**A Hybrid Deep Learning Approach for Breast Cancer Classification: VGG-19, Bi-LSTM, and Attention Mechanisms**

Meriem Touhami1, a), Mohammad Faizal Ahmad Fauzi1, 2, b), Md Serajun Nabi1, c) and Sarina Mansoor 1, d)

1Faculty of Artificial Intelligence and Engineering, Multimedia University, Persiaran Multimedia, 63100 Cyberjava, Malaysia

2Centre for Image and vision Computing, COE for Artificial Intelligence, Multimedia University, Persiaran Multimedia, 63100 Cyberjava, Malaysia

*d) Corresponding author: sarina.mansor@mmu.edu.my*

*a)*[*1231400795@student.mmu.edu.my*](mailto:1231400795@student.mmu.edu.my)

*b)*[*mohammad.faizal.ahmad.fauzi@mmu.edu.my*](mailto:mohammad.faizal.ahmad.fauzi@mmu.edu.my) *c)*[*md.serajun.nabi@student.mmu.edu.my*](mailto:md.serajun.nabi@student.mmu.edu.my)

**Abstract.** Breast cancer remains a serious and life-threatening disease worldwide, emphasizing the critical need for accurate and early diagnosis. Deep learning techniques have demonstrated strong potential in supporting medical professionals by enabling faster and more precise classification of breast cancer. A hybrid solution is proposed in this study for breast cancer classification that leverages the strengths of the VGG-19 convolutional neural network and a channel attention mechanism. Additionally, we integrate a Bidirectional Long Short-Term Memory (Bi-LSTM) network to capture long-term dependencies and preserve contextual information, enabling more robust feature learning. In this study, the BreakHis dataset was used, which contains histopathology images at four different magnifications. Our proposed hybrid model achieved an accuracy of 98% for 40x magnification, 97% for both 100x and 200x magnifications, and 96% for 400x magnification. These experimental results show that the developed hybrid model exceeds the performance of traditional deep learning architectures, offering a reliable and effective tool for accurate breast cancer diagnosis.

# Introduction

Breast cancer remains a major global health threat and a top cause of death in women, claiming over 40,000 lives annually [**Error! Reference source not found.**]. As a significant health challenge of the 21st century, cancer manifests in various forms across different organs and tissues, with breast cancer being a primary concern, particularly for women. As stated by the Global Cancer Observatory, breast cancer caused approximately 684,996 deaths globally in 2020, highlighting its devastating impact. In 2022 alone, it was responsible for 670,000 deaths, making it the most widespread form of cancer among women in 157 out of 185 countries [**Error! Reference source not found.**].

The diagnosis and classification of breast cancer rely heavily on examination of biopsy tissue through microscopic imaging, which help specialists determine the microscopic type, severity level, and progression stage of the cancer [**Error! Reference source not found.**] [**Error! Reference source not found.**]. This process is critical for tailoring effective treatment plans. Additionally, advancements in detection technologies have significantly improved early diagnosis. Conventional imaging techniques are widely used to detect breast cancer. When combined with histopathological imaging and genetic analysis, these methods serve a crucial function in diagnosis and treatment strategy formulation [**Error! Reference source not found.**].The evolution of medical imaging has revolutionized the field of diagnostics, enabling the detection of even the smallest cellular changes with remarkable precision. This progress has been instrumental in the fight against diseases like breast cancer.

One of the major hurdles in leveraging deep learning for medical applications is the process of choosing the appropriate model and optimizing its hyperparameters[**Error! Reference source not found.**],given the vast array of neural network architectures, identifying the one best suited for a particular dataset or diagnostic task can be overwhelming. Additionally, hyperparameter optimization, which entails fine-tuning settings such as learning rates, batch sizes, and layer configurations to enhance model performance, often requires significant computational resources and expertise. Despite these complexities, the potential benefits such as increased diagnostic precision and the ability to detect diseases at earlier stages justify the investment of time and effort, as they can lead to substantial advancements in patient care and outcomes [**Error! Reference source not found.**].

The fundamental objective of this study is focused on enhancing the precision and performance regarding breast cancer classification by leveraging state-of the art technologies, including VGG19, Bidirectional Long Short-Term Memory (Bi-LSTM) networks, and attention mechanisms. By using Convolutional Neural Networks (CNNs) like VGG19, the study focuses on extracting unique and discriminative features from histopathological images, particularly those from the BreakHis dataset. The integration of Bi-LSTM networks into the hybrid model addresses challenges such as the vanishing gradient problem and ensures the retention of long-term dependencies across training epochs. Additionally, the attention mechanism is employed to prioritize the most relevant features within the images, thereby refining the classification process.

# Proposed methodology

The proposed framework of the study presented in FIGURE 1begins by applying data augmentation to ensure balanced class representation. All images are resized to 224x224 pixels to match VGG-19’s input requirements. The dataset is subsequently divided into training, validation, and testing subsets for robust evaluation. A hybrid model is constructed, leveraging VGG-19’s feature extraction capabilities, enhanced by a channel attention mechanism, and integrated with LSTM to capture temporal dependencies. The model is compiled and trained using the training data, and its performance is evaluated on the test set to assess accuracy and generalization.

A diagram of a model

AI-generated content may be incorrect.

**FIGURE 1.** Overall framework

## The dataset

The BreaKHis dataset (images from the dataset are shown in FIGURE 2.) The dataset comprises a total of 7,909 instances, categorized into two groups: benign and malignant. The benign category includes 2,440 instances and is regarded as the underrepresented class, whereas the malignant category contains 5,429 instances, making it the dominant class. These samples were acquired from 82 individual patients and captured under four distinct magnification levels: 40×, 100×, 200×, and 400×. Each of the benign and malignant groups is further subdivided into four specific histological types. The benign group consists of Adenosis (A), Fibroadenoma (F), Tubular Adenoma (TA), and Phyllodes Tumor (PT). The malignant groups include Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC), and Papillary Carcinoma (PC). The detailed number of samples in each class is presented in TABLE 1 [**Error! Reference source not found.**].

**TABLE 1.** The Breakhis dataset description

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Class** | **Type** | **40x** | **100x** | **200x** | **400x** |
| Benign | Adenosis | 114 | 113 | 111 | 106 |
|  | Fibro | 253 | 260 | 264 | 237 |
|  | Tublar | 109 | 121 | 108 | 115 |
|  | phylodes | 149 | 150 | 140 | 130 |
| Malignant | Ductal | 864 | 903 | 896 | 788 |
|  | Lobular | 156 | 170 | 163 | 137 |
|  | Mucinous | 205 | 222 | 196 | 169 |
|  | Papillary | 145 | 142 | 135 | 138 |

A collage of images of cells

AI-generated content may be incorrect.

**FIGURE 2.** Sample images from the Breakhis dataset

### Dataset pre-processing

Initially, the histopathological image sizes were standardized to 224x224 Pixel dimensions modified to align with the input needs of the VGG-19 model, Following this, the dataset underwent a data augmentation process to address class imbalance, oversampling techniques were applied to the minority classes, artificially increasing their representation through producing more variants of existing images through transformations such as rotation, flipping, and scaling.

## Models

### VGG19 Model Structure

VGG-19 is an extensively layered convolutional network, totaling 19 layers, including 16 convolutional and 3 fully connected layers. It takes input images of size 224×224×3 and uses 3×3 convolutional filters with stride 1 and padding, followed by ReLU activations. The network has five convolutional blocks (with 2, 2, 4, 4, and 4 layers), each ending with 2×2 max-pooling for downsampling. After flattening the output, it is processed through pair of fully connected layers of 4096 units (ReLU), and a final layer with 1000 softmax units for classification. FIGURE 3. (a) Shows the architecture of the model. [**Error! Reference source not found.**]

### Bi-LSTM Model Architecture

Bidirectional LSTM (Bi-LSTM) is an advanced RNN architecture that captures both past and future context in sequential data by using two LSTM layers, one forward and one backward. Each LSTM unit includes input, forget, and output gates to control information flow. This bidirectional setup enhances feature extraction and improves performance in various deep learning tasks by leveraging context from both directions.model architecture presented in Figure 3(b) [**Error! Reference source not found.**].

|  |  |
| --- | --- |
|  |  |
|  | (b) |

**FIGURE 3**. (a) VGG-19 model architecture, (b) Bi-LSTM model architecture

### Attention mechanism

The Channel Attention mechanism (shown in FIGURE 4) was used in this study to enhance the representational power of a neural network by focusing on the most informative channels in a feature map. It works by applying global average pooling and global max pooling across spatial dimensions to capture channel-wise context. These pooled features are passed through a shared multi-layer perceptron, and the outputs are combined and activated using a sigmoid function to generate attention weights. These weights are then used to rescale the original feature map, allowing the network to emphasize important features [**Error! Reference source not found.**].

A diagram of a shared mlp

AI-generated content may be incorrect.

**FIGURE 4.** Channel attention architecture

## Hybrid model architecture

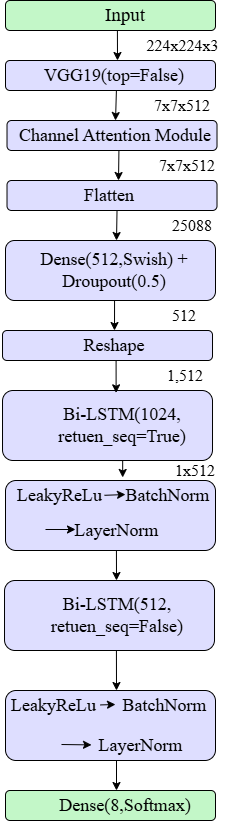
This model (FIGURE 5), leverages VGG19 as its backbone for feature extraction,The VGG19 model is loaded with ImageNet weights and fine-tuned by unfreezing the last 7 layers while keeping the earlier layers frozen to retain learned features.On top of the VGG19 backbone, the architecture incorporates a Channel Attention (CA) mechanism to enhance feature representation. This mechanism uses global average pooling and global max pooling to compute attention scores, which are then applied to the feature maps to emphasize important channels. After applying channel attention, the features are flattened and passed through a fully connected (Dense) layer for further processing, with 512 units and a Swish activation function, followed by a Dropout layer (with a 50% dropout rate) to prevent overfitting.The model then reshapes the output to a 3D tensor to feed into a Bidirectional LSTM layer, which captures sequential dependencies in the feature space. The Bi-LSTM layers are enhanced with LeakyReLU activation, Batch Normalization, and Layer Normalization to stabilize training and improve performance. The final output layer consists of a Dense layer with 8 units and a Softmax activation function, corresponding to the 8 classes in the dataset. Overall, this model combines the strengths of CNNs, attention mechanisms, and recurrent networks to achieve robust performance on the BreakHis dataset.

*Hyperparameter tuning*

This model uses the Adam optimizer with a learning rate of 0.0001 and a batch size of 32. Training runs for up to 30 epochs with EarlyStopping (patience=8) to prevent overfitting. ReduceLROnPlateau adjusts the learning rate when validation loss stops improving. Bidirectional LSTM layers with 1024 and 512 units, along with LeakyReLU, BatchNormalization, and LayerNormalization, help capture dependencies and stabilize training. A dropout rate of 0.5 is applied for regularization.

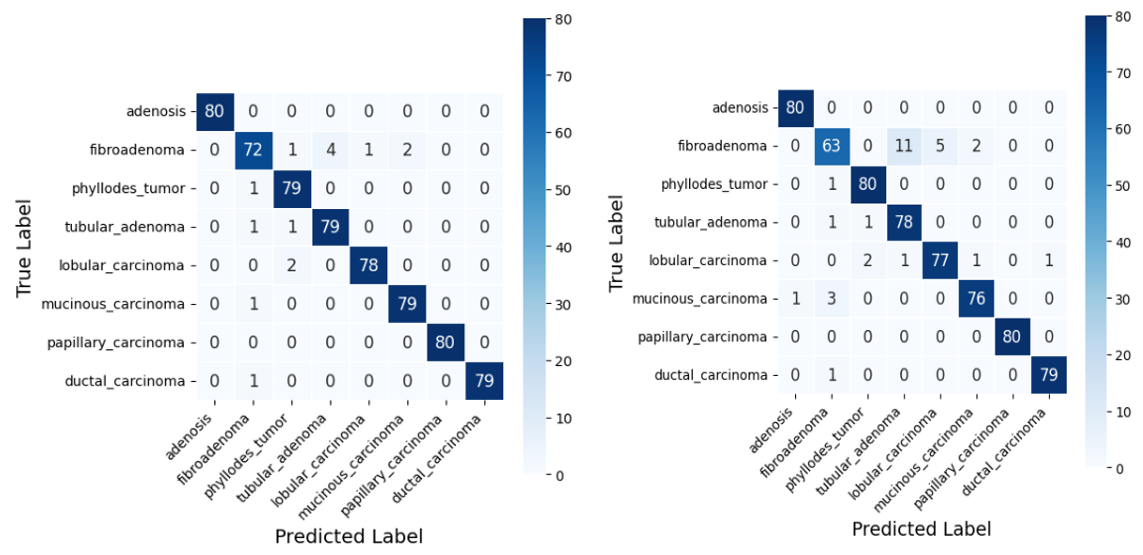
# Results and discussion

The proposed hybrid VGG-19, Channel Attention, and Bi-LSTM model demonstrates significant improvements over standalone models (comparative results presented in TABLE 2) in classifying histopathology images across multiple magnifications. While the standalone VGG-19 achieved accuracies of 94% (40x), 92% (100x), 96% (200x), and 91% (400x), and the Bi-LSTM alone scored lower (89%, 93%, 94%, and 83%, respectively), the hybrid model attained superior performance 98% (40x), 97% (100x and 200x), and 96% (400x). This enhancement can be attributed to the synergistic integration of the Channel Attention mechanism, which adaptively highlights discriminative features, and Bi-LSTM, which captures long-range spatial dependencies.

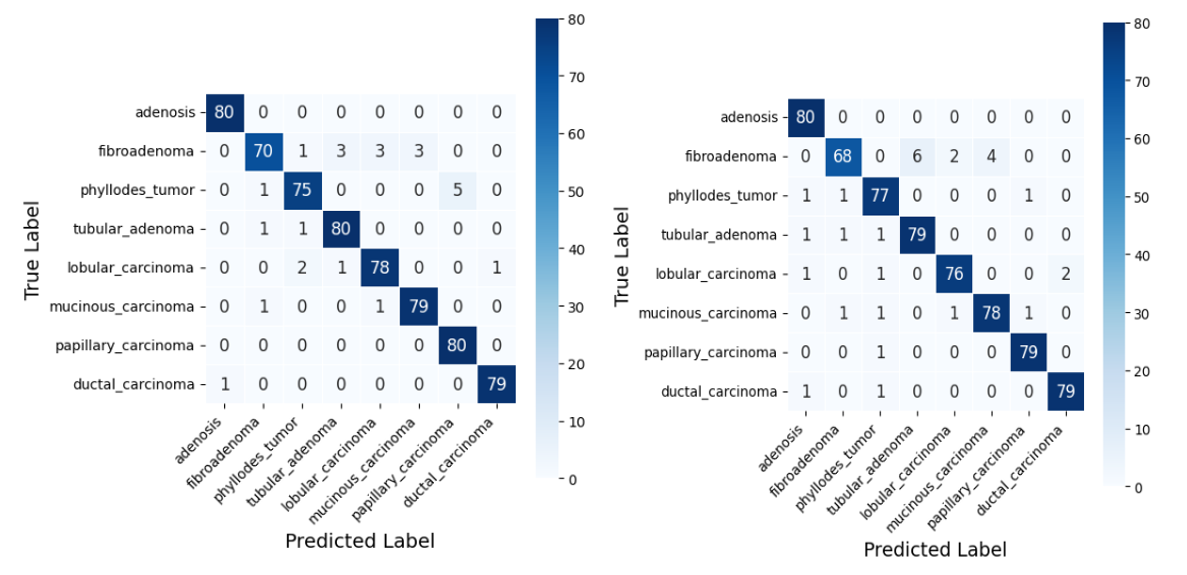


**FIGURE 5.** Hybrid model architecture

The confusion matrix, presented in FIGURE 6., provides a meticulous evaluation in terms of the model’s classification capabilityby highlighting the true positives, true negatives, false positives, and false negatives for each of the eight classes. This comprehensive evaluation enables a more comprehensive insight into the model’s capability to accurately identify the differences betweenthe various classes. The minimal number of misclassifications observed underscores the effectiveness of the Bi-LSTM component in retaining and utilizing relevant information across multiple epochs. By capturing sequential dependencies and understanding temporal relationships, the Bi-LSTM significantly enhances the internal logic of the model, leading to more accurate and consistent predictions. This detailed insight affirms the robustness of the hybrid approach in handling complex histopathology image classification tasks.



1. (b)



(c) (d)

**FIGURE 6.** Confusion matrix for all four magnifications, where (a) is 40x, (b) 100x, (c) 200x and (d) 400

**TABLE 2.** Accuracy comparison between our model and baseline models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Models** | **40X** | **100X** | **200X** | **400X** |
| VGG-19 | 0.94 | 0.92 | 0.96 | 0.91 |
| Bi-LSTM | 0.85 | 0.87 | 0.88 | 0.83 |
| Ours | 0.98 | 0.97 | 0.97 | 0.96 |

# Conclusion

In conclusion, the combination of VGG-19, the Channel Attention mechanism, and Bi-LSTM proved to be highly effective in classifying breast cancer types, achieving impressive accuracy across the four magnification levels. This demonstrates the strength of integrating spatial feature extraction, adaptive focus, and sequential modeling to accurately differentiate between various cancer types. The hybrid model’s ability to capture intricate patterns and contextual relationships contributed significantly to its robust performance. For future work, the model can be further enhanced to achieve even higher accuracy by exploring the integration of additional architectures or leveraging complementary techniques. Additionally, applying this approach to other medical imaging modalities could validate its versatility and effectiveness in broader clinical applications, enabling more precise and reliable diagnostic tool.

# References

1. A. El-Nabawy, N. El-Bendary, and N. A. Belal, “A feature-fusion framework of clinical, genomics, and histopathological data for METABRIC breast cancer subtype classification,” *Applied Soft Computing* **91**, 106238 (2020).
2. World Health Organization, “Breast cancer,” (n.d.). https://www.who.int/news-room/fact-sheets/detail/breast-cancer (accessed 2024).
3. M. Z. Alom, C. Yakopcic, Mst. S. Nasrin, T. M. Taha, and V. K. Asari, “Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network,” *Journal of Digital Imaging* **32**(4), 605–617 (2019).
4. R. R. Devi and G. S. Anandhamala, “Recent trends in medical imaging modalities and challenges for diagnosing breast cancer,” *Biomedical and Pharmacology Journal* **11**(3), 1649–1658 (2018).
5. A. B. Nassif, M. A. Talib, Q. Nasir, Y. Afadar, and O. Elgendy, “Breast cancer detection using artificial intelligence techniques: A systematic literature review,” *Artificial Intelligence in Medicine* **127**, 102276 (2022).
6. A. Pfob, S.-C. Lu, and C. Sidey-Gibbons, “Machine learning in medicine: A practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison,” *BMC Medical Research Methodology* **22**(1), 282 (2022).
7. A. J. Dunn, S. Coniglio, M. ElRefai, P. R. Roberts, B. M. Wiles, and A. B. Zemkoho, “Deep learning and hyperparameter optimization for assessing one’s eligibility for a subcutaneous implantable cardioverter-defibrillator,” *Annals of Operations Research* **328**(1), 309–335 (2023).
8. F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “A dataset for breast cancer histopathological image classification,” *IEEE Transactions on Biomedical Engineering* **63**(7), 1455–1462 (2016).
9. I. M. A. Agastya and A. Setyanto, “Classification of Indonesian batik using deep learning techniques and data augmentation,” in *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)* (IEEE, Yogyakarta, Indonesia, 2018), pp. 27–31.
10. A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” *Neural Networks* **18**(5–6), 602–610 (2005).
11. J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-excitation networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**(8), 2011–2023 (2020).