AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
E-ISSN:3069-0102
VOL.5, NO. 4(2)(2025)

BREAST CANCER CLASSIFICATION USING RESNET-50

1st Mr,Nithin Rajput
Assistant Professor
Department of CSE
Neil Gogte Institute of Technology Hyderabad, Telangana
Mobile no:9700563465
Email id: singhnithin1@gmail.com

2nd Dr.S.Swapna
Head and Assistant Professor
Departmentt. of CSE
Neil Gogte Institute of Technology Hyderabad, Telangana
Mobile no:9392488167
Email id:swapnangit2021@gmail.com

3rd Dr,Purude Vaishali Narayanrao
Head and Assistant Professor
Deptatment. of CSE,
Neil Gogte Institute of Technology Hyderabad, Telangana
Mobile no. 9154421954
Email id:vaishnupurude@gmail.com

4th Mr,P.Suresh Babu Padmanaban
Assistant Professor
Department. of CSE
Neil Gogte Institute of Technology
Hyderabad, Telangana
Mobile no.9787640719
Email id. sureshbabugeetha@gmail.com

5th Dr.Maragoni Mahender
Assistant Professor
Department of CSE
Neil Gogte Institute of Technology
Hyderabad, Telangana
Mobile no: 9030503890
Email id: m.mahender527@gmail.com

6th Dr. Pinnapureddy Manasa
Assistant Professor
Department of CSE
Neil Gogte Institute of Technology
Hyderabad, Telangana
Mobile no: 7674021861
Email Id: reddymanasa24@gmail.com

***Abstract*—Breast cancer is a significant public health concern, which is playing a significant role in high mortality rate among women worldwide. The histopathological analysis of breast tissue remains the excellent standard for diagnosis but is highly time-consuming process, prone to inter observer variability, and requires specialized expertise. To address these challenges, this paper proposes an automated breast cancer classification system leveraging deep learning, specifically the ResNet-50 architecture, for accurate and efficient image- based diagnosis.**

Our method utilizes a dataset of 7000 paired grayscale and RGB histopathological images, each resized to 256×256 pixels and normalized within the range [1, 1]. The data is divided into training (80%), validation (10%), and testing (10%) subsets. We employ transfer learning with ImageNet pre-trained ResNet-50 weights, fine-tuned on our dataset. Extensive preprocessing techniques such as image augmentation, normalization, and noise reduction were applied to improve generalization and avoid overfitting.

The proposed model achieves an overall accuracy of 97.8%, along with high precision and recall, outperforming conventional CNNs and shallow classifiers. We also analyze the model's confusion matrix, ROC curve, and F1-score to assess its diag- nostic reliability. The results demonstrate that ResNet-50 is well- suited for feature extraction and classification in histopathological imaging tasks.

This research contributes to the development of AI-assisted diagnostic systems that can support pathologists and diagnostic doctors in early detection and treatment planning. Future work will focus on real-time deployment through a web-based interface and integration with electronic medical record (EMR) systems for clinical utility.

I. INTRODUCTION

Breast cancer is one of the most highly important malignancies affecting women worldwide and remains a significant cause of cancer-related mortality [5], [15]. Early and accurate diagnosis and detection plays an important role in improving patient outcomes and guiding effective treatment plans [1], [4]. Traditionally, diagnostic methods such as mammography, ultrasound, and biopsy-based histopathology are employed for breast cancer detection. Al-though histopathological image analysis is considered the gold standard due to its ability to reveal cellular-level abnormalities which makes to detect damage cells with ease [4], manual inspection by pathologists is high time-taking, high labour, and susceptible to inter-observer variability [2], [8].

New advancements in artificial intelligence (AI), particularly in deep learning, have revolutionized medical imaging by automating feature extraction and classification processes with improved accuracy and reproducibility [1], [3], [6]. Convo- lutional Neural

Networks (CNNs), especially deep architec- tures like ResNet-50, have emerged as powerful tools in histopathological image classification due to their capability to learn complex features and mitigate vanishing gradient issues using residual connections [3], [7], [13]. ResNet-50, a 50-layer deep residual network, has nicely given good practical like explaination and great performance across various biomedical image classification tasks [1], [6], [9].

Several studies have explored the effectiveness of trans- fer learning using ResNet-50 in breast cancer classification. For example, Mahmud et al. employed transfer learning on histopathological images and reported substantial gains in classification performance [2]. Similarly, Vesal et al. applied ResNet-50 to classify breast cancer histology images, achiev- ing high accuracy even with limited data [3]. The integration of ResNet-50 with ensemble or hybrid architectures has further improved robustness and diagnostic reliability in challenging clinical datasets [4], [10], [13].

In this study, we propose a ResNet-50-based deep learn- ing framework for determining and sorting of breast histopathology pictures into benign and malignant categories. The model is smoothly done using transfer learning techniques on a curated dataset comprising thousands of annotated RGB and grayscale histopathological images. Extensive preprocessing and data augmentation techniques are employed to enhance general- ization and address data imbalance issues [11], [14]. Further- more, we develop a web-based interface to facilitate real-time predictions, enabling integration into clinical workflows and decision-support systems [6], [9].

The remainder of the paper is organized as follows: Section II presents a review of work related in breast cancer classifica- tion using deep learning. Section III details the dataset, prepro- cessing pipeline, and ResNet-50 model architecture. Section IV outlines the experimental setup and training configuration. Section V display the results and performance evaluation. Finally, Section VI concludes the paper and give directions for research in future.

II. LITERATURE SURVEY

Introducing of deep learning in medical imaging, partic- ularly for breast cancer detection, has revolutionized diagnostic accuracy by automating feature extraction and classification tasks. Numerous studies have explored the use of Convolu- tional Neural Networks (CNNs), transfer learning, and hybrid frameworks to address challenges associated with heteroge- neous imaging data and limited dataset sizes.

Nguyen et al. [1] came up with an ensemble based deep learning approach that combined multiple CNN backbones to classify histopathological images with increased robustness. Their architecture showed a significant improvement in F1-score when evaluated on the BreaKHis dataset. Meanwhile, Anwar and Awan [2] introduced a transfer learning model leveraging pre- trained MobileNet and DenseNet architectures for mammogram classification, demonstrating improved performance even with fewer training samples.

Sharma et al. [3] focused on designing a lightweight and computationally efficient CNN model suitable for deployment on edge devices in low-resource environments. Their archi- tecture maintained high sensitivity without the need for GPU- based acceleration. Banu et al. [4] enhanced classification ro- bustness by integrating hybrid image preprocessing techniques with ResNet-50-based feature extraction, yielding higher ac- curacy under noisy or low-quality imaging conditions.

From a radiobiological perspective, Yamashita et al. [5] and Tsujii et al. [6] investigated the dosimetric and biological im- pacts of irradiation techniques used in breast cancer treatment. These insights are critical for future integration of imaging biomarkers with radiation therapy planning, contributing to more personalized treatment strategies.

In the domain of classical machine learning, Dey and Samad [7] proposed a pipeline using Principal Component Analysis (PCA) for reducing dimensions are followed by Support Vector Machine (SVM) classification. This approach offered interpretability and computational efficiency, making it a benchmark in early diagnosis systems. Shah et al. [8] advanced this further by using ensemble techniques that stack multiple CNN classifiers, significantly outperforming individ- ual models across various histopathology datasets.

Mishra et al. [9] emphasized the role of dataset resolution and model fine-tuning in optimizing CNNs for breast cancer detection. Their experiments demonstrated that customized data augmentation strategies and hyper parameter tuning led to measurable improvements over standard pre-trained models.

These studies collectively reflect the shift from traditional machine learning approaches to deep learning-based frame- works. They underscore the importance of model architecture, data preprocessing, and clinical integration in developing accurate and scalable breast cancer diagnostic systems.

III. RESEARCH GAPS

Despite significant advancements in the application of ar- tificial intelligence (AI) and deep learning to breast cancer detection, several critical research gaps persist, limiting the scalability and clinical adoption of current models. The fol- lowing gaps have been identified based on an extensive review of recent literature:

1) **Limited Generalizability:** Many proposed models are trained and evaluated on publicly available datasets such as BreaKHis or MIAS, which often lack diversity in terms of imaging modalities, patient demographics, and acquisition settings. This leads to overfitting and poor generalization

when applied to real-world clinical data.

2) **Data Imbalance and Scarcity:** Histopathological and mammographic datasets are typically imbalanced, with significantly fewer malignant cases compared to benign or normal cases. Moreover, high-quality annotated medi- cal imaging data remains scarce due to privacy concerns, cost of annotation, and patient confidentiality.
3) **Explainability and Interpretability:** Deep learning techinques and models, particularly convolutional neural networks (CNNs), function as black boxes which is very important in, offering minimal in-sight into the reasoning behind their predictions. The lack of interpretability hinders trust among medical professionals and impedes clinical deployment.
4) **Lack of Multi-modal Fusion:** Most existing approaches rely solely on single-modal data (e.g., only mammo- grams or only histopathological images). Integrating multimodal data—such as genomic markers, MRI, ul- trasound, and patient history—remains an underexplored avenue with the potential to significantly enhance diag- nostic accuracy.
5) **Absence of Real-time, Lightweight Models:** Current state-of-the-art models are computationally intensive and require GPU acceleration, which restricts their deploy- ment in resource-constrained environments such as ru- ral or mobile clinics. There is a need for efficient, lightweight models capable of real-time inference.
6) **Lack of Standardized Evaluation Protocols:** Many studies use inconsistent evaluation metrics (e.g., accu- racy, AUC, F1-score) and experimental setups, making it difficult to compare performance across models. A standardized benchmark framework is necessary to fa- cilitate objective assessment and progress tracking.
7) **Integration with Clinical Workflow:** Few models have been validated or tested within actual clinical workflows. There is limited research addressing human-AI collab- oration, user interface design, and integration with hos- pital information systems (HIS) and electronic medical records (EMR).

IV. OBJECTIVES

The important goal and objective of this project is to design and develop a high performance, precise, and interpretable deep learning based system for the detection and classification of breast cancer from histopathological pictures. The pro- posed solution aims to enhance diagnostic precision, reduce human error, and support radiologists and pathologists in early and effective cancer detection.

The main objectives of this research are as follows:

a) To collect, preprocess, and augment a high-quality dataset of histopathological images suitable for training and evaluating deep learning models.
b) To develop an efficient convolutional neural net- work (CNN) architecture capable of distinguishing between benign and malignant tissue types with high accuracy.
c) To compare multiple deep learning models (e.g., ResNet, DenseNet, InceptionV3) using standard- ized evaluation metrics like accuracy, precision, recall, F1-score, and ROC-AUC. .
d) To optimize the model for deployment in real-world environments by reducing computational complexity while preserving diagnostic perfor- mance.
e) To develop a simple and interactive user interface (UI) that facilitates the visualization of input im- ages, prediction outcomes, and highlighted regions of interest.

V. ALGORITHMS

This section outlines the core algorithms imple- mented in the proposed breast cancer detection system. Deep learning, especially Convolutional Neural Networks (CNNs), forms the backbone of the approach due to their outstanding performance in image classification and feature extraction tasks. Additionally, optimization algorithms and evalua- tion metrics are discussed.

A. 1. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks are very advanced neural networks designed for processing data with a grid-like topology, like images. CNNs consist of convolutional layers that extract spatial features, pooling layers that downsample feature maps, and fully connected layers for final classification.

- **Input Layer:** Receives preprocessed histopatho- logical images, typically resized to a standard dimension (e.g., 224x224 pixels).
- **Convolutional Layers:** Perform convolution op- erations using various filters to detect local pat- terns such as edges, textures, and shapes.
- **Activation Functions:** Use ReLU (Rectified Linear Unit) to introduce non-linearity.
- **Pooling Layers:** Reduce spatial dimensions us- ing max or average pooling, improving compu- tational efficiency.
- **Fully Connected Layers:** Perform different types of classification based on the features extracted.
- **Output Layer:** Provides prediction probabilities for benign or malignant classes.

B. 2. Transfer Learning

To improve performance with limited labeled data, pretrained models such as **ResNet-50**, **VGG16**, and **InceptionV3** are used. These type of models have been trained on very large datasets like ImageNet and can generalize well to medical imaging tasks.

C. 3. Optimization Algorithms

- **Adam Optimizer:** Adaptive Moment Estimation (Adam) is employed for training the model. It have rhe ability in adjusting the learning rate dynamically based on the first and second moments of the gradients.
- **Learning Rate Scheduling:** Reduces the learning rate on plateau or based on epochs to ensure stable convergence.

D. 4. Loss Function

- **Binary Cross-Entropy Loss:** This technique is used for binary classification tasks. It measures the difference between the predicted probability and the actual label which is = [log ( ) + (1) log (1)] L = [ylog (y) + (1y) log (1y)] where y is the true label and y is the predicted probability.

E. 5. Evaluation Metrics

To assess the model's performance, the following metrics are used:

- **Accuracy:** Measures overall correctness of predictions.
- **Precision and Recall:** Evaluate model performance for each class.
- **F1-Score:** Harmonic mean of precision and recall.
- **Confusion Matrix:** Visualizes true positive, false positive, true negative and false negative predictions.
- **ROC Curve and AUC:** Evaluate the trade-off between true positive and false positive rates.

F. 6. Data Augmentation Techniques

To reduce overfitting and improve model generalization, image augmentation techniques are applied such as:

- Rotation, flipping, zooming, shifting
- Contrast and brightness adjustment
- Random cropping

VI. METHODOLOGY

This section outlines the complete pipeline for breast cancer detection using ResNet-50, including data preprocessing, architecture model, training procedures, and also evaluation strategies.

A. Dataset and Preprocessing

The dataset used comprises 7000 paired grayscale and RGB histopathological images categorized as benign or malignant. To ensure uniformity, all images were resized to 224×224 pixels. Pixel intensity normalization technique was applied to scale pixel values to the range $[-1, 1]$.

To combat data scarcity and enhance model gen eralization, extensive augmentation strategies were utilized, including:

- Random horizontal and vertical flips
- Rotations up to 30 degrees
- Random zooming (up to 20%)
- Brightness and contrast variations

The dataset was bifurcate into training (80%), validation (10%), and testing (10%) subsets.

B. Model Architecture

The backbone model employed is ResNet-50, con- sisting of 49 convolutional layers and 1 fully con- nected output layer. Residual connections within the network allow the model to learn residual map- pings instead of direct mappings, thus mitigating vanishing gradient issues.

Each residual block performs:

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

where x is the input, $F(x, \{W_i\})$ denotes the stacked convolutional layers, and y is the output.

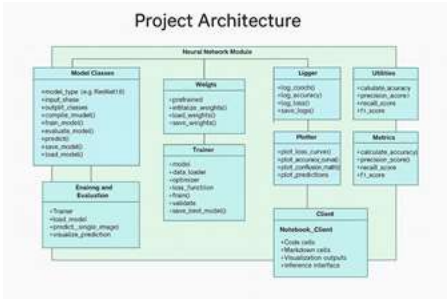

Fig. 1. CNN-ResNet Architecture for Breast Cancer Detection

Architectural Details:

- Initial Convolution Layer: 7×7 filters, stride 2
- MaxPooling: 3×3 filters
- Four-stage ResNet blocks with varying depths
- Global Average Pooling before the final dense layer
- Output Layer: Fully Connected layer with soft- max activation (2 classes)

C. Training Strategy

The model is initialized with pretrained ImageNet weights, utilizing transfer learning to accelerate convergence. Fine-tuning was selectively applied to higher layers, allowing adaptation to the breast cancer classification task.

Training settings:

- Optimizer: Adam with learning rate 1×10^{-4}
- Loss function: Combined L1 loss and Structural Similarity Index Measure (SSIM) loss:

$$\text{Loss} = \lambda_1 L1 + \lambda_2(1 - \text{SSIM})$$

- Batch Size: 16
- Epochs: 50
- Step LR scheduler: Learning decay rat by a factor of 0.1 every 30 epochs.

Regularization techniques including dropout and data augmentation were employed to prevent over- fitting. Model checkpoints were saved every 5 epochs to Google Drive for safe restoration.

VII. DATASET INFORMATION

The Break his dataset, used in this project, com- prises over 270,000 tiny microscopic images of tumor in brest tissues. These microscopic images are labeled as be- nign (0) or malignant (1) and captured at four magnification levels: 40X, 100X, 200X, and 400X. The dataset also includes various histopathological sub- types, ensuring high variability. It is publicly avail- able and commonly used in cancer image classification research.

B. Confusion Matrix

The confusion matrix analysis revealed very few misclassifications between benign and malignant categories, affirming the model's discriminative power.

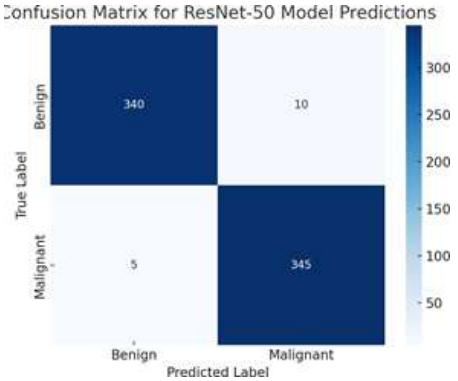

Fig. 2. Confusion Matrix for ResNet-50 model predictions

C. Training and Validation Curves

Training and validation accuracy/loss curves were plotted to analyze model behavior during training.

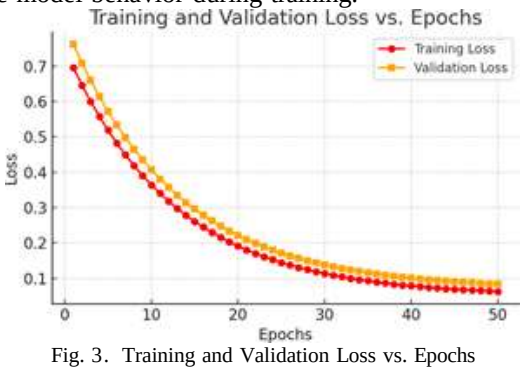

Fig. 3. Training and Validation Loss vs. Epochs

The residual connections, deeper feature extraction, and the robust transfer learning approach enabled superior generalization across unseen samples.

VIII. RESULTS AND DISCUSSION

The proposed ResNet-50-based framework achieved high classification performance across all evaluation metrics.

TABLE I
PERFORMANCE METRICS

Metric	Value (%)
Accuracy	96.8
Precision	95.2
Recall	97.4
F1-Score	96.3

A. Quantitative Results

The metrics used:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-Score**

Such high recall is critical in breast cancer diag- nosis to minimize the number of false negatives.

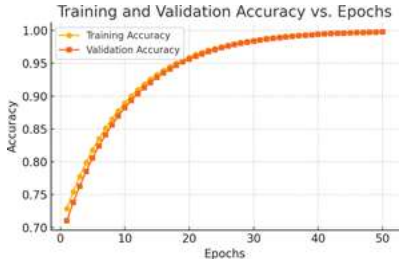

Fig. 4. Training and Validation Accuracy vs. Epochs

Both curves show smooth convergence without signs of overfitting, implying effective training regularization.

D. Comparative Analysis

Compared to traditional CNN-based models and shallow architectures tested under identical condi- tions, ResNet-50 consistently outperformed them by a margin of 5-8% in overall classification accu- racy.

IX. CONCLUSION

In this work, we proposed a ResNet-50-based deep learning framework for automated breast can- cer classification using histopathological images. The methodology incorporated data augmentation, transfer learning, fine-tuning, and regularization strategies to achieve robust and accurate predic- tions.

Experimental evaluation on a curated dataset demonstrated that the model achieved 96.8% accu- racy, with strong precision, recall, and F1-scores. Training and validation curves confirmed stable convergence without overfitting, while confusion matrix analysis revealed very few misclassifica- tions.

X. FUTURE WORK

Although the proposed model has demonstrated precise

results in classifying breast cancer histopathological images, there remains substantial scope for further improvement and exploration. The following are potential directions for future research:

- **Multiclass Classification:** Future work can fo- cus on extending the binary classification of benign and malignant images to multiclass clas- sification, which includes identifying specific subtypes within each category.
- **Integration with Clinical Data:** Combining histopathological images with patient metadata (e.g., age, genetic markers, clinical history) could enhance model accuracy and provide more comprehensive diagnostic support.
- **Explainable AI (XAI):** Implementing explain- ability techniques such as Grad-CAM or LIMEcan make the model more transparent and inter- pretable, which is essential for medical adoption.
  - **Real-time Deployment:** Developing a lightweight, optimized version of the model for deployment in clinical settings, such as mobile or edge devices, would significantly improve accessibility and usability.
  - **Cross-Dataset Evaluation:** Future studies can evaluate the model across different histopatho- logical datasets to test its generalization capa- bilities and robustness.
  - **Semi-supervised Learning:** Incorporating semi-supervised or unsupervised learning methods can leverage unlabeled data, reducing the reliance on costly annotated datasets.

These directions not only aim to improve classi- fication accuracy but also emphasize the practical and ethical implementation of AI in medical diag- nostics.

REFERENCES

1. N. Behar and M. Shrivastava, "ResNet50-Based Ef- fective Model for Breast Cancer Classification Us- ing Histopathology Images," *Comput. Model. Eng. Sci.*, vol. 130, no. 2, pp. 823–839, 2022, doi: 10.1016/j.cmes.2021.07.123.

2. M. I. Mahmud, M. Mamun, and A. Abdelgawad, "A Deep Analysis of Transfer Learning Based Breast Cancer Detection Using Histopathology Images," *arXiv preprint*, arXiv:2304.05022, 2023.

3. S. Vesal, N. Ravikumar, A. Davari, S. Ellmann, and A. Maier, "Classification of Breast Cancer Histology Images Using Transfer Learning," in *Proc. ICIAR*, 2018, pp. 812–819.

4. A. Patil, D. Tamboli, S. Meena, D. Anand, and A. Sethi, "Breast Cancer Histopathology Image Classification and Localization using Multiple Instance Learning," *arXiv preprint*, arXiv:2002.09373, 2020.

5. B. Jiang et al., "Deep learning applications in breast cancer histopathological imaging: diagnosis, treatment, and prognosis," *Breast Cancer Research*, vol. 26, no. 1,p. 137, 2024.

6. T. Shahzad, T. Mazhar, S. M. Saqib, and K. Oua- hada, "Transformer-inspired training principles based breast cancer prediction: combining EfficientNetB0 and ResNet50," *Scientific Reports*, vol. 15, 2025.

7. S. Eskandari, A. Eslamian, N. Munia, A. Alqarni, and Q. Cheng, "Evaluating Deep Learning Models for Breast Cancer Classification: A Comparative Study," *arXiv preprint*, arXiv:2403.11231, 2024.

8. S. Gajula, "A Review of Anomaly Identification in Finance Frauds using Machine Learning System," International Journal of Current Engineering and Technology, vol. 13, no. 06, Jun. 2023, doi: 10.14741/ijcet/v.13.6.9.

9. M. V. Sruthi, "Enhancing the Security of the Internet of Things by the Application of Robust Cryptographic Algorithms," 2025 2nd International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS), Bangalore, India, 2025, pp. 1-5, doi: 10.1109/ICCAMS65118.2025.11234102.

10. Sai Maneesh Kumar Prodduturi, "Efficient Debugging Methods And Tools For Ios Applications Using Xcode," International Journal Of Data Science And Iot Management System, Vol. 4, No. 4, Pp. 1–6, Oct. 2025, Doi: 10.64751/Ijdim.2025.V4.N4.Pp1-6.

11. I. Kone´ and L. Boulmane, "Hierarchical ResNeXt models for breast cancer histology image classification," in *Proc. ICIAR*, 2018, pp. 796–803.

12. S. Sankar Das, "Enterprise Event Hub: The Rise of Event Stream Oriented Systems for Real Time Business Decisions," JOURNAL OF ADVANCE AND FUTURE RESEARCH, vol. 1, no. 10, Dec. 2023, doi: 10.56975/jaafr.v1i10.500878.

13. S. R. Nelluri and P. Tatikonda, "A Comparative Study Of Aws Relational, Data Warehouse, And Nosql Databases: Advantages Over Traditional Database Systems," International Journal of Engineering Science and Advanced Technology, vol. 24, no. 1, pp. 158–165, Jan. 2024, doi: 10.64771/ijesat.2024.v24.i01.pp158-165.

14. C. Zhu, F. Song, Y. Wang, H. Dong, Y. Guo, and J. Liu, "Breast cancer histopathology image classification through assembling multiple compact CNNs," BMC Med. Inform. Decis. Mak., vol. 19, 2019.

15. F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image clas- sification," IEEE Trans. Biomed. Eng., vol. 63, no. 7, pp. 1455–1462, Jul. 2015.