Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2025 (43) 4** : 1–18.
ISSN-0037-8712
doi:10.5269/bspm.78926

GWO-Tuned Deep Learning Model with Memory Mechanism for Accurate Mammographic Breast Cancer Diagnosis

Mustafa Raad Ali, Normy Norfiza Abdul Razak and Rashid Jan*

ABSTRACT: Automated and accurate diagnosis of breast cancer from mammography images remains a critical challenge in medical imaging, necessitating advanced computational approaches to improve early detection rates. This paper proposes a hybrid deep learning framework designed to enhance the diagnostic precision of breast cancer. The methodology commences with a meticulous preprocessing pipeline, including Contrast Limited Adaptive Histogram Equalization (CLAHE) with optimized Rayleigh distribution and clip limits, a novel background cropping technique to focus on relevant tissue, pixel intensity adjustments for zero-valued regions, and data augmentation through rotations, flips, and zooming. Feature extraction is subsequently performed using a pre-trained ResNet-50 architecture, adapted via transfer learning with fine-tuning of its terminal layers and a custom dense layer. The resultant high-dimensional feature vectors are then refined using Linear Discriminant Analysis (LDA) to enhance class separability while reducing dimensionality. For classification, a Bidirectional Gated Recurrent Unit (BiGRU) network is employed, with its crucial hyperparameters (number of hidden units, learning rate, decay factor, and batch size) systematically optimized using the Grey Wolf Optimizer (GWO). The developed model demonstrates strong diagnostic performance, achieving an accuracy of 98.44% on the INbreast dataset, highlighting its potential as an effective tool for computer-aided breast cancer diagnosis.

Key Words: Breast Cancer, mammography, Grey Wolf optimizer, ResNet-50, BiLSTM, contrast limited adaptive histogram equalization.

Contents

* Corresponding author.
2010 *Mathematics Subject Classification:* 68T07.
Submitted September 12, 2025. Published December 20, 2025

1. Introduction

Breast cancer remains a leading cause of cancer-related mortality among women worldwide, underscoring the critical importance of early and accurate detection to improve patient prognosis and treatment efficacy. Mammography is the most widely utilized screening modality for breast cancer; however, its interpretation can be challenging due to the subtle nature of early malignant signs, high inter-observer variability among radiologists, and the sheer volume of images requiring analysis, which can lead to diagnostic errors or delays. These inherent limitations in manual screening processes highlight an urgent need for robust and reliable automated systems. Consequently, the development of advanced Computer-Aided Diagnosis (CAD) tools, particularly those leveraging the sophisticated pattern recognition capabilities of deep learning, has become an area of intense research, offering the potential to enhance diagnostic accuracy, reduce workload, and ultimately contribute to better clinical outcomes in the fight against breast cancer. Recent research highlights the potential of deep learning techniques for breast cancer diagnosis using mammographic images. Convolutional Neural Networks (CNNs) have shown promising results in classifying breast lesions as benign or malignant [1,2,3]. These approaches can assist radiologists in early detection and improve diagnostic accuracy. One study using a ResNet-50 CNN achieved 93% classification accuracy on the INbreast dataset [2]. Various CNN architectures like VGG19, Inception-Net, and ResNet50 have been applied successfully to this task [3]. The process typically involves image preprocessing, segmentation, and post-processing steps [1]. By leveraging transfer learning and pre-trained models, these deep learning methods can potentially reduce error rates in screening mammograms and contribute to improved breast cancer diagnosis and patient outcomes [4]. These techniques have been applied to multiple imaging modalities, including mammography, ultrasound, MRI, and nuclear medicine imaging [5,6]. While deep learning shows potential to improve breast cancer screening workflows and early detection, larger trials are needed to fully determine its clinical value, particularly for ultrasound and MRI [5]. Additionally, legal and ethical considerations must be addressed before widespread clinical implementation [5].

Lin et al. developed a two-stage model combining support vector machines and convolutional neural networks (CNNs), achieving 94.20% accuracy on mammography images [7]. Chorianopoulos et al. evaluated various CNN architectures, reporting accuracies of 96.82% on ultrasounds and 88.23–91.04% on histopathology images [8]. Suguna et al. compared CNN designs like VGG16 and ResNet50 to traditional machine learning techniques using a large breast cancer dataset [9]. All three studies emphasized the potential of deep learning methods to improve early detection and diagnosis of breast cancer. The high accuracies achieved across different imaging modalities (mammography, ultrasound, and histopathology) suggest that these approaches could serve as valuable tools to assist radiologists and pathologists in clinical settings [7,8,9]. Gupta et al. (2022) compared four pre-trained Convolutional Neural Network (CNN) models (VGG16, VGG19, Inception v3, and SqueezeNet) combined with various machine learning classifiers for early detection of breast cancer in mammograms [10]. Their research aimed to provide an overview of current diagnostic techniques to improve patient survival rates. Similarly, Noor Eldin et al. (2021) investigated deep learning methods for breast cancer diagnosis using microscopy biopsy images [11]. They evaluated several CNN architectures, with Densenet169, Resnet50, and Resnet101 achieving the highest accuracies without preprocessing. The study demonstrated that data augmentation and segmentation techniques could further improve model performance by up to 20%. By employing ensemble learning, they achieved a peak accuracy of 92.5%. Both studies highlight the potential of deep learning in enhancing breast cancer diagnosis through medical image analysis.

A U-net inspired network architecture was developed in reference [12], which demonstrated high sensitivity and specificity in identifying breast abnormalities. Sannasi Chakravarthy et al. (2022) conducted extensive experiments using various deep convolutional neural networks (CNNs) and feature fusion techniques, achieving classification accuracies of 97.93% and 96.646% on the MIAS and INbreast datasets, respectively [13]. Their study also investigated the use of principal component analysis to reduce computational costs. Similarly, the researchers in [14] employed automated deep learning methods for breast cancer diagnosis using biomedical mammogram images. Amrisha R R et al. (2023) developed a customized CNN with data augmentation and transfer learning using the Inception Net model, achieving over 94% accuracy across ultrasound, mammography, and histopathology datasets [15]. This approach demonstrated improved accuracy compared to conventional methods. Similarly, reference [16] evaluated

four deep CNN models (AlexNet, VGGNet, ResNet50, and GoogleNet) for breast cancer diagnosis using infrared thermal images. Their study showed high performance in classifying malignant and benign cancers using thermography. Both studies highlight the potential of deep learning algorithms in enhancing breast cancer detection across multiple imaging modalities, offering promising results for accurate and efficient diagnosis. Researchers have employed various CNN architectures, including VGG19, ResNet50, and DenseNet121, to analyze mammograms and classify lesions as benign or malignant [17,18].

Image augmentation techniques have been used to enhance model performance when working with limited datasets [17]. The YCbCr color space has been explored to potentially improve image quality and classification accuracy [18]. These deep learning approaches show promise in assisting radiologists, potentially increasing the precision and timeliness of breast cancer diagnosis and, consequently, improving survival rates [19]. Salman Zakareya et al. (2023) developed a model combining GoogLeNet, residual blocks, and granular computing, achieving 93% and 95% accuracy on ultrasound and histopathology images, respectively [20]. Liu et al. (2020) introduced DeepBC, integrating Inception, ResNet, and AlexNet, which attained 96.43% accuracy in image classification [21]. A Deep Learning assisted Efficient AdaBoost Algorithm (DLA-EABA) was proposed in reference [22] that achieved 97.2% accuracy, 98.3% sensitivity, and 96.5% specificity. Rajput et al. (2022) proposed a customized convolutional neural network (CNN) with a novel activation function, achieving 99% accuracy and 0.97 precision in tumor detection [23]. Ragab et al. (2021) presented a computer-aided diagnosis system using multiple deep CNNs, comparing various approaches including pre-trained fine-tuned networks, deep feature extraction with SVM classifiers, and feature fusion [24]. They achieved the highest accuracy using deep feature fusion on two datasets (CBIS-DDSM and MIAS). These studies demonstrate the potential of deep learning techniques in improving breast cancer detection and classification, offering promising alternatives to traditional diagnostic methods and potentially enhancing early detection rates. However, the reviewed literature demonstrates significant strides in applying deep learning to mammographic breast cancer diagnosis; several avenues for further enhancement remain. Many studies have successfully employed various Convolutional Neural Network (CNN) architectures; however, the synergistic potential of combining these powerful spatial feature extractors with sequence-aware models like Recurrent Neural Networks (RNNs) for mammogram classification is less explored. Furthermore, the performance of these complex models is often highly sensitive to hyperparameter configurations, yet systematic metaheuristic optimization of these parameters, particularly for the RNN component in such hybrid setups, is not consistently applied. Additionally, while preprocessing is common, a tailored pipeline addressing specific mammographic challenges like low contrast and non-informative background regions in a multi-stage, optimized manner, followed by supervised dimensionality reduction specifically aimed at enhancing class separability for subsequent sequential analysis, could further boost diagnostic accuracy.

This study aims to bridge these gaps by proposing a meticulously designed hybrid deep learning framework that leverages the strengths of ResNet-50 for feature extraction and a Bidirectional Gated Recurrent Unit (BiGRU) for classification, with the BiGRU's hyperparameters fine-tuned using the Grey Wolf Optimizer (GWO), all built upon a comprehensive preprocessing and feature refinement strategy. The main contributions of this paper are multi-fold and aim to advance the state-of-the-art in automated breast cancer diagnosis. Firstly, we introduce a novel hybrid deep learning architecture that synergistically combines a pre-trained ResNet-50 for robust feature extraction with a BiGRU network designed to capture complex patterns and dependencies within these deep features. Secondly, we demonstrate the efficacy of employing the Grey Wolf Optimizer for systematic and automated hyperparameter tuning of the BiGRU classifier, leading to an optimized model tailored for the mammography task. Thirdly, our work incorporates a specialized preprocessing pipeline, featuring optimized Contrast Limited Adaptive Histogram Equalization (CLAHE) and an innovative background cropping technique, coupled with Linear Discriminant Analysis (LDA) for effective dimensionality reduction and class discrimination enhancement. Finally, through rigorous evaluation on the public INbreast dataset, our proposed framework achieves a high diagnostic accuracy, showcasing its potential as a reliable and effective tool to assist clinicians in the early detection of breast cancer.

2. Basic Concepts

This section outlines the essential concepts and network architectures that form the foundation of the proposed method.

2.1. Gated Recurrent Unit

The Gated Recurrent Unit (GRU) is a type of Recurrent Neural Network (RNN) designed to effectively model sequential dependencies while addressing the limitations of traditional RNNs, such as the vanishing gradient problem. Introduced by Cho et al. (2014) [25], the GRU employs a gating mechanism to adaptively control the flow of information, thereby enhancing the model's ability to capture both short-term and long-term temporal dependencies. A GRU cell operates based on two principal gates: the update gate and the reset gate. The update gate determines the extent to which the previous hidden state is carried forward to the current hidden state. It is computed as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}), \tag{2.1}$$

where x_t is the input at time step t, W_z and U_z are weight matrices, and σ denotes the sigmoid activation function. The reset gate, calculated by:

$$r_t = \sigma(W_r x_t + U_r h_{t-1}), \tag{2.2}$$

controls how much of the previous hidden state should be forgotten before computing the candidate hidden state. This mechanism allows the model to discard irrelevant historical information when necessary.

Subsequently, the candidate hidden state, which incorporates the new input and the selectively reset past hidden state, is computed as:

$$\tilde{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1})), \tag{2.3}$$

where $\odot$ represents element-wise multiplication, and tanh is the hyperbolic tangent activation function. Finally, the new hidden state is derived as a linear interpolation between the previous hidden state and the candidate hidden state, modulated by the update gate:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t. \tag{2.4}$$

This formulation enables the GRU to dynamically balance between preserving past information and incorporating new input, depending on the task's temporal characteristics. Due to its relatively compact architecture—requiring fewer parameters than LSTM networks—GRU is computationally efficient and particularly well-suited for applications involving sequential data such as language modeling, speech recognition, biomedical signal processing, and time-series classification [26].

2.2. Gray Wolf Optimization

The Gray Wolf Optimization (GWO) algorithm is a nature-inspired metaheuristic proposed by Mirjalili et al. (2014) [27], mimicking the leadership hierarchy and hunting strategy of gray wolves (*Canis lupus*) in the wild. As a population-based optimization method, GWO has gained attention due to its simplicity, flexibility, and strong global search capability, making it well-suited for solving complex optimization problems in engineering, machine learning, and data mining.

In the social structure of gray wolves, the pack is led by four hierarchical levels: alpha (α), beta (β), delta (δ), and omega (ω). The α wolf represents the best candidate solution, followed by β and δ, which are the second and third best solutions, respectively, while the remaining wolves (ω) follow these three leaders. The hunting process in GWO is mathematically modeled to simulate three main stages: encircling prey, hunting, and attacking prey (exploitation).

The encircling behavior is modeled by updating the position of a search agent with respect to the position of the prey using the following equations:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)|, \tag{2.5}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D}. \tag{2.6}$$

Here, $\vec{X}_p$ represents the position of the prey, $\vec{X}$ is the current position of the gray wolf, and $\vec{A}$ and $\vec{C}$ are coefficient vectors calculated as:

$$\vec{A} = 2a \cdot \vec{r}_1 - a, \quad \vec{C} = 2 \cdot \vec{r}_2, \tag{2.7}$$

where a decreases linearly from 2 to 0 over the course of iterations, and $\vec{r}_1, \vec{r}_2$ are random vectors in [0,1].

To simulate the hunting mechanism, each search agent updates its position based on the positions of the top three wolves (α, β, and δ):

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)|, \tag{2.8}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D}, \tag{2.9}$$

$$\vec{A} = 2a \cdot \vec{r}_1 - a, \quad \vec{C} = 2 \cdot \vec{r}_2, \tag{2.10}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3}. \tag{2.11}$$

This cooperative strategy balances exploration and exploitation, guiding the search agents toward the global optimum. The reduction of the parameter a over time ensures a gradual shift from exploration to exploitation, allowing for refined convergence near optimal solutions.

2.3. ResNet-50

ResNet-50 is a deep convolutional neural network (CNN) that belongs to the Residual Network (ResNet) family, introduced by He et al. (2016) [28]. ResNet was developed to address the degradation problem in deep neural networks, where increasing network depth leads to higher training error due to vanishing gradients and optimization difficulties. The key innovation of ResNet is the introduction of residual learning through skip connections, which allow the network to learn identity mappings and thus facilitate the training of very deep architectures.

The ResNet-50 model consists of 50 layers, including one initial convolutional layer, followed by 16 residual blocks organized into four stages. Each residual block contains convolutional layers with batch normalization and ReLU activations, along with a shortcut connection that bypasses one or more layers. Formally, a residual block is expressed as:

$$y = F(x, W_i) + x, \tag{2.12}$$

where x is the input to the block, F represents the residual mapping to be learned (typically a stack of two or three convolutional layers), and W_i denotes the weights. The term x is added via identity mapping (skip connection) to the output of F, ensuring that the gradient can flow directly through the network without diminishing.

ResNet-50 adopts a bottleneck design within each residual block, where three convolutional layers are used: a 1×1 convolution to reduce dimensionality, a 3×3 convolution, and a 1×1 convolution to restore dimensionality. This design allows the network to be deep while maintaining computational efficiency. The final structure of ResNet-50 includes an average pooling layer and a fully connected layer for classification tasks. The architecture has demonstrated remarkable performance in large-scale visual recognition benchmarks, particularly on the ImageNet dataset, where it significantly outperformed earlier architectures without residual connections. The use of ResNet-50 is particularly advantageous in medical and scientific imaging tasks due to its proven generalization ability and resistance to overfitting, even when the training dataset is relatively limited.

3. Proposed Methodology

The methodology presented in this study introduces a robust deep learning pipeline for the automated classification of breast cancer from mammographic imagery. Our framework is designed to systematically process and analyze images, extracting salient features and performing classification with a focus on computational efficiency and effective pattern recognition. This approach integrates a Convolutional Neural Network (CNN) for its powerful spatial feature learning capabilities with a Gated Recurrent

Unit (GRU) network, chosen for its proficiency in modeling sequential data with reduced complexity compared to other recurrent architectures. The pipeline is meticulously structured into four distinct phases: (1) mammogram image conditioning and dataset preparation, (2) hierarchical feature derivation using a pre-trained CNN, (3) feature space refinement via dimensionality reduction, and (4) optimized classification using a Bidirectional GRU network whose hyperparameters are tuned by the Grey Wolf Optimizer. Figure 1 (not provided here, but you would include it) illustrates the sequential workflow of our proposed system.

3.1. Mammogram Image Conditioning and Dataset Preparation

The initial phase is dedicated to standardizing the input mammography images to enhance the quality of data fed into subsequent deep learning models, directly impacting their learning capacity and final performance. Mammograms inherently possess characteristics that can impede model training, notably poor contrast arising from dense, overlapping breast tissues and the presence of extensive zero-intensity pixel regions (background areas), which provide no gradient information during backpropagation. To address the challenge of poor visibility in crucial diagnostic regions, the contrast-limited adaptive histogram equalization (CLAHE) is employed. This technique sharpens local details by performing histogram equalization on contextual regions, thereby improving the discernibility of subtle lesions and calcifications, while the clipping mechanism prevents the undue amplification of noise often associated with global histogram equalization.

Following contrast enhancement, we tackle the issue of zero-valued pixels. These pixels can lead to vanishing gradients, effectively stalling the learning process. To counteract this, a pixel intensity adjustment is performed, where zero-valued pixels are systematically replaced with a minimal, non-zero constant. This is formalized as:

$$I'_{\text{img}}(x,y) = \begin{cases} I_{\text{clahe}}(x,y) + o', & \text{if } I_{\text{clahe}}(x,y) = 0, \\ I_{\text{clahe}}(x,y), & \text{otherwise,} \end{cases}$$

where $I(x,y)$ is the pixel intensity at coordinates (x,y) of the CLAHE-processed image, and ϵ is a small constant, empirically set to 1. This adjustment ensures that all pixels contribute, at least minimally, to the gradient computation, fostering more stable and effective weight updates. To further bolster the model's ability to generalize from the available training data and to enhance its robustness against variations typically encountered in clinical settings, we incorporate a suite of data augmentation techniques. These transformations are applied on-the-fly during the training process. The augmentation strategies include random rotations within a range of $\pm 10^\circ$ and horizontal mirroring (flips). These operations synthetically expand the diversity of the training set by simulating minor differences in patient positioning, breast compression, and imaging equipment calibration, thereby reducing the risk of the model overfitting to specific characteristics of the original training samples.

Moreover, a cropping approach is also adopted to reduce the black area in the image, zooming in just on the breast tissues. This is especially helpful for further resizing of images to be fed to the ResNet-50 network. The details of this cropping approach are presented in the Simulation Results section. Finally, the dataset is divided using a stratified sampling method into training (80%) and testing (20%) sets, ensuring that the class distribution is maintained across both splits, which is crucial for unbiased model evaluation. Finally, all images are resized to match the input requirements of the subsequent feature extraction network (ResNet-50) and normalized based on its ImageNet pretraining statistics.

3.2. Hierarchical Feature Derivation using Transfer Learning

For the feature extraction stage, we leverage the power of transfer learning by employing the ResNet-50 architecture, a convolutional neural network renowned for its depth and effectiveness, largely due to its residual connections which facilitate the training of very deep networks by alleviating the vanishing gradient problem. Pre-trained on the large-scale ImageNet dataset, ResNet-50 has learned a rich hierarchy of visual features, many of which are transferable to medical imaging tasks.

Our strategy for adapting ResNet-50 involves a cautious fine-tuning process. The majority of the network's earlier layers, responsible for learning generic features like edges and textures, are kept frozen

to retain their robust, pre-learned weights. We unfreeze only the terminal three convolutional blocks and the subsequent layers, allowing them to adapt specifically to the nuanced patterns found in mammograms, such as mass morphologies and calcification clusters. The original classification head of ResNet-50 is supplanted by a custom dense layer consisting of 2000 neurons with ReLU activation, designed to capture high-level abstractions from the preceding convolutional features. This can be represented as:

$$h = ReLU(W_h c + b_h), \tag{3.1}$$

where c is the feature map output from the last unfrozen ResNet-50 block, W_h and b_h are the weights and biases of our custom dense layer, and h is the resulting feature vector. The fine-tuning process is driven by the Adam optimizer. To balance knowledge retention with specialization, a differential learning rate scheme is adopted. In this regard, a conservative rate of 1×10^{-5} is applied to the initial feature extraction layers, while a more aggressive rate of 1×10^{-4} is used for the unfrozen terminal layers. Training is conducted for 100 epochs with a batch size of 64, and the training data is shuffled at the beginning of each epoch to prevent the model from learning any spurious sequential order. The feature vectors, denoted by F_{train} and F_{test}, are then extracted from the custom dense layer added to the network for subsequent processing.

3.3. Feature Space Refinement via Dimensionality Reduction

The feature vectors obtained from the ResNet-50 are high-dimensional (2000 features according to our design), which can lead to increased computational burden for the subsequent classifier and a heightened risk of overfitting. To address this, we implement linear discriminant analysis (LDA) for dimensionality reduction. LDA is a supervised technique chosen for its ability to project features onto a lower-dimensional space while maximizing the separation between classes, making it particularly suitable for classification tasks.

The LDA transformation seeks to find a projection matrix P that reshapes the feature space. This matrix is derived by optimizing the ratio of between-class scatter to within-class scatter using the training features F_{train} and their corresponding labels L_{train}. Eigenvectors corresponding to the largest eigenvalues of the scatter matrix computation define the new, lower-dimensional space. We set a reduction target to retain 20% of the original feature dimensions, a coefficient determined through preliminary experimentation to balance information preservation with computational efficiency. The transformation is applied as:

$$F'_{\text{train}} = F_{\text{train}} P_{\text{reduced}}, \tag{3.2}$$

$$F'_{\text{test}} = F_{\text{test}} P_{\text{reduced}}, \tag{3.3}$$

where $P_{reduced}$ contains the selected eigenvectors. This step not only makes the subsequent classification more computationally tractable but also aims to enhance model generalization by filtering out less discriminative feature components.

3.4. Optimized Classification using Bidirectional GRU

For the final classification task, we employ a Bidirectional Gated Recurrent Unit (BiGRU) network. GRUs are a type of recurrent neural network with gating mechanisms that allow them to capture dependencies in sequential data more effectively than simple RNNs, while being computationally more efficient than LSTMs due to their simpler architecture (having two gates instead of three). The bidirectional nature allows the network to process the input feature sequences in both forward and backward directions, capturing contextual information from past and future elements within the sequence of features.

The performance of a BiGRU network is heavily dependent on its architectural hyperparameters. To find an optimal configuration, we utilize the Grey Wolf Optimizer (GWO), a meta-heuristic algorithm inspired by the hunting behavior and social hierarchy of grey wolves. GWO was selected for its reported strong global search capabilities and convergence speed in various optimization problems. We aim to optimize four critical hyperparameters: the number of hidden units in each GRU layer ($N_{units} \in [20, 200]$), the initial learning rate ($\eta \in [1 \times 10^{-5}, 1 \times 10^{-2}]$), the learning rate decay factor ($\gamma_{decay} \in [0.55, 0.99]$), and the mini-batch size ($S_{batch} \in [32, 256]$). The GWO is configured with a population of 10 wolves and

run for 50 iterations. The fitness function guiding the GWO is the classification accuracy on a dedicated validation set (comprising 20% of the training data):

$$Maximize : Accuracy(\Psi) = \frac{1}{\mid D_{val} \mid} \sum_{(x_i, y_i) \in D_{val}} \mathbf{I}[g_\Psi(x_i) = y_i], \tag{3.4}$$

where $\Psi = [N_{units}, \eta, \gamma_{decay}, S_{batch}]$ is the vector of hyperparameters, g_Ψ is the BiGRU classifier parameterized by Ψ, (x_i, y_i) is the validation dataset, and $\mathbf{I}(.)$ is the indicator function. While GWO is a powerful optimizer, its effectiveness can be sensitive to its own parameter settings (population size, iterations), and the search landscape for deep learning hyperparameters remains complex. The optimized

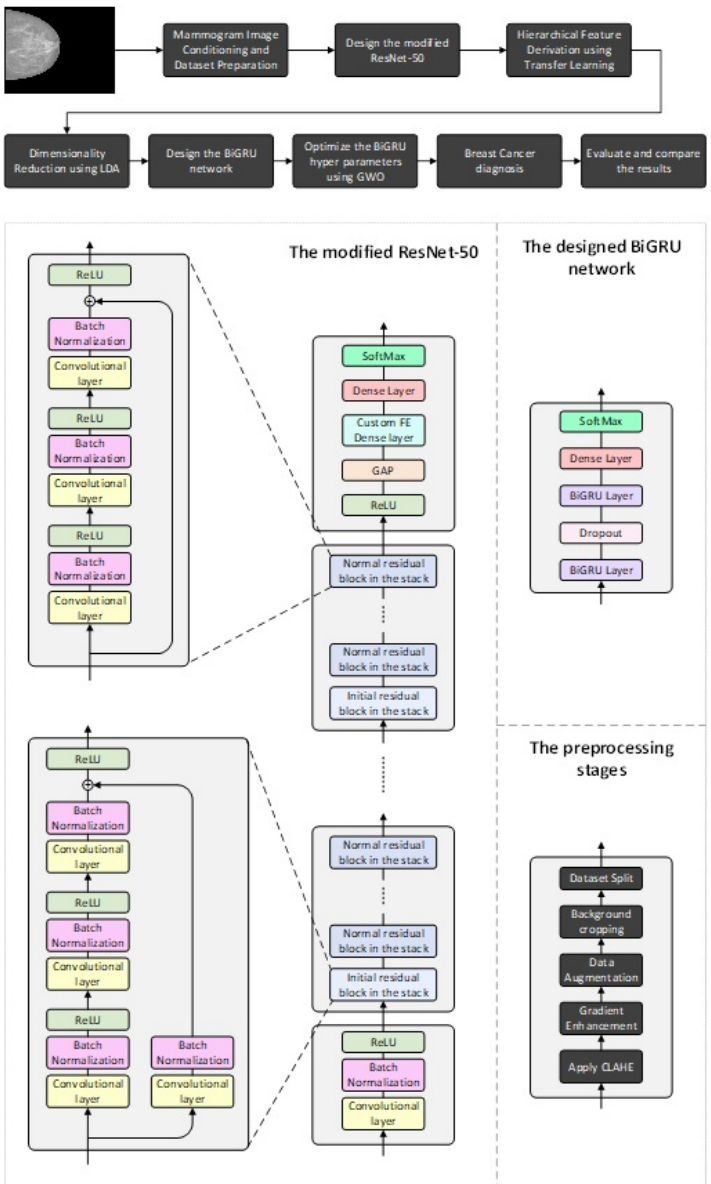

Figure 1: Block diagram of the proposed method.

BiGRU architecture comprises a first BiGRU layer with N_{units} units, a dropout layer with a fixed rate of 0.25 (to mitigate overfitting, a persistent challenge in models with many parameters), a second BiGRU

layer also with N_{units} units, followed by a fully connected layer that maps the BiGRU outputs to the number of classes, and a softmax activation function for probabilistic output and final classification of data. The network is trained using the Adam optimizer with the hyperparameters identified by GWO, for a maximum of 200 epochs, with early stopping implemented based on validation accuracy to prevent overfitting and select the best-performing model. The model's ultimate performance is then assessed on the independent test set. This comprehensive approach, leveraging refined features and optimized recurrent neural networks, aims to provide an effective solution for mammography-based breast cancer diagnosis. The block diagram of the proposed method has been illustrated in Figure 1.

4. Dataset

This study utilized the INbreast dataset, a publicly available and high-quality full-field digital mammography (FFDM) database widely used in breast cancer research. The dataset was selected due to its superior resolution, detailed annotations, and comprehensive representation of real-world clinical scenarios, making it suitable for developing and validating deep learning models for breast cancer detection and classification. The INBreast dataset was created by the Breast Research Group of the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), and is composed of 115 cases comprising 410 mammography images in DICOM format. Each case includes both mediolateral oblique (MLO) and craniocaudal (CC) views of the left and right breasts, covering a broad range of diagnostic categories. The images are annotated by expert radiologists with BI-RADS assessment scores (ranging from 1 to 5), as well as lesion contours and mass classification (such as calcifications, asymmetries, and architectural distortions).

For the purpose of this study, the dataset was preprocessed to focus specifically on the binary classification of normal versus breast cancer cases. Images were converted from DICOM to a standard image format, resized for computational efficiency, and normalized to enhance contrast. Labeling was guided by the BI-RADS scores and lesion annotations, where BI-RADS categories 1, 2, and 3 were considered normal, and categories 4 and 5 were considered indicative of breast cancer.

5. Evaluation Metrics

To quantitatively evaluate the performance of the proposed breast cancer classification framework, a set of standard evaluation metrics was employed, including Accuracy, Precision, Recall, and F1-score. These metrics are widely used in medical image analysis to assess the diagnostic capability of machine learning models, particularly in binary classification tasks such as distinguishing between normal and breast cancer cases. The evaluation relies on four key quantities derived from the confusion matrix: True Negatives (TN) correspond to correctly identified normal cases, False Positives (FP) refer to normal cases misclassified as breast cancer, and False Negatives (FN) are breast cancer cases incorrectly labeled as normal.

Accuracy: The Accuracy metric evaluates the overall correctness of the model by measuring the ratio of correctly classified instances to the total number of instances:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{5.1}$$

Precision: Precision, also referred to as the positive predictive value, quantifies the proportion of correctly predicted Breast Cancer cases among all cases classified as Breast Cancer:

$$Precision = \frac{TP}{TP + FP}. \tag{5.2}$$

Recall: Recall, or sensitivity, reflects the model's ability to correctly identify actual Breast Cancer cases, representing the proportion of true Breast Cancer cases that were correctly classified:

$$Recall = \frac{TP}{TP + FN}. \tag{5.3}$$

F1-score: The F1-score is the harmonic mean of Precision and Recall, providing a single metric that balances both false positives and false negatives, which is particularly important in medical diagnostics where both types of errors carry significant consequences:

$$F1\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{5.4}$$

Together, these metrics offer a comprehensive assessment of the proposed model's performance, ensuring not only its overall accuracy but also its reliability and robustness in correctly identifying Breast Cancer cases while minimizing diagnostic errors.

6. Simulation Results

This section presents and analyzes the simulation results obtained from the implementation of the proposed hybrid deep learning framework for breast cancer diagnosis. The evaluation is structured to provide a comprehensive understanding of the model's performance, beginning with a stage-wise assessment of key pipeline components, followed by an in-depth analysis of the overall diagnostic capabilities. The initial part examines the outcomes of critical preprocessing steps, such as the tuning of Contrast Limited Adaptive Histogram Equalization (CLAHE) and the effectiveness of the background cropping technique, the impact of Linear Discriminant Analysis (LDA) on feature separability, and the convergence behavior of the Grey Wolf Optimizer (GWO) during the hyperparameter optimization of the Bidirectional GRU (BiGRU) network. Subsequently, the overall diagnostic performance of the complete framework on the INbreast dataset is rigorously evaluated using standard metrics including accuracy, precision, recall, F1-score, as well as detailed visualizations through confusion matrices and Receiver Operating Characteristic (ROC) curves. All experiments were conducted on a system configured with an Intel® Core™ i7-12700H processor, 32GB of RAM, and an NVIDIA GeForce RTX 3070 Laptop GPU with 8GB of dedicated memory, utilizing MATLAB 2024b for implementation.

6.1. Pipeline Component Performance

This subsection details the experimental results and analyses pertaining to the key stages of our proposed mammography image processing and feature engineering pipeline, leading up to the final classification. We present the outcomes of critical preprocessing steps, including the empirical tuning of Contrast Limited Adaptive Histogram Equalization (CLAHE) and the application of our background cropping technique. Furthermore, we evaluate the efficacy of Linear Discriminant Analysis (LDA) in transforming the feature space for enhanced class separability. Finally, this section covers the hyperparameter optimization process for the Bidirectional Gated Recurrent Unit (BiGRU) classifier, showcasing the convergence behavior of the Grey Wolf Optimizer (GWO) and the resulting optimal hyperparameters. The systematic evaluation of these components is crucial for validating their individual contributions and ensuring that each stage effectively prepares the data for robust and accurate breast cancer diagnosis by the subsequent classification model.

Figure 2 presents a comparative visual analysis of various Contrast Limited Adaptive Histogram Equalization (CLAHE) settings applied to a representative mammogram from the dataset. This investigation was conducted to empirically determine the optimal CLAHE parameters for enhancing image contrast, crucial for improving the visibility of subtle pathological indicators. The figure displays the outcomes of applying three distinct histogram distributions—uniform, Rayleigh, and exponential—each tested with clip limits of 0.01, 0.02, and 0.03. Visual inspection reveals noticeable differences in image quality and feature discernibility across these configurations. For instance, the exponential distribution (right column) tends to produce overly bright regions and potentially harsh contrast shifts, which could obscure finer details. Conversely, the uniform distribution (left column) provides a more moderate enhancement. After careful evaluation, the Rayleigh distribution with a clip limit of 0.02 (center image in the second row) was selected for subsequent preprocessing steps. This specific combination was deemed to offer the most balanced enhancement, effectively improving the definition of internal breast structures and tissue variations while preserving smooth gradients in fatty areas. It provides a discernible improvement in clarity over the Rayleigh 0.01 setting without introducing the potential over-enhancement

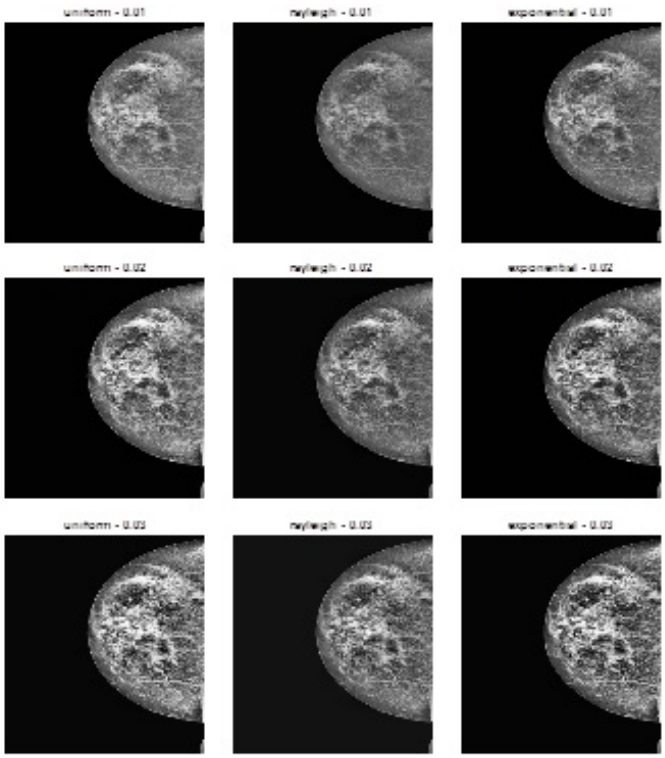

Figure 2: Comparing different histogram distributions and clip limit adjustments for CLAHE.

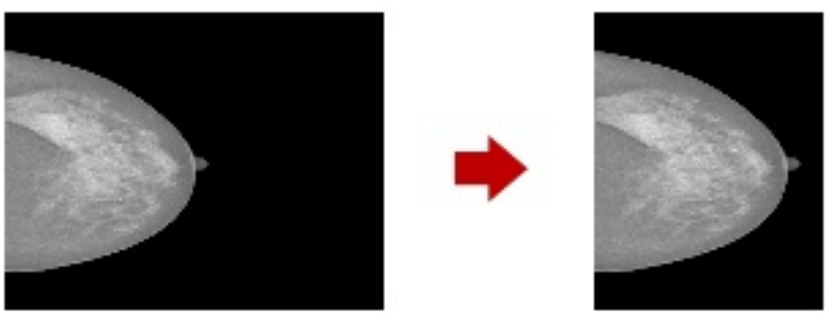

Figure 3: The proposed cropping results for one mammography sample from the dataset.

or artifact introduction observed at the higher clip limit of 0.03, thereby ensuring that diagnostically relevant information is made more accessible for the feature extraction phase.

Figure 3 illustrates the outcome of the proposed background cropping stage, a critical step in our preprocessing pipeline designed to isolate the relevant breast tissue from non-informative background regions. The figure presents a side-by-side comparison of a sample mammogram before (left) and after (right) the application of this technique, visually confirming the effective removal of the extraneous black background. This process involves an initial analysis to identify the side of the image predominantly containing breast tissue by comparing the cumulative pixel intensities, thereby accommodating both left and right-sided mammograms. Subsequently, a predetermined horizontal crop is applied to standardize

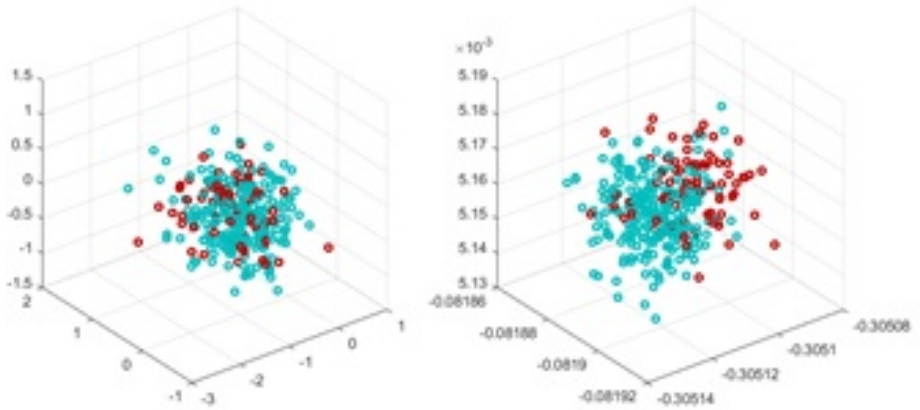

Figure 4: The class discriminability using the first three extracted features before and after applying LDA.

the image width while ensuring no significant tissue information is lost. The primary advantage of this targeted cropping is particularly evident when preparing images for the fixed input dimensions (e.g., 224×224 pixels) of the ResNet architecture. By eliminating large swathes of background, this step ensures that the subsequent resizing operation preserves a higher resolution of the actual breast tissue, concentrating the feature extraction capabilities of the CNN on diagnostically pertinent areas and optimizing the quality of input data for the deep learning model.

Figure 4 provides a visual assessment of the Linear Discriminant Analysis (LDA) stage by displaying 3D scatter plots of the dataset's first three feature dimensions, both prior to and following LDA application. The left panel illustrates the distribution of the two classes (represented by distinct red and cyan markers) in the original feature space derived from ResNet-50, where a considerable degree of overlap between the classes is evident, indicating limited inherent separability based solely on these initial features. In contrast, the right panel depicts the same data points projected onto the first three components derived from LDA. A noticeable improvement in class discriminability is observed in this transformed space; the clusters of red (normal class) and cyan (cancerous class) points appear more distinctly separated, with reduced inter-class mingling compared to the original feature representation. This enhanced separation underscores LDA's effectiveness in identifying a subspace that maximizes inter-class variance while minimizing intra-class variance. While these plots offer a visualization based on only three dimensions, and complete class separation typically relies on the full set of retained LDA components, the improved visual distinction strongly suggests that LDA successfully enhances the feature representation for the subsequent GRU-based classification, thereby contributing to a more robust diagnostic model.

The efficacy of the GWO in fine-tuning the Bidirectional GRU network's hyperparameters is demonstrated in Figure 5 and Table 1. Figure 5 illustrates the convergence curve of the GWO algorithm over 40 iterations, plotting the minimization of the objective loss function. The curve exhibits a characteristic stepwise reduction in loss, with significant improvements achieved in the initial iterations—most notably around the 4th and 14th iterations—followed by periods of stability. This behavior indicates that the GWO effectively explored the search space and rapidly converged towards an optimal set of hyperparameter values. By approximately the 15th iteration, the algorithm had identified a solution yielding a low loss value of approximately 0.013, which remained largely unchanged for the subsequent iterations, signifying stable convergence. Table 1 enumerates the specific hyperparameter values identified by this GWO process as optimal for the BiGRU classifier. These include 86 hidden units per GRU layer, an initial learning rate of 0.0044, a learning rate decay factor (drop factor) of 0.97, and a mini-batch size of 33. This systematic optimization is paramount, as it tailors the BiGRU architecture and training dynam-

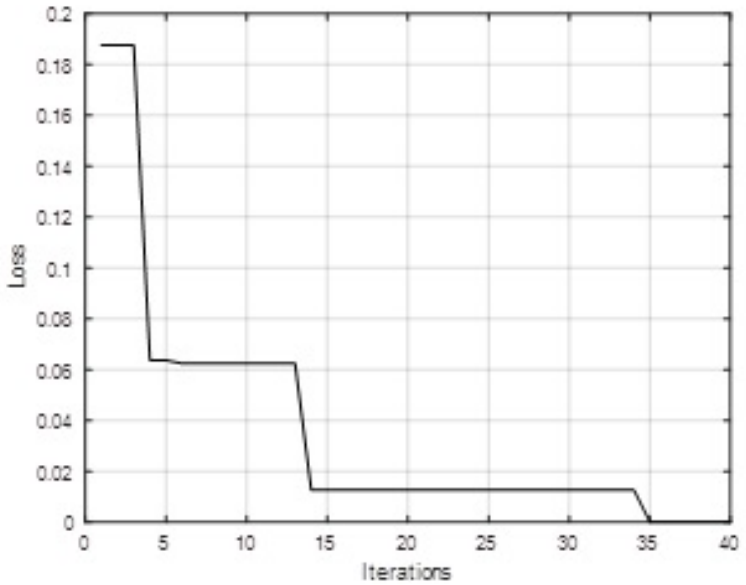

Figure 5: The convergence curve of the GWO algorithm.

ics to the specific characteristics of the mammography feature data, thereby maximizing the classifier's potential for accurate breast cancer diagnosis in the subsequent evaluation phase.

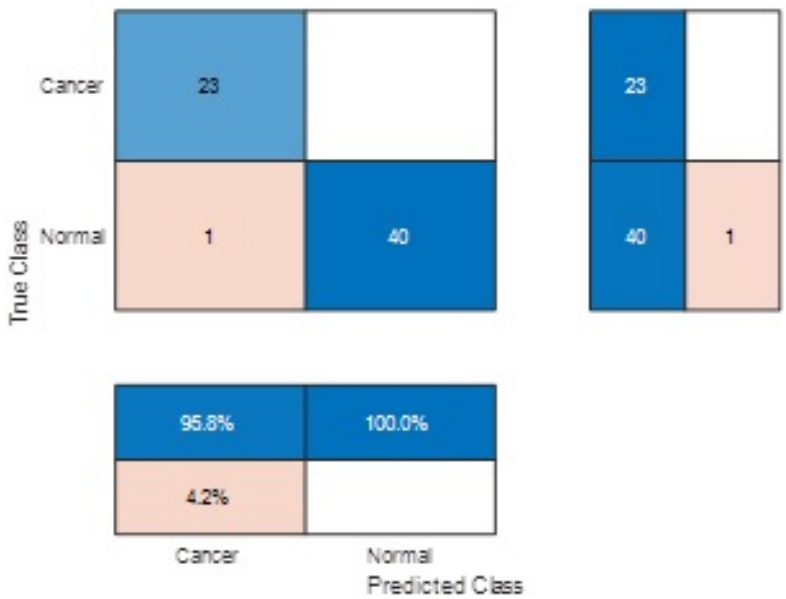

Figure 6: The confusion matrix of the proposed method evaluation.

Table 1: The obtained optimal hyperparameters of the BiGRU network.

Parameter	Value
Number of Hidden Units	86
Learning Rate	0.0044
Learning Rate Drop Factor	0.97
Mini-Batch Size	33

6.2. Breast Cancer Diagnosis Performance Evaluation

Having established the performance of individual components and the optimized configuration of the BiGRU classifier in the preceding subsection, we now turn to the comprehensive evaluation of the end-to-end breast cancer diagnosis system. This section quantifies the diagnostic capabilities of the fully developed hybrid deep learning model when applied to the unseen independent test set. The model's ability to accurately distinguish between benign and malignant cases is assessed through several key performance indicators. Specifically, we will present a detailed breakdown of classification metrics—namely, accuracy, precision, recall (sensitivity), and F1-score—summarized in Table 2. Furthermore, a visual analysis of the classification results will be provided through the confusion matrix (Figure 6), offering insights into the specific types of correct and incorrect predictions. Finally, the model's discriminative power across various diagnostic thresholds will be illustrated by its Receiver Operating Characteristic (ROC) curve (Figure 7), from which the Area Under the Curve (AUC) can also be determined. Together, these results will provide a robust measure of the proposed system's diagnostic performance.

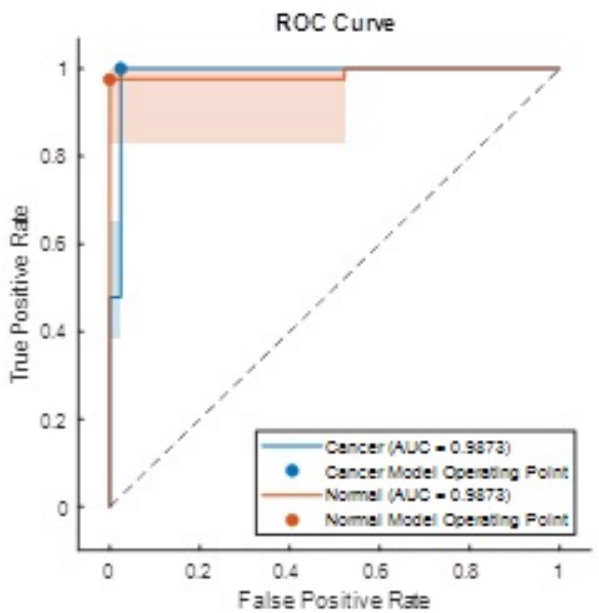

Figure 7: The ROC curve of the proposed method evaluation.

Figure 6 presents the confusion matrix detailing the proposed model's classification performance on the independent test set. This provides a clear view of the specific outcomes for each class. According to the matrix, the model correctly identified all 23 actual 'Cancer' cases as 'Cancer' (True Positives = 23), and critically, there were no 'Cancer' cases misclassified as 'Normal' (False Negatives = 0). This achievement of zero false negatives is of paramount clinical importance, as it indicates the model did not miss any actual instances of cancer within this test cohort, thereby avoiding potentially fatal consequences associated with missed diagnoses. On the other side, out of the cases that were actually 'Normal,' 40

Table 2: Performance metrics of the proposed model on the test set.

Metric	Value
Accuracy	98.44%
Precision	97.92%
Recall	98.78%
F1-Score	98.35%

were correctly identified as 'Normal' (True Negatives = 40), while only 1 'Normal' case was misclassified as 'Cancer' (False Positives = 1). While a false positive may lead to unnecessary follow-up examinations and patient anxiety, it is clinically preferable to a false negative. These raw counts (TP=23, FN=0, FP=1, TN=40) demonstrate a strong diagnostic capability, particularly highlighting the model's perfect sensitivity in detecting cancer cases within this specific test set.

Figure 7 displays the Receiver Operating Characteristic (ROC) curve for the proposed breast cancer diagnostic model, illustrating its discriminative ability across the spectrum of decision thresholds. The curve plots the True Positive Rate (Sensitivity) against the False Positive Rate (1-Specificity). The Area Under the Curve (AUC) for detecting the 'Cancer' class is an excellent 0.9873, approaching the ideal value of 1.0. This high AUC value signifies the model's outstanding capacity to distinguish between malignant and benign cases effectively. Visually, the ROC curve rises sharply towards the top-left corner of the plot, indicating that the model achieves high sensitivity while maintaining a low false positive rate across various thresholds. The specific operating point chosen for the 'Cancer' class (indicated by the blue dot) further highlights this strong performance, corresponding to a True Positive Rate approaching 1.0 (perfect sensitivity for cancer detection) at a False Positive Rate of approximately 0.024 (Specificity of approximately 97.6%). This specific point aligns well with the results presented in the confusion matrix (FN=0, FP=1). In a nutshell, the ROC analysis provides robust evidence of the model's high diagnostic accuracy and reliability.

6.3. Comparison

To evaluate the effectiveness of the proposed method, we compare its performance with recent state-of-the-art approaches for breast cancer classification using mammographic images. Among the latest contributions, the study by Hameedur Rahman et al. (2023) [1] presents a deep learning framework based on ResNet-50 pretrained on ImageNet, used to classify the INbreast dataset into benign and malignant categories. Their model achieved an accuracy of 93%, demonstrating the capability of deep convolutional networks in extracting discriminative features from mammograms and supporting early detection in clinical settings.

Table 3: Comparative table of the proposed method with other reviewed references.

References	Method	Dataset	Accuracy (%)
[1]	ResNet-50	INbreast	93
[13]	Multi-deep CNNs, SVM, PCA	MIAS, INbreast	MIAS: 97.93, INbreast: 96.646
Proposed method	ResNet-50, GRU, GWO	INbreast	98.44

Another noteworthy study is by Sannasi Chakravarthy et al. (2023) [13], which explored multiple classification strategies using deep learning. Their research included experiments with deep CNNs, SVM classifiers on extracted features, fusion techniques, and PCA, evaluated on both MIAS and INbreast datasets. The best results on the INbreast dataset were obtained using fused deep features, yielding an accuracy of 96.646%, while PCA reduced computational cost without improving accuracy. This study emphasizes the importance of hybrid methods that integrate feature learning and optimization to enhance diagnostic precision.

Building upon these foundations, our proposed method utilizes ResNet-50 for feature extraction, followed by a GRU for classification, whose parameters are optimized using the GWO algorithm. This hybrid approach achieved an accuracy of 98.44% on the INbreast dataset, surpassing both previous

frameworks. The integration of deep spatial feature extraction and optimized sequential classification enables more robust learning and enhances diagnostic accuracy, thereby offering a promising solution for computer-aided diagnosis systems in breast cancer detection. Table 3 provides a comparison of the reviewed studies.

7. Conclusion

This paper addressed the critical challenge of enhancing automated breast cancer diagnosis from mammography images by proposing and evaluating a meticulously designed hybrid deep learning framework. The primary objective was to develop a robust system capable of high diagnostic accuracy. Our methodology integrated several key stages: comprehensive image preprocessing involving optimized CLAHE, a novel background cropping technique, pixel value adjustments, and data augmentation; feature extraction via a fine-tuned ResNet-50 model; dimensionality reduction using Linear Discriminant Analysis (LDA) to refine feature representation; and finally, classification employing a Bidirectional Gated Recurrent Unit (BiGRU) network whose hyperparameters were rigorously optimized using the Grey Wolf Optimizer (GWO). The proposed framework, when evaluated on the publicly available INbreast dataset, demonstrated exceptional performance. It achieved an overall accuracy of 98.44%, a precision of 97.92%, recall (sensitivity) of 98.78%, and an F1-score of 98.35%. Notably, the confusion matrix analysis highlighted the model's strength in correctly identifying all cancer cases in the test set breakdown (False Negatives = 0), a crucial factor for clinical reliability, alongside a high AUC value of 0.9873. These results underscore the efficacy of the synergistic combination of advanced image processing, deep feature learning, and optimized recurrent neural networks. The significance of this work lies in the successful application of a GWO-tuned BiGRU classifier on LDA-refined ResNet-50 features for mammography analysis, showcasing a promising avenue for developing highly accurate Computer-Aided Diagnosis (CAD) tools. The careful pipeline construction, particularly the preprocessing steps and the meta-heuristic optimization, contributed significantly to the final performance. While the results are promising, potential limitations include the evaluation being conducted on a single, albeit challenging, dataset. The computational cost associated with meta-heuristic optimization, such as GWO, could also be a consideration for broader implementation. Future research could focus on validating the model's robustness and generalization capabilities across larger and more diverse multi-center datasets. Further investigations might also explore alternative deep learning architectures, other optimization algorithms, and the incorporation of explainable AI (XAI) techniques to enhance transparency and clinical trust. In conclusion, the hybrid deep learning framework presented in this study offers a highly effective and accurate approach for breast cancer detection in mammograms, holding considerable potential to support radiologists and improve patient outcomes through earlier and more reliable diagnosis.

References

1. Rahman, H., Naik Bukht, T. F., Ahmad, R., Almadhor, A., and Javed, A., *Efficient breast cancer diagnosis from complex mammographic images using deep convolutional neural network*, Comput. Intell. Neurosci. 2023(1), 7717712, (2023).

2. Urabinahatti, S., and Jayadevappa, D., *Breast cancer detection using deep learning technique*, Proc. 2023 Int. Conf. on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), IEEE, pp. 1-5, (2023).

3. Ajantha Devi, V., and Nayyar, A., *Fusion of deep learning and image processing techniques for breast cancer diagnosis*, in Deep Learning for Cancer Diagnosis, Springer, Singapore, pp. 1-25, (2020).

4. Saber, A., Sakr, M., Abo-Seida, O. M., Keshk, A., and Chen, H., *A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique*, IEEE Access 9, 71194-71209, (2021).

5. Balkenende, L., Teuwen, J., and Mann, R. M., *Application of deep learning in breast cancer imaging*, Semin. Nucl. Med. 52(5), 584-596, (2022).

6. Mahoro, E., and Akhloufi, M. A., *Applying deep learning for breast cancer detection in radiology*, Curr. Oncol. 29(11), 8767-8793, (2022).

7. Lin, R. H., Kujabi, B. K., Chuang, C. L., Lin, C. S., and Chiu, C. J., *Application of deep learning to construct breast cancer diagnosis model*, Appl. Sci. 12(4), 1957, (2022).

8. Chorianopoulos, A. M., Daramouskas, I., Perikos, I., Grivokostopoulou, F., and Hatzilygeroudis, I., *Deep learning methods in medical imaging for the recognition of breast cancer*, Proc. 2020 11th Int. Conf. on Information, Intelligence, Systems and Applications (IISA), IEEE, pp. 1-8, (2020).

9. Suguna, R., Sathik Raja, M., and Priya, S., *Breast cancer detection using deep learning*, Int. Res. J. Comput. Sci. 11(4), (2024).

10. Gupta, S., Panwar, A., Yadav, R., Aeri, M., and Manwal, M., *Employing deep learning feature extraction models with learning classifiers to diagnose breast cancer in medical images*, Proc. 2022 IEEE Delhi Section Conf. (DELCON), IEEE, pp. 1-6, (2022).

11. Eldin, S. N., Hamdy, J. K., Adnan, G. T., Hossam, M., Elmasry, N., and Mohammed, A., *Deep learning approach for breast cancer diagnosis from microscopy biopsy images*, Proc. 2021 Int. Mobile, Intelligent, and Ubiquitous Computing Conf. (MIUCC), IEEE, pp. 216-222, (2021).

12. Rashed, E., and El Seoud, M. S. A., *Deep learning approach for breast cancer diagnosis*, Proc. 8th Int. Conf. on Software and Information Engineering, pp. 243-247, (2019).

13. Sannasi Chakravarthy, S. R., Bharanidharan, N., and Rajaguru, H., *Multi-deep CNN based experimentations for early diagnosis of breast cancer*, IETE J. Res. 69(10), 7326-7341, (2023).

14. Escorcia-Gutierrez, J., Mansour, R. F., Beleño, K., Jiménez-Cabas, J., Pérez, M., Madera, N., and Velasquez, K., *Automated deep learning empowered breast cancer diagnosis using biomedical mammogram images*, Comput. Mater. Continua 71(3), (2022).

15. Amrisha, R. R., Monika, M., Yamuna, V., Arun Kumar, S., and Sasikala, S., *Enhanced breast cancer detection in multiple imaging modalities using deep learning*, Proc. 2023 2nd Int. Conf. on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), IEEE, pp. 1-5, (2023).

16. Grigore, M. A., and Neagoe, V. E., *A deep CNN approach using thermal imagery for breast cancer diagnosis*, Proc. 2021 13th Int. Conf. on Electronics, Computers and Artificial Intelligence (ECAI), IEEE, pp. 1-6, (2021).

17. Rajpal, D., Mishra, S., and Kumar, A., *Lesion detection and classification for breast cancer diagnosis based on deep CNNs from digital mammographic data*, Comput. Anal. Deep Learn. Med. Care: Principles, Methods, and Applications, pp. 257-292, (2021).

18. Hamid, M. A., Mondher, H. M., and Ayoub, B., *Deep learning CNNs for breast cancer classification and detection: enhancing diagnostic accuracy in medical practice*, Proc. 2024 2nd Int. Conf. on Electrical Engineering and Automatic Control (ICEEAC), IEEE, pp. 1-6, (2024).

19. Kumar, M. S., and Kumar, K. P., *Breast cancer tissue identification using deep learning in mammogram images*, Proc. 2023 Int. Conf. on Advances in Computing, Communication and Applied Informatics (ACCAI), IEEE, pp. 1-6, (2023).

20. Zakareya, S., Izadkhah, H., and Karimpour, J., *A new deep-learning-based model for breast cancer diagnosis from medical images*, Diagnostics 13(11), 1944, (2023).

21. Liu, W., Li, H., Hua, C., and Zhao, L., *Classifications of breast cancer images by deep learning*, medRxiv preprint, 2020-06, (2020).

22. Zheng, J., Lin, D., Gao, Z., Wang, S., He, M., and Fan, J., *Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis*, IEEE Access 8, 96946-96954, (2020).

23. Rajput, G., Agrawal, S., Biyani, K., and Vishvakarma, S. K., *Early breast cancer diagnosis using cogent activation function-based deep learning implementation on screened mammograms*, Int. J. Imaging Syst. Technol. 32(4), 1101-1118, (2022).

24. Ragab, D. A., Attallah, O., Sharkas, M., Ren, J., and Marshall, S., *A framework for breast cancer classification using multi-DCNNs*, Comput. Biol. Med. 131, 104245, (2021).

25. Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y., *Learning phrase representations using RNN encoder-decoder for statistical machine translation*, arXiv preprint arXiv:1406.1078, (2014).

26. Chung, J., Gulcehre, C., Cho, K., and Bengio, Y., *Empirical evaluation of gated recurrent neural networks on sequence modeling*, arXiv preprint arXiv:1412.3555, (2014).

27. Mirjalili, S., Mirjalili, S. M., and Lewis, A., *Grey wolf optimizer*, Adv. Eng. Softw. 69, 46-61, (2014).

28. He, K., Zhang, X., Ren, S., and Sun, J., *Deep residual learning for image recognition*, Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, (2016).

Mustafa Raad Ali,
Institute of Energy Infrastructure (IEI),
Department of Civil Engineering, College of Engineering,
Universiti Tenaga Nasional (UNITEN),
Putrajaya Campus, Jalan IKRAM-UNITEN,
43000 Kajang, Selangor, Malaysia
E-mail address: m99448910@gmail.com

and

Normy Norfiza Abdul Razak,
Institute of Energy Infrastructure (IEI),
Department of Civil Engineering, College of Engineering,
Universiti Tenaga Nasional (UNITEN),
Putrajaya Campus, Jalan IKRAM-UNITEN,
43000 Kajang, Selangor, Malaysia
E-mail address: normy@uniten.edu.my

and

Rashid Jan[1,2,3,*],
[1]*Department of Mathematics,*
Saveetha School of Engineering (SIMATS),
Thandalam 600124, Chennai, Tamil Nadu, India.
[2]*Department of Mathematics,*
Khazar University, AZ1096, Baku Azerbaijan.
[3]*Institute of Energy Infrastructure (IEI),*
Department of Civil Engineering, College of Engineering,
Universiti Tenaga Nasional (UNITEN),
Putrajaya Campus, Jalan IKRAM-UNITEN,
43000 Kajang, Selangor, Malaysia
E-mail address: Corresponding author: rashid_ash2000@yahoo.com