditions, staining variations, and background noise, reflecting the visual diversity encountered in real clinical settings. Such heterogeneity makes it ideal for evaluating the generalization capability of deep learning models under non-uniform imaging conditions.

C. Preprocessing Operations

To ensure consistency and optimize learning during model training, the following preprocessing pipeline was implemented:

- **Resizing:** All images were resized to a fixed input dimension compatible with each CNN architecture.
- **Normalization:** Pixel intensities were normalized to a standard dynamic range, ensuring uniform input scaling.
- **Data Augmentation:** Random horizontal and vertical flips, small rotations, and controlled brightness/contrast variations were applied to expand dataset variability and minimize overfitting.

These preprocessing steps ensured that all models were trained on balanced, high-quality input data while maintaining the inherent diagnostic characteristics of the original histopathological images.

The procedure for obtaining images at different magnifications is as follows: first, the pathologist locates a tumor and

TABLE II
IMAGE DISTRIBUTION BY MAGNIFICATION FACTOR AND CLASS

Magnification	Benign	Malignant	Total
40×	625	1370	1995
100×	644	1437	2081
200×	623	1390	2013
400×	588	1232	1820
Total	2480	5429	7909
# Patients	24	58	82

TABLE III
BENIGN IMAGE DISTRIBUTION BY MAGNIFICATION FACTOR AND HISTOLOGICAL SUBTYPES

Magnification	A	F	TA	PT	Total
40×	114	253	109	149	598
100×	113	260	121	120	614
200×	111	264	108	140	594
400×	106	237	131	88	562
Total	444	1014	469	497	2424
# Patients	4	10	3	7	24

identifies the region of interest (ROI). To adequately cover the entire ROI, multiple images are acquired through the lowest maginifcation, namely 40×. The pathologist generally chooses images with only one tumor type (most of the time); however, images can also include transitional tissue (i.e., a mixture of normal-pathological). On average, a total of aaveraged 24 images per patient are taken through each slide at low power (i.e., 40×, see Table II). Following this, the pathologist manually zooms to 100× and takes a similar quantity of images from within the original ROI. This is further repeated at both 200× and 400× magnification, consecutively. Images that are out-of-focus are then ultimately eliminated through final visual (i.e, manual) review.

IV. METHODOLOGY

The experimental framework for this study is designed with a systematic workflow that encompasses dataset preparation, data preprocessing, model selection, training, evaluation, and comparative analysis.

A. Dataset Utilization

For the experiment, we used the BreakHis (Breast Cancer Histopathological Database), which contains microscopic biopsy images classified as benign or malignant. Each image is taken at four different magnification levels 40×, 100×, 200×, and 400× to assess how robust the models are at various scales. This dataset offers a balanced representation of different tumor subtypes, facilitating effective supervised learning for binary classification tasks.

TABLE IV
MALIGNANT IMAGE DISTRIBUTION BY MAGNIFICATION FACTOR AND HISTOLOGICAL SUBTYPES

Magnification	DC	LC	MC	PC	Total
40×	864	156	205	145	1370
100×	903	170	222	142	1437
200×	896	163	193	138	1390
400×	788	137	192	115	1232
Total	3451	626	812	540	5429
# Patients	38	5	9	6	58

B. Data Preprocessing

To enhance compatibility with the model and boost convergence rates, several preprocessing steps were undertaken:

- **Image Resizing:** All images were resized to a consistent dimension (for instance, 224×224 pixels) to satisfy the input requirements of deep learning models.
- **Normalization:** Pixel intensity values were converted to a range of [0,1], helping stabilize gradient updates during training.
- **Data Augmentation:** Techniques such as rotation, flipping, and zooming were employed to enrich dataset diversity while reducing overfitting risks.
- **Dataset Splitting:** The data was divided into three subsets: training (70%), validation (15%), and testing (15%) to objectively assess model generalization capabilities.

C. Model Selection

A variety of both well-established CNN and recent transfer learning architectures were used to compare and evaluate. The models that were selected included:

- **CNN (Custom model)**: A convolutional network that was designed from first principles to serve as the base model.
- **ResNet50**: This deep residual network used skip connections to advance the flow of gradients.
- **DenseNet121**: An architecture that encouraged feature reuse based on dense connections.
- **EfficientNetV2-B0**: In this model, a trade-off was made between computational efficiency and accuracy.
- **MobileNetV1**: This model was developed as a lightweight model that may not perform at its best, but is suited for resource-constrained environments.
- **InceptionNetV3**: The Inception model used inception modules to capture multi-scale features.

All models were initialized with ImageNet weights (where applicable) and subsequently trained on the BreakHis dataset.

D. Model Training

All models trained in a controlled environment utilizing Google Colab, while taking advantage of GPU acceleration. The hyperparameters were implemented consistently amongst the models for an accurate comparison:

- **Optimizer:** Each model used the Adam optimizer, because of its adaptive learning rate.
- **Loss Function:** Binary Cross-Entropy loss was used as the function for a two-class classifiers measure.
- **Epochs and Batch Size:** Each model trained using a size of 32 for Batch Size and a maximum of 50 epochs, with consideration to early stopping.
- **Learning Rate Schedule:** A learning rate decay schedule would be employed to adjust the learning rate after a plateau in the validation loss.

- **Regularization Techniques:** Early stopping and dropout layers were used as a means to reduce overfitting.

V. Results and Discussion

The various academic deep learning models were evaluated in terms of their ability to classify breast histopathological images as benign or malignant. Each model was trained under the same conditions, using the same learning rate, batch size, optimizer, and preprocessing pipeline, and on the same BreakHis dataset, in order to compare the performance of the models fairly.

This section presents results based on two dimensions: quantitative performance measures and qualitative model behavior through the dimension of a confusion matrix. Quantitative measures included calculated metrics of Accuracy, Precision, Recall, F1-Score and Loss to evaluate and compare the models' classification strength, sensitivity to malignant cases, and generalization performance. Table 6.1 provides a comparison across the models of the results of all models.

TABLE V
Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	Loss
CNN	0.86447	0.88525	0.92875	0.90647	0.43515
ResNet	0.90550	0.90013	0.97230	0.93483	0.26385
MobileNet	0.80940	0.53000	0.61000	0.55000	0.42140
InceptionV3	0.83000	0.54000	0.58000	0.55000	0.52830
EfficientNet	0.91410	0.91000	0.91000	0.91000	0.22430
DenseNet	0.7571	0.57000	0.62000	0.59000	0.56810

A. Discussion

Based on the comparative assessment, EfficientNet displayed the strongest accuracy (91.41%), with a proportional trade-off between precision and recall ratings, conveying its high-level ability to learn complex multi-scale features in a computationally effective way. Its loss value of 0.2243 detected a solid convergence with minimal overfitting.

ResNet was the second-highest to EfficientNet, with a recall of 0.9723, indicating its good ability to locate malignant tissue regions this is vital in any medical setting. ResNets skip connections promoted good gradient flow and better representation of deeper features hence the solid performance.

In general, MobileNet and InceptionV3 had lower performance, likely due to them being lightweight networks and generally less deep networks. These networks suffered to learn the quite stand-offish histopathological patterns because of this.

The custom CNN performed reasonably well, however, its loss value was relatively high, indicating room for improvement in the depth and strength of regularization parameters.

After all experimental results and performance were considered, it is validated that current transfer learning models consistently outperformed standard CNNs (EfficientNet and ResNet) based on the accuracy and stability of training loss. The performance, complexity and scalability of these networks rationalized them as powerful candidates for real world clinical use in digital pathology.

B. Computational Time Analysis

The study also investigated the time efficiency of each architecture together with the classification accuracy metrics. Understanding the runtime is important for assessing the deployability of deep learning methods, particularly in clinical or diagnostic settings, where the speed of inference is crucial for usability. Each model architecture was trained and tested, while configured to the same hyperparameters, on the same (Google Colab, Tesla T4 GPU, 16 GB RAM) GPU-enabled environment. The measured training time and testing time for all architecture in order of time-based efficiency ranking are shown in Table VI, with Rank 1 being the most efficient model based on time of training and time to test efficiency score.

TABLE VI
Model Training and Testing Time Comparison with Efficiency Rank

Model	Training Time (s)	Testing Time (s)	Efficiency Rank
EfficientNetV2-B0	**69.88**	73	**1**
DenseNet121	376.18	27.29	2
CNN	599	30	3
ResNet50	1000	90	4
InceptionV3	7870.50	127	5
MobileNetV1	8878.26	37	6

Based on the findings, **EfficientNetV2-B0** had the shortest training time overall (69.88 s), while testing efficiency remained competitive, further reinforcing its enabling computational scalability. When looking at the other models, **MobileNetV1** and **InceptionV3** had training times of 8878.26 s and 7870.50 s, respectively, due in large part to the overall complexity of their model architectures (depth of the model and multi-scale feature extraction layers). For **DenseNet121** and the custom **CNN**, both achieved moderate training efficiencies where training efficiency is still feasible within a resource-limited environment. Despite the longer training time for **ResNet50** (1000 s), the model is still an excellent measure of accuracy, recall, and inference speed indicating that ResNet50 fits within a diagnostic pipeline where reliability is the more important factor over a specific speed of training time. Overall, **EfficientNetV2-B0** had the best overall relationship of accuracy over computational time, earning the highest efficiency reading, and becoming the most promising choice for scalability within a clinical environment.

VI. Conclusion

This research offers a thorough comparative evaluation of six deep learning models for the classification of breast histopathological images into benign and malignant types using the BreakHis dataset. All models underwent training and assessment in an identical experimental apparatus to abide by an equitable analysis involving learning ability, generalization ability, and computational efficiency.

In the present study, EfficientNetV2-B0 showed the best overall accuracy at 91.41%, illustrating an adequate trade-off between precision, recall, and F1-score. The above-normal accuracy can be attributed to EfficientNetV2-B0's compound

scaling approach where depth, width, and resolution of the architecture are optimally scaled at the same time to effectively represent multi-scale histopathological features. The ResNet50 model also performed well, especially in recall, which shows promise for identifying malignant tissue regions, a valuable aspect in clinical settings for the importance of sensitivity.

In contrast, lighter architectures such as MobileNetV1 and InceptionV3 demonstrated slightly lower accuracies because they presented with a lower representational capacity for complex tissue textures and appeal to slight visual differences. The custom CNN produced reasonable accuracy alongside a higher loss value which shows room for improvement in regularization and hyperparameter adjustments.

TABLE VII
COMPARISON OF DEEP LEARNING ARCHITECTURES

Characteristic	Architecture	Performance	Complexity
CNN	Baseline	Varies	Low
ResNet50	Residual connections	Very Good	Medium
DenseNet	Dense feature reuse	Very Good	High
EfficientNetV2-B0	Scaled efficiency	Excellent	High
MobileNetV1	Lightweight	Moderate	Low
InceptionNetV3	Multi-scale features	Good	Medium

Overall, the findings support the potential of transfer learning–based deep neural networks to automate the diagnosis of breast cancer in histopathological images. These models can be interpreted as CAD devices in digital pathology, which reduce subjectivity in medical diagnosis and improve speed and consistency in predictions. Current limitations include the use of imbalanced datasets across magnification levels, and a lack of explainable features to demonstrate how models arrived at their decisions.

In addition, the analysis of computational efficiency demonstrated that EfficientNetV2-B0 achieved the greatest accuracy and was also ranked as the most efficient model with the lowest training time and a good performance at inference. The tradeoff of diagnostic accuracy and lower computational time of EfficientNetV2-B0 further supports its hoard as a candidate for medical imaging applications that would involve real-time analysis or applications with limited computational resources.

Future work may focus on the use of explainable AI methods, increasing diversity in datasets, and developing hybrid ensemble methods to improve diagnostic reliability and clinical usability.

VII. FUTURE WORK

Future research could be directed toward advancing both the interpretability and generalizability effectiveness of deep learning procedures for histopathological imaging. Explainable AI (XAI) methods such as Grad-CAM or LIME could be used to visualize the specific tissue area(s) that informed the models' predictions, which could enhance model transparency and trust in the model for clinical use. In addition, balanced, diverse, and expansive datasets across multiple levels of magnification and tumor subtypes would provide additional robustness while minimizing bias. Future studies could also advance this framework into multi-class classification to assist with identifying cancer subtypes with greater specificity.

(a) The exploration of hybrid or ensemble models that capitalize on the functionality of multiple architectures may yield superior feature representation, as well as improved accuracy.

(b) In addition, compression or quantization of the models may optimize the architecture for real-time clinical or mobile diagnostics. Lastly, collaboration with pathologists for expert validation and testing through institution-level cross-dataset studies would enhance both the accessibility and clinical relevance of the proposed system.

REFERENCES

[1] J. M. Dolezal, S. Kochanny, E. Dyer, *et al.*, "Slideflow: Deep learning for digital histopathology with real-time whole-slide visualization," *BMC Bioinformatics*, vol. 25, p. 134, 2024. doi:10.1186/s12859-024-05758-x.

[2] A. Obeida, S. Boumarafa, A. Sohail, T. Hassan, S. Javed, J. Dias, M. Bennamoun, and N. Werghi, "Advancing Histopathology with Deep Learning Under Data Scarcity: A Decade in Review," *arXiv preprint* arXiv:2410.19820, 2024. Available: https://arxiv.org/abs/2410.19820.

[3] D. Komura, M. Ochi, and S. Ishikawa, "Machine learning methods for histopathological image analysis: Updates in 2024," *Computational and Structural Biotechnology Journal*, vol. 23, pp. 1–15, Jan. 2025. doi:10.1016/j.csbj.2024.12.045.

[4] Y. Wu, M. Cheng, W. Shao, *et al.*, "Recent advances of deep learning for computational histopathology: Principles and applications," *Cancers*, vol. 14, no. 5, p. 1152, Feb. 2022. doi: 10.3390/cancers14051152.

[5] A. H. Song, G. Jaume, D. F. K. Williamson, M. Y. Lu, A. Vaidya, T. R. Miller, and F. Mahmood, "Artificial Intelligence for Digital and Computational Pathology," *arXiv preprint*, arXiv:2401.06148, Dec. 2023.

[6] B. Jiang, L. Bao, S. He, X. Chen, Z. Jin, and Y. Ye, "Deep learning applications in breast cancer histopathological imaging: Diagnosis, treatment, and prognosis," *Breast Cancer Research*, vol. 26, no. 137, 2024.

[7] M. Nasser and U. K. Yusof, "Deep learning based methods for breast cancer diagnosis: A systematic review and future direction," *Diagnostics (Basel)*, vol. 13, no. 1, p. 161, Jan. 2023, doi: 10.3390/diagnostics13010161.

[8] R. G. Martinez and D. M. van Dongen, "Deep learning algorithms for the early detection of breast cancer: A comparative study with traditional machine learning," *Journal of Biomedical Informatics*, Aug. 2023, doi: 10.1016/j.jbi.2023.104331.