TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214
DOI: 10.26599/TST.2026.9010016

WCMA-Net: Enhancing Mammographic Cancer Diagnosis Using Wavelet-Driven Channel-Spatial Mamba Attention

Khalil ur Rehman, Yibin Tian*, Li Jianqiang, Anaa Yasin, Weiwei Jiang, Sushil Kumar Singh, Mohammed Aloraini, and Inam Ullah

Abstract: Microcalcification Clusters (MCCs) are radiological markers of breast cancer. However, accurate diagnosis of these minute calcium deposits remains a significant challenge due to their spatial obscurity, particularly in dense breast tissues. Reducing false positives is crucial for detecting MCCs in mammograms, as traditional methods often yield erroneous cases. We present WCMA-Net (Wavelet-based Channel-wise Mamba Attention), an interpretable deep learning framework that integrates Discrete Wavelet Transform (DWT), channel and spatial attention, and a state-space Mamba attention mechanism to improve MCC detection. It isolates high-frequency diagnostic cues through wavelet decomposition and enhances feature discriminability via dual attention mechanisms. The Mamba attention further captures long-range dependencies and temporal-spatial dynamics, facilitating precise classification. We incorporate Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations to explain model decisions and highlight diagnostically relevant regions. Experiments on two benchmark datasets demonstrate state-of-the-art performance, achieving AUCs of 0.99 and 0.96, with high sensitivity and specificity. Compared to Swin Transformer and Vision-Mamba models, WCMA-Net delivers superior accuracy with lower computing cost Giga Floating Point Operations Per Second (0.43 GFLOPs), suitable for real-time applications. The results establish WCMA-Net as a practical, interpretable system for breast cancer screening.

Keywords: Breast Cancer, Microcalcification Cluster (MCC), Wavelet, Neural Network, Mamba, Pattern Recognition

Just Accepted

1 Introduction

Breast cancer is the leading cause of mortality among women worldwide, with particularly high death rates in Asia and Africa compared to Europe [1]. Mammogram screening is a well-established method for the early detection of breast cancer in both men and women. Among critical indicators, breast Microcalcification Clusters (MCCs) are small calcium deposits that are highly indicative of breast cancer [2]. However, due to their small size and irregular distribution, particularly in dense breast tissues, detecting MCCs remains challenging. This complicates the diagnosis, especially for radiologists who rely on a manual mammogram examination [3]. MCCs are detectable on mammograms due to the higher density of breast tissues compared to surrounding structures [4]. These tiny calcium accumulations, typically less than 1 mm in diameter and corresponding to approximately 10-20 pixels on digital mammograms, are often obscured by high-frequency noise, which complicates their detection and presents a significant challenge for radiologists [5].

Identifying MCC lesions in dense breast tissue is a considerable challenge due to their poor contrast, uneven

• Khalil ur Rehman, Yibin Tian and Inam Ullah are with the College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China. Email: khalilur@szu.edu.cn, ybtian@szu.edu.cn, Inam160685@szu.edu.cn.
• Li Jianqiang and Anaa Yasin are with the College of Software Engineering, Beijing University of Technology, Beijing 100124, China. Email: lijianqiang@bjut.edu.cn, yasinanaa@emails.bjut.edu.cn.
• Weiwei Jiang is with School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China. Email: jww@bupt.edu.cn.
• Sushil Kumar Singh is with the Department of Computer Engineering, Marwadi University, Rajkot, India. Email: sushilkumar.singh@marwadieducation.edu.in.
• Mohammed Aloraini is with the Department of Electrical Engineering, Qassim University, Buraydah 52571, Saudi Arabia. Email: MO.ALORAINI@qu.edu.sa.
* To whom correspondence should be addressed.
Manuscript received: 2025-07-01; revised: 2025-11-05; accepted: 2026-01-16

morphology, and dispersed distribution. Current Computer-Assisted Diagnostic (CAD) systems, which aim to assist radiologists, often generate numerous False Positives (FP) in their efforts to enhance sensitivity [6][7]. The key factors contributing to these errors include imaging noise, artifacts, and variability between patients [8]. Reducing noise while maintaining the integrity of essential features is a crucial but yet unsolved problem. The primary objective of CAD systems is to assist radiologists by identifying and localizing potentially abnormal regions, such as masses or MCCs, within the breast [9][10]. Morphology plays a crucial role in the diagnostic process and encompasses physical attributes such as the shape, size, brightness, and roughness of the calcifications [11]. Conventionally, the process of detecting breast cancer using a digital mammogram involves two distinct processes. The first stage focuses on identifying MCCs by segmenting the Regions Of Interest (ROIs) of the mammogram. The second stage involves classifying these clustered ROIs as malignant or benign.

The manual extraction of features and their classifications using conventional methods is a burden and diminishes the sensitivity of cancer diagnosis. Numerous efforts have been made, including the shape-based method by Mabrouk et al. [12]) and the global detection and feature extraction model by Wang et al. [13], fail to accurately segment and classify MCCs due to issues such as inter-cluster variability, noise, and localization challenges. These deficiencies decrease the True Positive (TP) rate and increase the FP rate. Consequently, a resilient automated system that can overcome these hurdles is desired. Hadjidi et al. [14] proposed a watershed morphology method for segmenting mammograms, effectively detecting edges but struggling with noise reduction. Chakravarthy et al. [15] utilized a firefly algorithm with extreme learning machines for MCC detection, although it faced overfitting. Basile et al. [16] introduced a CAD system that combines the Hough transform and clustering algorithms, which improved detection but increased computational complexity. Long et al. [17] employed a contrast-enhanced mammogram to enhance the precision of the Breast Imaging Reporting and Data System (BI-RADS) for calcification lesions. Touil et al. [18] applied mathematical morphology to improve MCC detection by addressing low contrast in digitized mammograms.

Hao et al. [19] used a data-driven deep learning approach to identify cancerous regions on mammograms, while Sun et al. [20] utilized a Deep Neural Network (DNN) to analyze multidimensional mammographic data to predict breast cancer. Shu et al. [21] proposed a region-based DNN for the segmentation and classification of mammograms to predict breast cancer. Abdelsamea et al. [22] introduced a DNN-based strategy using a mapping technique to classify malignant and benign mammograms according to neuron topology. Adiga et al. [23] developed artificial neural networks that use self-organizing maps to minimize the distance between neurons, enhancing the performance of model training. Suresh et al. [24] designed a diagnostic system that applies deep learning-based pattern recognition to differentiate between normal and abnormal mammograms. Fan et al. [25] used a faster-region pooling architecture of DNNs to accurately categorize digital mammograms to predict breast cancer. Yurdusev et al. [26] emphasizes the detectability of MCCs in mammograms by enhancing the difference filter with a Yolov4 model. Ghug et al. [27] used the multi-scale DenseNet for mammogram segmentation. Chouhan et al. [28] employed an emotional learning-based DNN for the detection of breast cancer. Finally, Qian et al. [29] utilized a profound extreme learning technique with a DNN to diagnose breast cancer from mammograms. However, segmenting suspicious regions and classifying them remain major challenges in mammography analysis, even with these advanced deep learning techniques.

In this study, we address the research gap discussed above by introducing a novel automated de-noising method for detecting MCCs in digital mammograms, aimed at improving breast cancer prediction. The primary challenge lies in the accurate detection of MCCS in dense breast tissues, where the inter-cluster spacing, noise suppression, and contrast enhancement between individual MCCs, and precise localization significantly affect classification performance, thereby lowering the TP rate. In addition, traditional DNNs struggle to effectively classify these MCCs into malignant or benign categories. The proposed approach overcomes the limitations of manual mammogram analysis by radiologists and traditional DNNs. It employs a Gaussian notch filter to suppress noise and enhance background contrast, effectively amplifying these MCCs for better visibility. This preprocessing step localizes the signal characteristics in the transformed domain, enabling precise feature extraction. The Discrete Wavelet Transform (DWT) is then utilized to decompose the mammogram into multiple frequency subbands, isolating high-frequency components associated with MCCs. This hierarchical decomposition facilitates robust feature extraction, emphasizing the fine details critical for accurate classification. The wavelet features are fed into a state-space Mamba attention network, which integrates channel and spatial attention mechanisms to refine and amplify the diagnostic features [30]. The channel attention module focuses on enhancing the most relevant channels in the feature map. Using global average pooling and max pooling, the network computes channel-wise importance scores, which are refined through fully connected layers. Spatial attention complements channel attention by focusing on spatially important regions in the feature map. It aggregates spatial context using average and max pooling across the channel dimension and applies a convolutional layer to produce spatial attention weights. Finally, the Mamba network combines these attention-enhanced features and models them within a state-space framework. This structure captures complex interdependencies between wavelet-derived features, channel relationships, and spatial context, enhancing the network's ability to distinguish benign from malignant MCCs. The classification layer further processes the refined features, outputting probabilities for diagnostic categories. The main contributions are as follows.

- A Wavelet-based Channel-wise Mamba Attention Network (WCMA-Net) is proposed for MCC detection and the prediction of breast cancer. First, Gaussian notch filtering

Under Review

and DWT are applied to suppress noise and enhance the high-frequency components of mammograms. A novel channel-wise attention mechanism is employed to emphasize relevant ROIs, followed by a spatial attention mechanism to refine the representation of spatial features. Subsequently, the enhanced features are processed through a Mamba network. Finally, enriched features are classified into benign and malignant categories through fully connected layers.

- Experimental evaluations conducted on the private PINUM and public CBIS-DDSM breast cancer datasets demonstrate the superiority of WCMA-Net, achieving state-of-the-art performance in terms of precision, accuracy, sensitivity, specificity, and Area Under the Curve (AUC).

2 Materials and Method

2.1 Data Preprocessing and Augmentation

Two datasets are utilized for the study, the private Punjab Institute of Nuclear Medicine (PINUM) [31], and the public Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [32]. The PINUM dataset was collected with the approval from the Diagnostic Imaging Nuclear Medicine and Radiology Research and Development Committee, consisting of 288 mammograms. Patients between 32 and 73 years with an average age of 48.5 years contributed a total of 588 mammograms, including 162 malignant and 426 benign cases, captured in both Mediolateral Oblique (MLO) and Craniocaudal (CC) views. The data collection and labeling team members are experienced radiologists. The CBIS-DDSM dataset is a curated and enhanced version of the DDSM, sourced from the University of Florida. It includes 5000 mammograms, consisting of 3000 benign and 2000 malignant instances, establishing a solid basis for assessing the suggested methodology in various cases.

The PINUM images were originally obtained in Digital Imaging and Communications in Medicine (DICOM) format and subsequently converted to Portable Network Graphics (PNG). The images were resized with linear interpolation in OpenCV to maintain the aspect ratio, and all pertinent patient information was recorded in a CSV file. Automated cropping was applied to remove extraneous black pixels from the image borders, ensuring a cleaner dataset. In addition, any imperfections from the digitized scans were addressed by converting the images to binary form to eliminate smaller artifacts. Contrast-Limited Adaptive Histogram Equalization (CLAHE) [33] was applied to the images of both datasets. It enhances contrast by dividing each image into smaller tiles and applying localized histogram equalization, avoiding the pitfalls of global adjustments. CLAHE also limits contrast amplification to prevent noise amplification, ensuring improved image quality.

To enhance the generalization of the proposed model and mitigate overfitting, we quadrupled the PINUM dataset by data augmentation, resulting in a total of 2,940 pictures. Standard augmentation methods, including vertical and horizontal flipping and rotation in combination with cropping, were utilized. The images were rotated by 15° and 45° with cropping to generate rotational variants with consistent resolution. In addition, before these operations, we applied a random perturbation based interpolation of pixel values to generate images closely aligned with its original pixel value distribution. The augmentation strategy significantly improves the accuracy of the neural network by enhancing its resilience and mitigating domain shift [34]. Algorithm 1 outlines the data augmentation process.

Algorithm 1 Image augmentation algorithm.

ENSURE: INPUT = **In(x,y)**%Input Data
ENSURE: OUTPUT = **Out(x,y,4)**%Augmented Data
Initialize(P_1, P_2) %Row perturbation values
augTypeVec = [vFlip, hFlip, Rotate15, Rotate45]
T(x,y) = Normalize[In(x,y)]
(Col,Row) = size[T(x,y)] %Number of columns and rows
X = 1:Col
for r = 1:Row
$X_p = X + Randn(P_1, P_2)$ %Random perturbation
$temp = Interp[T(r,:), X_p]$ %Row interpolation
$T(r,:) = Resize(temp, Col)$ %To the original size
end
for k = 1:4 %Iterate by augumentation types
$augType = augTypeVec(k)$
$Out(x, y, k) = Augument[T(x, y), augType]$
end
Return $\mathbf{Out(x, y, 4)}$ Augmented Data

2.2 The Overall Workflow

The block diagram of the proposed breast cancer classification framework is shown in Fig.1. It consists of several key stages: image preprocessing and augmentation, de-noising and contrast enhancement for improved MCC detection, 2D DWT for feature extraction, and finally classification with the attention based Mamba network WCMA-Net to distinguish between malignant and benign cases. Further details of each step are provided in the following sections.

2.3 De-Noising with Gaussian Notch Filter

Mammograms often contain artifacts and noise due to acquisition systems. A Gaussian Notch Filter (GNF) [35] is employed to suppress background noise and enhance the visibility of bright MCC spots against normal tissue. In the initial step, the mean and standard deviation of the lesion (positive) and background (negative) pixels are calculated using the Gaussian center difference filter. The GNF rejects frequencies within a predetermined region around the center of the frequency rectangle. It translates the center of the frequency rectangle.

$$H(u,v) = \prod_{k=1}^{Q} H_k(u,v) H_{-k}(u,v) \tag{1}$$

where Q denotes the number of notch filter pairs $H_k(u,v)$ and $H_{-k}(u,v)$ are high pass filters whose centers are at (u_k, v_k) and $(-u_k, -v_k)$. The frequency rectangles of these centers are

Under Review

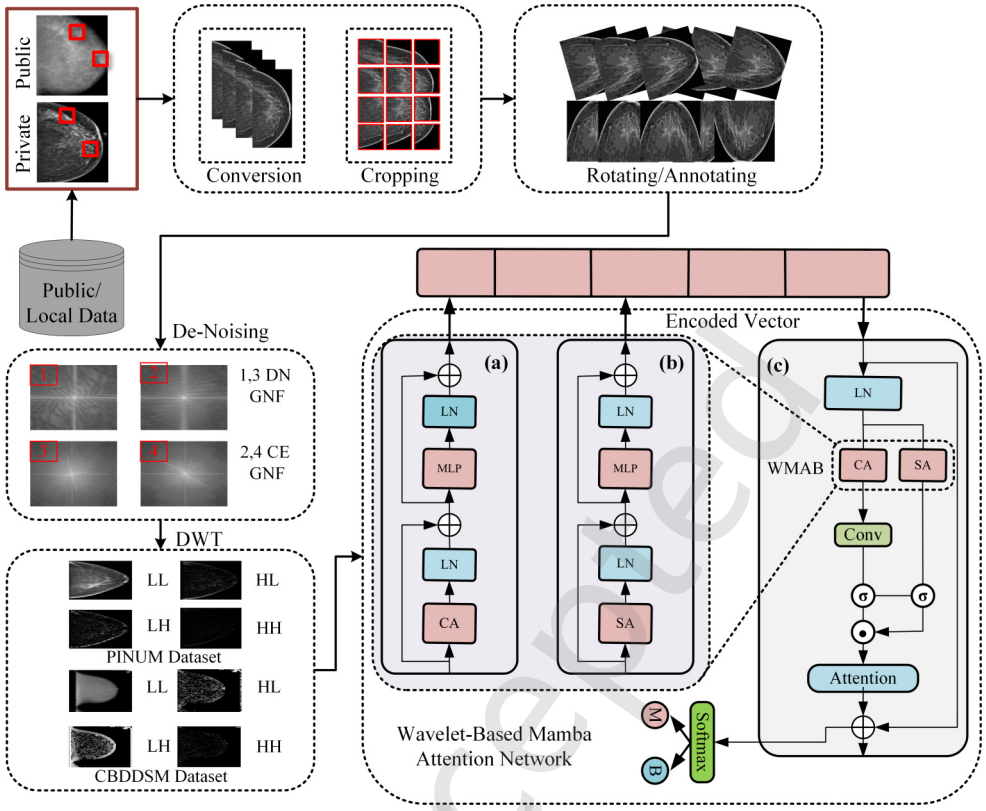


Fig. 1 The overall workflow of the proposed method. It includes preprocessing, augmentation, de-noising, wavelet feature extraction and classification of MCCs with WCMA-Net. (DN: denoised; CE: contrast enhanced; WMAB: wavelet-based Mamba attention block)

$\left(\frac{M}{2}, \frac{N}{2}\right)$. The distance between them can be computed as:

$$\begin{aligned} D_k(u,v) &= \left[\left(\frac{u-M}{2-u_k}\right)^2 + \left(\frac{v-N}{2-v_k}\right)^2\right]^{\frac{1}{2}} \\ D_{-k}(u,v) &= \left[\left(\frac{u-M}{2+u_k}\right)^2 + \left(\frac{v-N}{2+v_k}\right)^2\right]^{\frac{1}{2}} \end{aligned} \tag{2}$$

Here, M and N represent the dimensions (height and width) of the frequency rectangle, and (u_k, v_k) denotes the center coordinates of the k-th notch. The GNF of order n contains three pairs of notches whose distance from the center is the same for each pair:

$$H(u,v) = \prod_{k=1}^{3}\left[\frac{1}{1+\left[\frac{D_{k0}}{D_k(u,v)}\right]^{2n}}\frac{1}{1+\left[\frac{D_{k0}}{D_{-k}(u,v)}\right]^{2n}}\right] \tag{3}$$

In this formulation, D_{k0} is the cutoff frequency distance for the k-th notch pair, and n is the filter order controlling the sharpness of the notch transition.

2.4 Feature Extraction with Wavelet Transform

The two-dimensional DWT is employed to reduce noise while preserving essential features for contrast enhancement and feature extraction. The DWT can isolate crucial wavelet coefficients that represent the presence of MCCs. It ensures that the high-frequency components, corresponding to the MCCs, are retained while the low-frequency noise is effectively suppressed. By transforming the signal into its wavelet coefficients, we maintain clarity and accuracy in detecting suspicious regions, particularly in dense breast tissue. The DWT is essentially multi-scale decomposition:

$$\begin{aligned} a_{j+1,k} &= \sum_n LP[n-2k]a_{j,n} \\ d_{j+1,k} &= \sum_n HP[n-2k]a_{j,n} \end{aligned} \tag{4}$$

where $LP[n]$ and $HP[n]$ are low-pass and high-pass filters associated with the scaling and wavelet functions, $a_{j+1,k}$ and $d_{j+1,k}$ are the approximate and detailed coefficients at level j. Inverse DWT reconstruct the original signal by upsampling and filtering:

$$a_{j,n} = \sum_k LP[n-2k]a_{j+1,k} + \sum_k HP[n-2k]d_{j+1,k} \tag{5}$$

We use 2-level DWT coefficients to detect MCC features from the mammogram. Using 2-level DWT, a mammogram can be decomposed into subbands LL (Low-Low), LH (Low-High), HL (High-Low) and HH (High-High). LL provides the background intensity of the mammograms while HL, LH and HH contain horizontal, vertical, and diagonal features, respectively. We employ multi-scale correlation of mammograms to select the most significant wavelet coefficients using adaptive threshold de-noising.

2.5 Channel-Wise Attention and Spatial Attention

The Channel-Wise Attention (CWA) mechanism enhances the significance of critical features across input channels. For an input feature map $X \in \mathbb{R}^{C\times H\times W}$, where C, H and W are the

Just Accepted

number of image channels, height and width, respectively, CWA operates as two types of global pooling across the channels:

$$P_{\text{avg}}(c) = \frac{1}{H \cdot W} \sum_{h=1}^{H} \sum_{w=1}^{W} X(c,h,w) \tag{6}$$

$$P_{\text{max}}(c) = \max_{h,w} X(c,h,w) \tag{7}$$

The pooled features are passed through two shared fully connected layers to compute attention scores:

$$A_{\text{chl}} = \sigma\left(W_2 \cdot \text{ReLU}\left(W_1 \cdot P_{\text{avg}} + W_2 \cdot P_{\text{max}}\right)\right) \tag{8}$$

where W_1 and W_2 are learnable weights, and $\sigma()$ is the sigmoid activation function. The original input is modulated by the attention scores:

$$X_{\text{ref}}(c,h,w) = A_{\text{chl}}(c) \cdot X(c,h,w) \tag{9}$$

This mechanism selectively amplifies channels critical for identifying MCCs while suppressing irrelevant features. It should be noted that in the current study the multiple channels are from the outputs of DWT subbands, which is very different from the compositions of conventional color images or fused multimodal images where CWA has been widely applied [36].

The Spatial Attention (SA) mechanism identifies the salient spatial regions within each feature map. For the refined feature map X_{ref}, a single-channel spatial descriptor is computed using the average and maximum pooling across the channels:

$$S_{\text{avg}}(h,w) = \frac{1}{C} \sum_{c=1}^{C} X_{\text{ref}}(c,h,w) \tag{10}$$

$$S_{\text{max}}(h,w) = \max_{c} X_{\text{ref}}(c,h,w) \tag{11}$$

The pooled descriptors S_{avg} and S_{max} are concatenated and passed through a convolutional layer to compute the SA map

$$A_{\text{spl}} = \sigma\left(\text{Conv}\left([S_{\text{avg}}, S_{\text{max}}]\right)\right) \tag{12}$$

where Conv() denotes a convolutional operation. The feature map is spatially refined by modulating it with the SA map

$$X_{\text{spl}}(c,h,w) = A_{\text{spl}}(h,w) \cdot X_{\text{ref}}(c,h,w) \tag{13}$$

This ensures that spatial regions critical for the detection task are emphasized.

2.6 Selective State-Space Mamba Block

The Mamba attention mechanism is built upon a selective State-Space Model (SSM). It operates by selectively updating the hidden state h_t for an input sequence

$$h_t = \Delta_t \cdot A \cdot h_{t-1} + B \cdot X_t \tag{14}$$

where h_t is the hidden state at time t, Δ_t input-dependent gating mechanism, A forget gate that attenuates prior states, B learnable input projection matrix and X_t current input. The gating mechanism is computed as

$$\Delta_t = \text{SP}(W_3 \cdot X_t + W_4) \tag{15}$$

where W_3 and W_4 are learnable weight matrices and $\text{SP}(x) = \ln(1 + e^x)$ ensures non-negative gating values. The forget gate modulates the influence of prior states

$$A = \text{diag}\left(\exp(-\lambda \cdot t)\right) \tag{16}$$

where λ is a decay parameter controlling state attenuation and t time step; the function $\text{diag}(\cdot)$ constructs a diagonal matrix. The output of the Mamba layer y_t is given by

$$y_t = M_h \cdot h_t + M_x \cdot X_t \tag{17}$$

where M_h and M_x are learnable projection matrices. The Mamba attention mechanism combines the temporal dynamics of SSM (via h_t) and SA to ensure only the most relevant features propagate through the network.

2.7 Combining Attention Mechanism with Mamba

The connection of all the attention mechanisms within the Mamba framework is shown in Fig.2. The multiple frequency bands of DWT are used as inputs.

$$X \rightarrow \{X_{LL}, X_{LH}, X_{HL}, X_{HH}\} \tag{18}$$

High-frequency components X_{LH}, X_{HL}, X_{HH} improve the visibility of MCC. These components are processed through the CWA and SA modules to extract salient features. The refined features are fed into the Mamba attention network, capturing long-range dependencies and dynamic feature relationships. The final representation is passed through fully connected layers for binary classification using the Softmax layer.

2.8 The Network Training Process

The DWT mammogram subbands are normalized and resized to 64×64 to standardize the input dimensions. The dataset is divided into training, testing and validation sets in a ratio of 70:20:10. The batches are created for gradient-based optimization using a DataLoader. The WCMA-Net is instantiated with the hierarchical architecture. The cross-entropy loss is used to measure the classification error:

$$L_{\text{CE}} = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)\right] \tag{19}$$

where y_i is the ground truth label and $\hat{y}_i$ the predicted probability. The Adam optimizer is used to update the model weights. The weight update rules for parameter w_t at time t are:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) \nabla L \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2)(\nabla L)^2 \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \\ w_t &= w_{t-1} - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \end{aligned} \tag{20}$$

where m_t and v_t are the first and second moments of gradients, β_1 and β_2 momentum parameters and η the learning rate; ϵ prevents division by zero. Hyperparameters were carefully tuned, with a batch size of 32, a learning rate of 0.001, and a dropout rate of 0.5.

After each epoch the model is evaluated to computer the metrics such as accuracy, specificity, sensitivity, AUC, precision and F1-score. The early stopping is employed to stop the training to avoid overfitting if the validation loss does not improve after the specific number of epochs. This process ensures robust training of the proposed network.

3 Experimental Results

All algorithms are implemented in Python 3.10 using the PyTorch library. The experiments were carried out on an NVIDA GEFORCE RTX 4050 GPU, 64 GB RAM, and Windows 11 operating system. The computation time was 30 min for training

Just Accepted

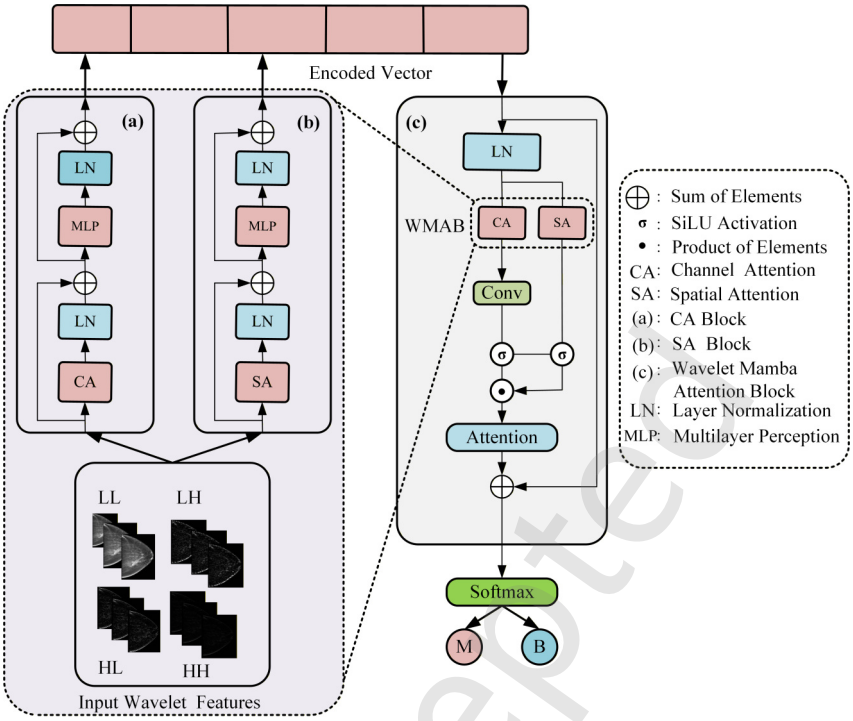


Fig. 2 Proposed WCMA-Net for MCC classification. The input wavelet features (LL, LH, HL, HH) are first processed through two separate attention blocks: (a) the CA block and (b) the SA block, each consisting of layer normalization (LN) and multilayer perception (MLP) units. Their outputs are fused and passed into (c) the Wavelet Mamba Attention Block (WMAB), which integrates CA, SA, convolutional processing, and an efficient state-space attention mechanism. The resulting encoded vector is classified via a Softmax head into malignant (M) or benign (B).

and testing. The proposed method was evaluated on the PINUM and CBIS-DDSM datasets.

3.1 Ablation Study

The performance of WCMA-Net demonstrates its robustness through an ablation study on the PINUM and CBIS-DDSM datasets. The WCMA-Net ablation was compared in seven configurations: (1) Baseline Mamba, (2) GNF denoising only, (3) DWT only, (4) Multi-scale convolution only, (5) CWA only, (6) SA only and (7) the full model. Table 1 shows that the full model significantly outperforms the other configurations on the PINUM dataset in all key metrics, including specificity (0.97), F1 score (0.90), precision (0.94), sensitivity (0.93), accuracy (0.95), and AUC (0.96). These results highlight its ability to minimize FPs and improve diagnostic precision. The sensitivity of 0.93 indicates excellent detection of TPs. In terms of computational efficiency, the proposed method achieves this performance with 0.43 GFLOPs and a parameter size of 40.5MB, balancing high accuracy with computational cost. The Multi-scale convolution only and DWT only models perform well with a specificity of 0.95 and 0.94, respectively. However, their lower F1-scores (0.83 and 0.84) and sensitivity values (0.78 and 0.82) reveal the critical role of the integrated wavelet-channel-spatial attention mechanism. The SA only variant has the lowest F1-score (0.77) and sensitivity (0.72).

For the CBIS-DDSM dataset, the full model continues to outperform all partial variants with a specificity of 0.95, an F1-score of 0.96, a precision of 0.94 and an outstanding AUC of 0.99, confirming its generalizability and superiority. The high sensitivity of 0.97 indicates its exceptional ability to identify malignant cases, a critical requirement for early cancer detection. The Multi-scale convolution only and CWA only variants also demonstrate competitive performance with a specificity of 0.94 and 0.92, respectively. Their AUC values (0.96 and 0.94) remain high, but their F1-scores (0.90 and 0.89) fall short of the full model. The SA only variant performs the worst with a specificity of 0.74 and an F1-score of 0.82.

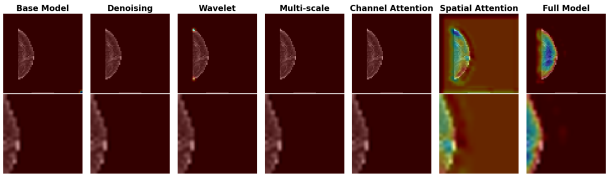


Fig. 3 Grad-CAM visualizations for ablation study on a benign case from the PINUM dataset. For each model configuration (left to right), the top image displays the full mammogram with Grad-CAM overlay, and the bottom image shows the zoomed-in ROI focusing on suspected MCC locations. (Blue = low activation, Red = high activation)

Figs.3-6 present comparative Grad-CAM visualizations of ablation results in four representative cases: benign and malignant images selected from the PINUM and CBIS-DDSM

Just Accepted

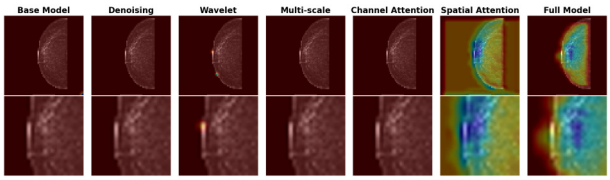


Fig. 4 Grad-CAM visualizations for ablation study on a malignant case from the PINUM dataset. For each model configuration (left to right), the top image displays the full mammogram with Grad-CAM overlay, and the bottom image shows the zoomed-in ROI focusing on suspected MCC locations. (Blue = low activation, Red = high activation)

datasets. Each figure shows the full-view and zoomed-in ROIs to evaluate how various WCMA-Net variants focus on diagnostically relevant structures. The Base model consistently shows low activation across the datasets, indicating poor localization of the lesion. While adding Denoising or DWT modules slightly enhances contrast, they lack strong attention focus. The Multi-scale module better captures contextual boundaries but still diffuses attention. The CWA and SA modules begin to emphasize suspicious regions more clearly, especially in malignant cases. However, the Full model, which integrates all modules, demonstrates the highest intensity and sharpest localization around MCCs or masses, strongly aligning with clinical expectations. This supports the generalizability and robustness of the full WCMA-Net model for mammographic interpretation under diverse imaging conditions.

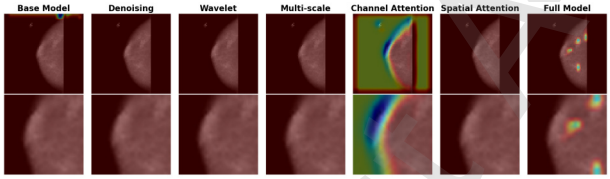


Fig. 5 Grad-CAM visualizations for ablation study on a benign case from the CBIS-DDSM dataset. For each model configuration (left to right), the top image displays the full mammogram with Grad-CAM overlay, and the bottom image shows the zoomed-in ROI focusing on suspected MCC locations. (Blue = low activation, Red = high activation)

The above ablation studies clearly demonstrate the efficacy of the proposed method. The results across both datasets highlight the importance of incorporating DWT, multi-scale analysis, and selective attention mechanisms for improved diagnostic accuracy. The proposed method not only outperforms its partial variants in terms of performance metrics but also balances computational efficiency, making it a practical solution for clinical applications.

3.2 Comparison with Existing Methods

The proposed WCMA-Net achieves superior results in multiple metrics compared to the state-of-the-art methods tested on the PINUM and CBIS-DDSM datasets as seen in Table 2 and Table 3. On the PINUM dataset, WCMA-Net achieves the highest specificity of 0.97, significantly higher than those of the

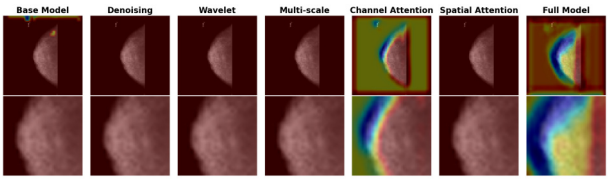


Fig. 6 Grad-CAM visualizations for ablation study on a malignant case from the CBIS-DDSM dataset. For each model configuration (left to right), the top image displays the full mammogram with Grad-CAM overlay, and the bottom image shows the zoomed-in ROI focusing on suspected MCC locations. (Blue = low activation, Red = high activation)

second best Bio-Fusion-Net (0.93) and the third best Vision-Mamba (0.91). This indicates its exceptional ability to minimize FPs in identifying non-cancerous regions. WCMA-Net's F1-Score of 0.90 surpasses those of the second best Convx-Net (0.87) and the third best Mamba-Out (0.85), demonstrating its capability to achieve a balance between precision and sensitivity. It achieves a precision of 0.94 and a sensitivity of 0.93, showing its robustness in detecting MCCs accurately without overfitting to positive cases. It also achieves the highest accuracy 0.95 and AUC 0.96, indicating its reliability in different scenarios. It requires only 0.43 GFLOPs, and thus is computationally efficient.

On the CBIS-DDSM dataset, WCMA-Net also shows exceptional performance, outperforming all other models. Achieving a specificity of 0.95, it significantly outperforms the second best Swin-Transformer (0.92). Its F1-Score of 0.96 far exceeds that of the second best EV-Mamba (0.88). It achieves an outstanding precision of 0.94 and a sensitivity of 0.97, which are critical for the accurate detection of malignant regions. It reaches the highest accuracy (0.96) and AUC (0.99), which are vital for clinical applications that require high reliability.

In the analysis of the proposed WCMA-Net against existing state-of-the-art methods on both the PINUM and CBIS-DDSM datasets, an intriguing observation is that some models exhibit lower parameter requirements but significantly higher computational cost in terms of GFLOP. The Swin-Transformer has 5.59 GFLOP which is 13 times higher than the 0.43 GFLOP of WCMA-Net, while it has fewer parameters (17.8M vs 40.5M). Despite its lower parameter count, the Swin-Transformer incurs significantly higher GFLOP due to its complex multi-head self-attention mechanism, which scales quadratically with the image size. This leads to increased computational overhead without significantly improving performance and as a result has lower accuracy (0.74) and AUC (0.78) compared with WCMA-Net on the PINUM dataset. The Local-Mamba has a higher GFLOP (0.72) and a smaller parameter size (22.8M) than WCMA-Net on both datasets. Local-Mamba leverages localized processing, which reduces parameter usage but increases GFLOP due to repeated localized operations, particularly in multi-scale feature aggregation. This results in suboptimal performance. Vision-Mamba's GFLOP is high due to its use of additional vision transformers for feature extraction. However, this does not translate into significantly better performance. WCMA-Net

Just Accepted

Table 1 Ablation study of the proposed approach on two datasets

Method	Specificity	F1-Score	Precision	Sensitivity	Accuracy	AUC	GFLOP	Param(M)	Dataset
The full model	**0.97**	**0.90**	**0.94**	**0.93**	**0.95**	**0.96**	0.43	40.5	**PINUM**
Baseline Mamba	0.94	0.79	0.88	0.72	0.89	0.93	0.41	41.5	PINUM
GNF Denoising Only	0.94	0.81	0.84	0.78	0.90	0.94	0.42	39.5	PINUM
DWT Only	0.94	0.84	0.85	0.82	0.91	0.93	0.43	40.5	PINUM
Multi-scale Conv Only	0.95	0.83	0.87	0.78	0.90	0.94	0.40	42.5	PINUM
CWA Only	0.93	0.82	0.83	0.82	0.90	0.93	**0.39**	**38.5**	PINUM
SA Only	0.94	0.77	0.83	0.72	0.88	0.92	0.45	39.5	PINUM
The full model	**0.95**	**0.96**	**0.94**	**0.97**	**0.96**	**0.99**	0.43	40.5	**CBIS-DDSM**
Baseline Mamba	0.90	0.86	0.87	0.85	0.88	0.94	0.41	41.5	CBIS-DDSM
GNF Denoising Only	0.91	0.87	0.89	0.85	0.88	0.93	0.42	39.5	CBIS-DDSM
DWT Only	0.91	0.84	0.88	0.80	0.86	0.92	0.43	40.5	CBIS-DDSM
Multi-scale Conv Only	0.94	0.90	0.92	0.87	0.91	0.96	0.40	42.5	CBIS-DDSM
CWA Only	0.92	0.89	0.90	0.88	0.91	0.94	**0.39**	**38.5**	CBIS-DDSM
SA Only	0.74	0.82	0.73	0.93	0.82	0.92	0.45	39.5	CBIS-DDSM

Table 2 Comparison of WCMA-Net and other deep learning approaches on the PINUM dataset

Algorithm	Specificity	F1-Score	Precision	Sensitivity	Accuracy	AUC	GFLOP	Param(M)
WCMA-Net (Ours)	**0.97**	**0.90**	**0.94**	**0.93**	**0.95**	**0.96**	**0.43**	40.5
V-Mamba [37]	0.90	0.71	0.73	0.68	0.83	0.89	6.29	20.3
Plain-Mamba [38]	0.88	0.77	0.73	0.82	0.86	0.93	5.5	28.9
Local-Mamba [39]	0.89	0.69	0.72	0.66	0.82	0.88	0.72	22.8
EV-Mamba [40]	0.87	0.73	0.86	0.74	0.76	0.83	34.71	46.8
Swin-Transformer [41]	0.93	0.71	0.83	0.61	0.74	0.78	5.59	**17.8**
Mamba-ND [42]	0.73	0.75	0.78	0.87	0.88	0.74	2.78	18.9
Convx-Net [43]	0.88	0.87	0.90	0.81	0.91	0.89	17.69	86.3
Mamba-Out [44]	0.89	0.85	0.79	0.92	0.90	0.92	15.59	80.6
Vision-Mamba [45]	0.91	0.82	0.81	0.83	0.89	0.92	0.74	60.4
Bio-FusionNet [46]	0.93	0.72	0.80	0.87	0.85	0.91	30.71	27.7
Alex-Net-BC [47]	0.83	0.79	0.78	0.81	0.86	0.82	2.27	58.2

Just Accepted

demonstrates an optimal balance, achieving the best results while requiring only 0.43 GFLOP and 40.5M parameters. Models with higher GFLOP may face deployment challenges in real-time systems or resource-constrained environments, where computational efficiency is critical. The proposed WCMA-Net, with its lower GFLOP, is more suitable for clinical applications that require real-time analysis.

Fig.7 shows the analysis of the runtime and training accuracy and highlights the clear superiority of the proposed method, which achieves the highest accuracy (99.5%) with the fastest runtime (7.7 ms for PINUM and 8.0 ms for CBIS-DDSM) across both datasets. This balance between computational efficiency and diagnostic precision makes it highly suitable for real-time applications in clinical settings. Compared to competing methods such as V-Mamba, Plain-Mamba, and EV-Mamba, which show moderate accuracy but higher runtimes, or Bio-FusionNet and Alex-Net-BC, which demonstrate the poorest runtime-accuracy trade-offs, the proposed method consistently excels. While some methods, such as Convx-Net, achieve competitive accuracy, their significantly higher runtime (e.g., 17.69 GFLOP) render them less practical for real-time deployment. The results emphasize that the proposed method offers an outstanding solution, leveraging both speed and accuracy to enhance the detection of breast MCCs in mammograms, thereby making a significant advancement over existing techniques.

3.3 Results Comparison with Other Approaches

Table 4 demonstrates that the performance of WCMA-Net outperforms previous studies, achieving a sensitivity of 0.93 and 0.97 on the PINUM and CBIS-DDSM datasets, respectively. Wang et al. [3] used deep learning approaches for MCC detection on a private dataset and reported sensitivities of 0.90. The machine learning models for MCC identification are used in [5] and [9] on the FDDM and DDSM datasets and obtain

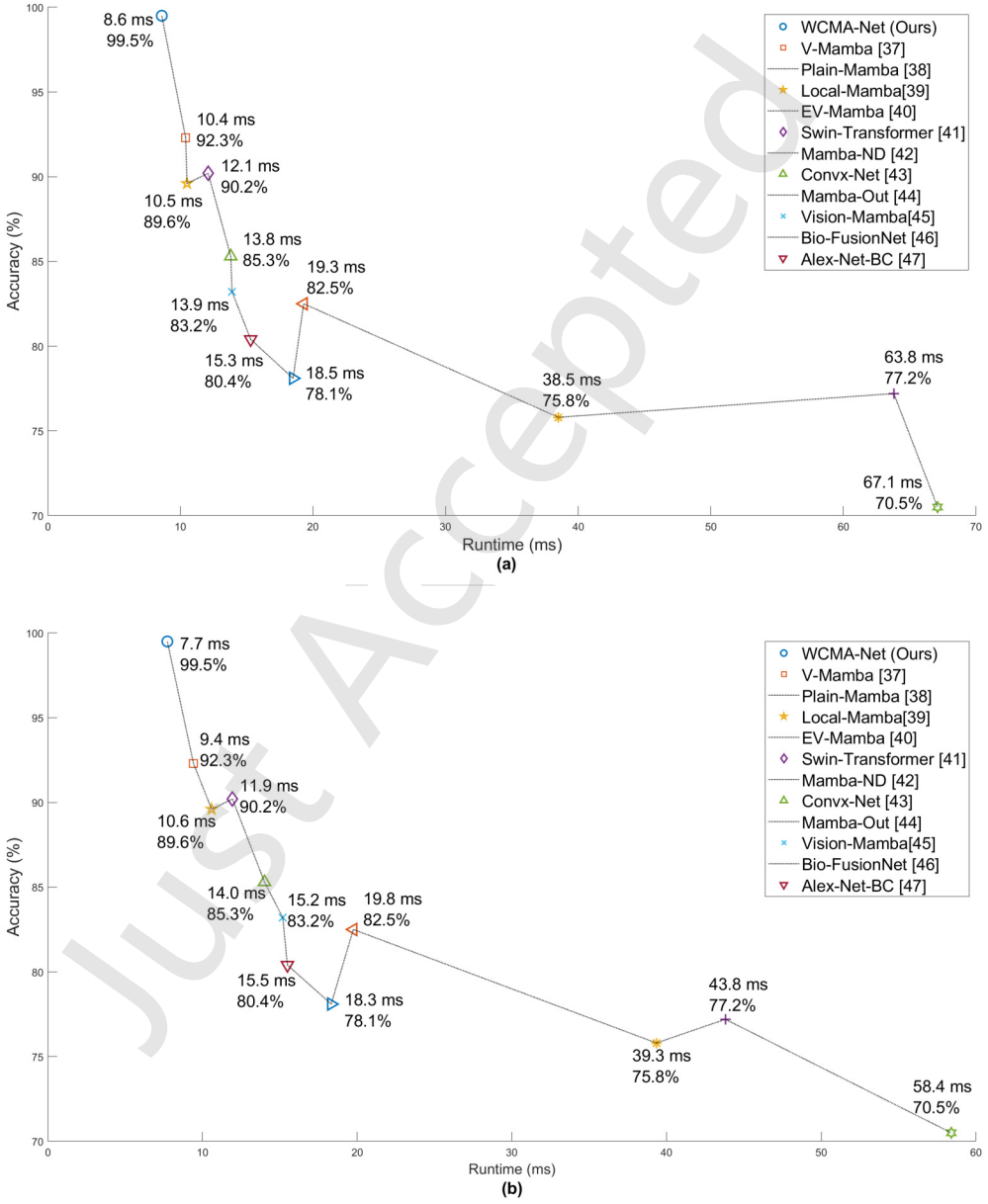


Fig. 7 Comparison of the training accuracy and runtime measures. (a) PINUM. (b) CBIS-DDSM.

Just Accepted

Table 3 Comparison results of our approach and other deep learning approaches on the CBIS-DDSM dataset

Algorithm	Specificity	F1-Score	Precision	Sensitivity	Accuracy	AUC	GFLOP	Param(M)
WCMA-Net (Ours)	**0.95**	**0.96**	**0.94**	**0.97**	**0.96**	**0.99**	**0.43**	40.5
V-Mamba [37]	0.74	0.69	0.87	0.83	0.81	0.74	6.29	20.3
Plain-Mamba [38]	0.76	0.66	0.68	0.63	0.70	0.76	5.5	28.9
Local-Mamba [39]	0.79	0.85	0.71	0.81	0.76	0.86	0.72	22.8
EV-Mamba [40]	0.83	0.88	0.86	0.81	0.82	0.89	34.71	46.8
Swin-Transformer [41]	0.92	0.81	0.68	0.72	0.88	0.76	5.59	**17.8**
Mamba-ND [42]	0.86	0.63	0.76	0.79	0.86	0.78	2.78	18.9
Convx-Net [43]	0.87	0.67	0.77	0.61	0.76	0.83	17.69	86.3
Mamba-Out [44]	0.88	0.61	0.84	0.86	0.77	0.82	15.59	80.6
Vision-Mamba [45]	0.75	0.73	0.72	0.88	0.71	0.79	0.74	60.4
Bio-FusionNet [46]	0.85	0.77	0.71	0.72	0.79	0.87	30.71	27.7
Alex-Net-BC [47]	0.87	0.73	0.69	0.75	0.78	0.85	2.27	58.2

sensitivities of 0.85 and 0.90, respectively. We conducted a comparison on both datasets with the suggested parameter values by replicating some relevant papers[11, 48, 49]. Table 4 reveals that the performance of the proposed method is better than these approaches.

Table 4 Comparison of our approach with approaches

Authors[Ref]	Methods	Dataset	Sensitivity
Wang et al.[3]	Deep Learning	Private	0.90
Brian et al.[5]	Machine Learning	FDDM	0.85
Quellec et al.[9]	Machine Learning	DDSM	0.90
Du et al.[11]	Deep Learning	CBIS-DDSM	0.84
Wang et al.[51]	Deep Learning	Private	0.91
Wang et al.[48]	Deep Learning	CBIS-DDSM	0.89
Gallardo et al.[49]	Deep Learning	CBIS-DDSM	0.81
Our approach	Deep Learning	CBIS-DDSM	0.97
Our approach	Deep Learning	PINUM	0.93

4 Discussion And Conclusion

Breast cancer detection through regular mammography screening is recognized as the most effective strategy for early identification. However, traditional mammogram diagnoses by radiologists often result in high FP rates, leading to unnecessary biopsies and patient anxiety. Although deep learning methods have shown promise in improving MCC detection, previous research has not adequately addressed FP reduction to minimize unwarranted interventions. Our proposed WCMA-Net aims to enhance the diagnostic process by accurately distinguishing between benign and malignant MCCs, supporting radiologists in making informed decisions, and potentially reducing unnecessary biopsies.

One of the key challenges in mammographic analysis is the accurate detection and classification of MCCs within dense breast tissue, where distinguishing between malignant and benign lesions is particularly complex. Conventional CAD systems often struggle with low-level features and MCC ROIs [50]. Our study introduces a new solution employing a DWT feature extraction module, integrated with the CWA and SA mechanisms. DTW decomposes mammograms into frequency subbands to isolate high-frequency components such as MCCs, improving visibility and clarity. Subsequently, the CWA selectively enhances critical feature maps, while the SA emphasizes salient spatial regions within the feature maps. These mechanisms are further refined using the selective SSM Mamba attention network, which effectively captures long-range dependencies and dynamic relationships between features, enabling a robust classification of dense and heterogeneous breast tissue. The results showed that the integration of wavelet-based denoising significantly improved classification performance over conventional techniques. The proposed method effectively addresses the class imbalance issue prevalent in clinical datasets, where benign calcifications are more frequent than malignant ones. We utilized advanced data augmentation techniques, including perturbation interpolation, rotations and flipping, to balance the training dataset and optimize model performance. These configurations allowed the network to learn robust features and mitigate overfitting, resulting in superior performance compared to the latest published models. The proposed WCMA-Net not only outperforms existing models in terms of sensitivity, specificity, accuracy, F1-score, precision, and recall, but also reduces the computational burden associated with deep learning networks.

Although the results are promising, the study has some limitations due to the lack of large-scale mammographic datasets. Future work will focus on expanding the dataset and further refining the model to improve diagnostic capabilities and generalizability. However, current findings highlight the potential of the proposed WCMA-Net as a reliable and efficient tool for early breast cancer detection, providing radiologists with enhanced diagnostic support while reducing FP and unnecessary procedures. By improving the visibility and localization of subtle MCCs, especially in dense breast tissue, WCMA-Net can help radiologists reduce diagnostic oversight and

Just Accepted

improve clinical workflow efficiency.The code is available at: https://github.com/khalilur1/WCMA-Net-

Acknowledgments

This study is partially supported by Shenzhen Talent Startup Funds (No. 827-000954) and Shenzhen University (No. 86902-000248).

References

[1] WHO, *Fact Sheet World Health Organization*, 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer

[2] P. Henrot, A. Leroux, C. Barlier, and P. Génin, "Breast microcalcifications: the lesions in anatomical pathology," *Diagnostic and interventional imaging*, vol. 95, no. 2, pp. 141–152, 2014.

[3] J. Wang, R. M. Nishikawa, and Y. Yang, "Global detection approach for clustered microcalcifications in mammograms using a deep learning network," *J. of Medical Imaging*, vol. 4, no. 2, pp. 501–511, 2017.

[4] A. Aminzadeh, B. Arhatari, A. Maksimenko, C. Hall, D. Hausermann, A. Peele, J. Fox, B. Kumar, Z. Prodanovic, M. Dimmock *et al.*, "Imaging breast microcalcifications using dark-field signal in propagation-based phase-contrast tomography," *IEEE Transactions on Medical Imaging*, vol. 41, no. 11, pp. 2980–2990, 2022.

[5] A. Bria, C. Marrocco, L. R. Borges, M. Molinara, A. Marchesi, J.-J. Mordang, N. Karssemeijer, and F. Tortorella, "Improving the automated detection of calcifications using adaptive variance stabilization," *IEEE transactions on medical imaging*, vol. 37, no. 8, pp. 1857–1864, 2018.

[6] M. V. S. de Cea, R. M. Nishikawa, and Y. Yang, "Estimating the accuracy level among individual detections in clustered microcalcifications," *IEEE transactions on ME*, vol. 36, no. 5, pp. 1162–1171, 2017.

[7] J. Zheng, H. Sun, S. Wu, K. Jiang, Y. Peng, X. Yang, F. Zhang, and M. Li, "3d context-aware convolutional neural network for false positive reduction in clustered microcalcifications detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 764–773, 2020.

[8] A. Mencattini, M. Salmeri, R. Lojacono, M. Frigerio, and F. Caselli, "Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing," *IEEE transactions on instrumentation and measurement*, vol. 57, no. 7, pp. 1422–1430, 2008.

[9] G. Quellec, M. Lamard, M. Cozic, G. Coatrieux, and G. Cazuguel, "Multiple-instance learning for anomaly detection in digital mammography," *IEEE transactions on ME*, vol. 35, no. 7, pp. 1604–1614, 2016.

[10] M. Tardy and D. Mateus, "Looking for abnormalities in mammograms with self-and weakly supervised reconstruction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 10, pp. 2711–2722, 2021.

[11] H. Du, M. M.-S. Yao, S. Liu, L. Chen, W. P. Chan, and M. Feng, "Automatic calcification morphology and distribution classification for breast mammograms with multi-task graph convolutional neural network," *IEEE JBHI*, vol. 27, no. 8, pp. 3782–3793, 2023.

[12] M. S. Mabrouk, H. M. Afify, and S. Y. Marzouk, "Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques," *Ain Shams Engineering Journal*, vol. 10, no. 3, pp. 517–527, 2019.

[13] J. Wang and Y. Yang, "A context-sensitive deep learning approach for microcalcification detection in mammograms," *Pattern recognition*, vol. 78, pp. 12–22, 2018.

[14] I. Hadjidj, A. Feroui, A. Belgherbi, and A. Bessaid, "Microcalcifications segmentation from mammograms for breast cancer detection," *International JBET*, vol. 29, pp. 1–16, 2019.

[15] S. Sannasi Chakravarthy and H. Rajaguru, "Detection and classification of microcalcification from digital mammograms with firefly algorithm, extreme learning machine and non-linear regression models: A comparison," *International Journal of Imaging Systems and Technology*, vol. 30, no. 1, pp. 126–146, 2020.

[16] T. Basile, A. Fanizzi, L. Losurdo, R. Bellotti, U. Bottigli, R. Dentamaro, V. Didonna, A. Fausto, R. Massafra, M. Moschetta *et al.*, "Microcalcification detection in full-field digital mammograms: A fully automated computer-aided system," *Physica Medica*, vol. 64, pp. 1–9, 2019.

Just Accepted

[17] R. Long, K. Cao, M. Cao, X.-T. Li, F. Gao, F.-D. Zhang, Y.-Z. Yu, and Y.-S. Sun, "Improving the diagnostic accuracy of breast bi-rads 4 microcalcification-only lesions using contrast-enhanced mammography," *Clinical Breast Cancer*, vol. 21, no. 3, pp. 256–262, 2021.

[18] A. Touil, K. Kalti, P.-H. Conze, B. Solaiman, and M. A. Mahjoub, "Automatic detection of microcalcification based on morphological operations and structural similarity indices," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1155–1173, 2020.

[19] D. Hao, L. Zhang, J. Sumkin, A. Mohamed, and S. Wu, "Inaccurate labels in weakly-supervised deep learning: Automatic identification and correction and their impact on classification performance," *IEEE journal of biomedical and health informatics*, vol. 24, no. 9, pp. 2701–2710, 2020.

[20] D. Sun, M. Wang, and A. Li, "A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 16, no. 3, pp. 841–850, 2018.

[21] X. Shu, L. Zhang, Z. Wang, Q. Lv, and Z. Yi, "Deep neural networks with region-based pooling structures for mammographic image classification," *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 2246–2255, 2020.

[22] M. M. Abdelsamea, M. H. Mohamed, and M. Bamatraf, "Automated classification of malignant and benign breast cancer lesions using neural networks on digitized mammograms," *Cancer informatics*, vol. 18, p. 1176935119857570, 2019.

[23] U. Adiga, R. Malladi, R. Fernandez-Gonzalez, and C. O. de Solorzano, "High-throughput analysis of multispectral images of breast cancer tissue," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2259–2268, 2006.

[24] R. Suresh, A. N. Rao, and B. E. Reddy, "Detection and classification of normal and abnormal patterns in mammograms using deep neural network," *Concurrency and Computation: Practice and Experience*, vol. 31, no. 14, p. e5293, 2019.

[25] M. Fan, Y. Li, S. Zheng, W. Peng, W. Tang, and L. Li, "Computer-aided detection of mass in digital breast tomosynthesis using a faster region-based convolutional neural network," *Methods*, vol. 166, pp. 103–111, 2019.

[26] A. A. Yurdusev, K. Adem, and M. Hekim, "Detection and classification of microcalcifications in mammograms images using difference filter and yolov4 deep learning model," *Biomedical Signal Processing and Control*, vol. 80, p. 104360, 2023.

[27] K. Ghuge and D. Saravanan, "Srmadnet: Swin resunet3+-based mammogram image segmentation and heuristic adopted multi-scale attention based densenet for breast cancer detection," *Biomedical Signal Processing and Control*, vol. 88, p. 105515, 2024.

[28] N. Chouhan, A. Khan, J. Z. Shah, M. Hussnain, and M. W. Khan, "Deep convolutional neural network and emotional learning based breast cancer detection using digital mammography," *Computers in Biology and Medicine*, vol. 132, p. 104318, 2021.

[29] L. Qian, J. Bai, Y. Huang, D. Q. Zeebaree, A. Saffari, and D. A. Zebari, "Breast cancer diagnosis using evolving deep convolutional neural network based on hybrid extreme learning machine technique and improved chimp optimization algorithm," *Biomedical Signal Processing and Control*, vol. 87, p. 105492, 2024.

[30] H. Qu, L. Ning, R. An, W. Fan, T. Derr, H. Liu, X. Xu, and Q. Li, "A survey of mamba," *arXiv preprint arXiv:2408.01129*, 2024.

[31] P. A. E. Commission, *Punjab Institue of Nuclear Medicine (PINUM Faisalabad)*, 2020. [Online]. Available: http://www.paec.gov.pk/Medical/Centres/

[32] C. I. Archive, *Archive, C.I.Digital Database for Screening Mammography(CBIS-DDSM)*, 2021. [Online]. Available: https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM

[33] P. SM, "Adaptive histogram equalization and its variations," *Computer Graphics and Image Processing*, vol. 6, pp. 184–195, 1977.

[34] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu *et al.*, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2531–2540, 2020.

Just Accepted

[35] I. Aizenberg and C. Butakoff, "A windowed gaussian notch filter for quasi-periodic noise removal," *Image and Vision Computing*, vol. 26, no. 10, pp. 1347–1353, 2008.

[36] Y. Cai, B. Cui, H. Deng, Z. Zeng, Q. Wang, D. Lu, Y. Cui, and Y. Tian, "Cherry tomato detection for harvesting using multimodal perception and an improved yolov7-tiny neural network," *Agronomy*, vol. 14, no. 10, p. 2320, 2024.

[37] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, and Y. Liu, "Vmamba: Visual state space model 2024," *arXiv preprint*, vol. arXiv:2401.10166, 2024. [Online]. Available: https://arxiv.org/abs/2401.10166

[38] C. Yang, Z. Chen, M. Espinosa, L. Ericsson, Z. Wang, J. Liu, and E. J. Crowley, "Plainmamba: Improving non-hierarchical mamba in visual recognition," *arXiv preprint*, vol. arXiv:2403.17695, 2024, available at https://arxiv.org/abs/2403.17695.

[39] T. Huang, X. Pei, S. You, F. Wang, C. Qian, and C. Xu, "Localmamba: Visual state space model with windowed selective scan," *arXiv preprint*, vol. arXiv:2403.09338, 2024, https://arxiv.org/abs/2403.09338.

[40] X. Pei, T. Huang, and C. Xu, "Efficientvmamba: Atrous selective scan for light weight visual mamba," *arXiv preprint arXiv:2403.09977*, 2024. [Online]. Available: https://arxiv.org/abs/2403.09977

[41] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 10 012–10 022. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2021/html/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper.html

[42] S. Li, H. Singh, and A. Grover, "Mamba-nd: Selective state space modeling for multi-dimensional data," in *European Conference on Computer Vision*. Springer, Cham, 2025, pp. 75–92. [Online]. Available: arXiv:2402.05892v5[cs.CV]13Jul2024

[43] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11 976–11 986.

[44] W. Yu and X. Wang, "Mambaout: Do we really need mamba for vision?" *arXiv preprint arXiv:2405.07992*, 2024. [Online]. Available: https://arxiv.org/abs/2405.07992

[45] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision mamba: Efficient visual representation learning with bidirectional state space model," *arXiv preprint arXiv:2401.09417*, 2024. [Online]. Available: https://arxiv.org/abs/2401.09417

[46] R. K. Mondol, E. K. A. Millar, A. Sowmya, and E. Meijering, "Biofusionnet: Deep learning-based survival risk stratification in er+ breast cancer through multifeature and multimodal data fusion," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 9, pp. 5290–5301, 2024. [Online]. Available: https://doi.org/10.1109/JBHI.2024.3418341

[47] M. Liu, L. Hu, Y. Tang, C. Wang, Y. He, C. Zeng, K. Lin, Z. He, and W. Huo, "A deep learning method for breast cancer classification in the pathology images," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 10, pp. 5025–5032, 2022. [Online]. Available: https://doi.org/10.1109/JBHI.2022.3145678

[48] K. Wang, Z. Ye, X. Xie, H. Cui, T. Chen, and B. Liu, "Mln-net: A multi-source medical image segmentation method for clustered microcalcifications using multiple layer normalization," *Knowledge-Based Systems*, vol. 283, p. 111127, 2024.

[49] R. Gallardo-Caballero, C. García-Orellana, A. García-Manso, H. González-Velasco, M. Macías-Macías *et al.*, "Independent component analysis to detect clustered microcalcification breast cancers," *The Scientific World Journal*, vol. 2012, 2012.

[50] K. U. Rehman, L. Jianqiang, A. Yasin, A. Bilal, S. Basheer, I. Ullah, M. K. Jabbar, and Y. Tian, "A feature fusion attention-based deep learning algorithm for mammographic architectural distortion classification," *IEEE Journal of Biomedical and Health Informatics*, 2025.

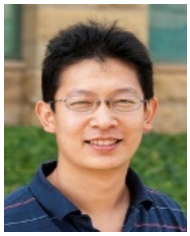
[51] J. Wang, H. Sun, K. Jiang, W. Cao, S. Chen, J. Zhu, X. Yang, and J. Zheng, "Capnet: Context

Under Accepted

attention pyramid network for computer-aided detection of microcalcification clusters in digital breast tomosynthesis," *CMPB*, vol. 242, p. 107831, 2023.

Khalil Ur Rehman working at College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China. He received a Ph.D. degree in Software Engineering from Beijing University of Technology, Beijing China in 2023. He specializes in medical imaging, deep learning, and machine learning, with over eight years of experience in developing data-driven solutions to real-world problems. His research interest include image processing, Medical image processing, computer vision, Breast Cancer, Deep learning, Control Engineering, OCT and software Engineering .

Yibin Tian Yibin Tian received his M.S. from Zhejiang University in 2002, and Ph.D. from the University of California at Berkeley in 2007. He is a Professor with the State Key Laboratory of Radio Frequency Heterogenous Integration and College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China, and a Visiting Professor with Xi'an Jiaotong University. He has 12 years of research and development experience in leading and pioneering Silicon Valley companies in sensors, optical imaging, and micro-nanoprecision manufacturing. He also founded Litemaze Technology. His current research covers micro-nanostructures and applications, optics, computational sensing and imaging, signal and image processing, machine learning/AI, and applied robotics.



Li Jianqiang received his B.S. degree in Mechatronics from Beijing Institute of Technology, Beijing, China in 1996, M.S. and Ph.D degrees in Control Science and Engineering from Tsinghua University, Beijing, China in 2001 and 2004, respectively. From 2004 to 2005, he worked as a full time researcher in Digital Enterprise Research Institute, National University of Ireland, Galway. From 2005 to 2013, he worked in NEC Labs China as Associate Researcher, Researcher, and Senior Researcher, and Department of Computer Science, Stanford University, as a visiting scholar in 2009-2010. He joined Beijing University of Technology, Beijing, China, in 2013 as Beijing Distinguished Professor. His research interests are in Petri nets, enterprise information system, business process, data mining, information retrieval, semantic web, privacy protection, and big data.

Anaa Yasin is currently pursuing his Ph.D. in the College of Software Engineering, Beijing University of Technology, Beijing, China. She received her Master degree from University of Agriculture Pakistan in 2014. Her research interests span across Artificial Intelligence (AI), Computer Vision, Medical Imaging, Pattern Recognition and Deep learning.With a strong passion for innovation and a commitment to advancing technology, her pioneering work continues to contribute to cutting-edge developments in these domains.

Weiwei Jiang received the B.Sc. Degree of Electronic Engineering and Ph.D. Degree of Information and Communication Engineering from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2013 and 2018, respectively. He is currently an associate professor with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China and Key Laboratory of Universal Wireless Communications, Ministry of Education. His current research interests include artificial intelligence, machine learning, big data, wireless communication and edge computing.

Sushil Kumar Singh is an Associate Professor in the Department of Computer Engineering at Marwadi University, Rajkot, India. He earned his Ph.D. from Seoul National University of Science and Technology (SeoulTech), Seoul, South Korea, in 2023. His research interest including Computer C Programming, Cyber Security, Big Data Analytics, Mobile Computing, and Secure and Intelligent IoT-Enabled Smart Cities, among others.

Mohammed Aloraini received his B.S. degree in Electrical Engineering from Qassim University in 2011, and his M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Illinois at Chicago in 2014 and 2020, respectively. In 2020, he joined Qassim University, where he currently serves as an Associate Professor in the Department of Electrical Engineering, Qassim University, Buraydah, Saudi Arabia. His research interests include image and video analysis, computer vision, multimedia forensics, and information security.

Just Accepted

Inam Ullah received the Ph.D. degree in Information and Communication Engineering from the College of Internet of Things (IoT) Engineering, Hohai University (HHU), Changzhou Campus, China, in 2022. He is working at the College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China. His research interests include Robotics, Internet of Things (IoT), Internet of Vehicles (IoV), ITS, Wireless Sensor Networks (WSNs), Underwater Communication and Localization, Artificial I ntelligence (AI), Machine Learning, etc.

JustAccepted