https://doi.org/10.31436/iiumej.v27i2.3387

Magnetic Resonance Imaging and Ultrasound-based Breast Tumor Classification using Soft Voting Ensemble Deep Learning Method

Suresh Ramayanam*, Vijaya Bathini, Sankar Babu Jangili, Santosh Kumar Reddy Thikkam

Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Tirupati, India

**Corresponding author: ramsuri42@gmail.com*

(Received: 6 August 2024; Accepted: 4 February 2026; Published online: 10 May 2026)

ABSTRACT: In recent years, breast cancer (BC) has been considered the most common cause of death in women worldwide. Commonly, BC occurs in breast cells or fatty tissues within the breast, and these tissues tend to grow faster and lead to death. The survival of patients with BC can be enhanced with early and accurate diagnosis. Therefore, this research proposes an ensemble learning approach that uses Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and a soft-voting mechanism to make early predictions and diagnoses of BC. The input sample images are initially passed through the convolutional layer to extract features from each modality separately. The Multi-Modality Adaptive Feature Fusion (MMAFF) technique combines generic features across modalities. Compared with existing techniques such as Deep Convolutional Neural Networks (DCNN) and Resolution Adaptive Network (RANet), the proposed ensemble learning technique achieves 89.6% accuracy for clinical indicator prediction and 99.25% accuracy for final tumor classification on the MRI-US dataset.

ABSTRAK: Dalam beberapa tahun kebelakangan ini, kanser payudara (Breast Cancer, BC) telah dikenal pasti sebagai antara punca utama kematian dalam kalangan wanita di seluruh dunia. Secara umum, BC berlaku pada sel-sel payudara atau tisu lemak dalam payudara, di mana sel-sel ini cenderung membiak dengan cepat dan berpotensi membawa kepada kematian. Kadar kelangsungan hidup pesakit BC dapat ditingkatkan melalui diagnosis awal yang tepat dan boleh dipercayai. Sehubungan itu, kajian ini mencadangkan satu pendekatan pembelajaran ansambel yang menggabungkan Rangkaian Neural Berulang (Recurrent Neural Networks, RNN), Rangkaian Neural Konvolusi (Convolutional Neural Networks, CNN), Memori Jangka Pendek Panjang (Long Short-Term Memory, LSTM), serta mekanisme pengundian lembut (soft voting) bagi tujuan ramalan awal dan diagnosis BC. Imej sampel input pada peringkat awal diproses melalui lapisan konvolusi untuk mengekstrak ciri daripada setiap modaliti secara berasingan. Seterusnya, teknik Multi-Modality Adaptive Feature Fusion (MMAFF) digunakan untuk menggabungkan ciri-ciri umum daripada setiap modaliti tersebut. Berbanding dengan teknik sedia ada seperti Rangkaian Neural Konvolusi Mendalam (Deep Convolutional Neural Networks, DCNN) dan Resolution Adaptive Network (RANet), pendekatan pembelajaran ansambel yang dicadangkan menunjukkan prestasi yang lebih baik dengan mencapai ketepatan sebanyak 89.6% bagi ramalan penunjuk klinikal dan 99.25% bagi pengelasan akhir tumor menggunakan set data MRI-US.

https://doi.org/10.31436/iiumej.v27i2.3387

KEYWORDS: *Breast Tumor Classification, Dynamic Contrast-Enhanced Magnetic Resonance Imaging, Ensemble Deep Learning, Multi-Modality Adaptive Feature Fusion Framework, Ultrasound.*

1. INTRODUCTION

Breast Cancer (BC) remains a severe health concern among females since it accounts for a high percentage of cancer-related mortality worldwide [1]. So far, several studies have investigated the use of diverse medical imaging modalities to identify BC; however, the precision of each model varies with the available dataset and the nature of the problem [2]. There is significant dependence on breast cancer imaging, making the quality of the developed images paramount for enabling accurate pathological results [3]. Additionally, several other variables significantly impact the accuracy of pathology tests [4]. These include the clinicians' degree of experience and the amount of attention they dedicate to handling samples, among other factors [5]. The diagnosis of BC using deep learning models has been extensively researched because it can improve patient survival, reduce the disease burden, and advance healthcare technology [6]. Hence, there is a need to incorporate the learning model, given advances in computer and artificial intelligence technologies, to help doctors provide precise, faster breast cancer treatment [7-8].

Deep learning models have gained further adoption within the computer vision community, particularly in biomedical image analysis, where they provide a rich set of complementary features and automatically extract both high- and low-level features from images [9-10]. Convolutional neural networks, a type of deep learning, are applied to handle histories of breast cancer images, where the primary classification is into two classes [11-12]. CNNs have also been used to learn deep representations from BC images, which are then used in models built with other classical DL methods to classify features derived from images [13]. Performance analysis of CNNs compared with traditional classifiers for pathology image classification of elite breast cancer images suggests that the former has emerged as the best among them [14]. Breast pathology image analysis using semi- or fully automated methods helps pathologists achieve better results, optimize the diagnostic workflow, and speed it up. DL models capitalize on the small features in images that are hard to overlook, given their nature [15]. Such an analysis is motivated by three key challenges widely present in current breast tumor classification methods: (1) most breast tumor classification methods rely on single-modality imaging, which may compromise diagnostic reliability; (2) many multi-modality fusion techniques cannot effectively disentangle modality-specific from modality-agnostic features; and (3) individual deep learning models often exhibit unstable performance across heterogeneous clinical data. Therefore, in this research, the aim is to: (i) design a discrimination–adaptation mechanism for separating modality-specific and modality-agnostic features with adversarial learning; (ii) propose an MMAFF framework to improve the compactness and robustness of the learned features; and (iii) build a reliable breast tumor classifier using MRI-US information based on a soft-voting ensemble strategy incorporated with CNN, RNN, and LSTM networks.

The major contributions are listed as follows:

- The ensemble deep learning method is used for classification from MRI-US. The features are extracted from various modalities distinctly by 2D and 3D convolutional layers.

- A discrimination-adaptation part is presented to attain agnostic-specific features using adversarial learning with 3 distinct discriminators.
- The feature fusion module is proposed by using MAFF to combine the modality-agnostic features and increase feature solidity by exploiting an Affinity Matrix (AM) with Nearest Neighbor Selection (NNS).

The structure of this research is as follows: Section 2 analyzes the related works. Section 3 expands the proposed ensemble learning method, and Section 4 presents the results analysis and comparison. The conclusion of this research is demonstrated in Section 5.

2. LITERATURE REVIEW

The goal of breast tumor classification is to extract and identify tumors. The relevant research focused on DL techniques for segmenting and classifying liver tumors is presented here.

Kranti Kumar Dewangan et al. [16] presented a Back Propagation Boosting Recurrent Wienmed (BPBRW) with Hybrid Krill Herd African Buffalo Optimization (HKH-ABO) model for identifying BC. This was achieved by the optimal performance of the remaining models in terms of low energy consumption and precise severity analysis. However, the combination of multiple optimization techniques and models increases system complexity, making implementation harder and requiring a large number of parameters.

Muhammad Junaid Umer et al. [17] implemented a 6B-Net Deep CNN with a feature fusion and selection model for the multi-class classification of BC. This model was efficiently applied to early detection, using Particle Swarm Optimization (PSO) and Ant Colony Search (ACS) for feature selection before performing classification. These methods converge faster and often require fewer iterations to achieve accurate breast tumor classification, effectively finding optimal paths and solutions. However, these optimization techniques struggle with high-dimensional optimization, leading to large-scale issues due to the need to update pheromone trails and evaluate multiple solutions.

Muhammad Junaid Umer et al. [18] proposed a multi-class breast cancer classification framework that incorporates 6B-Net with deep feature fusion and a selection approach. Thereby, the recommended framework comprises three phases: firstly, a 35-layer deep CNN model with a simultaneous processing block was demonstrated. Subsequently, to enable effective feature learning, the recommended approach was initially trained on the CIFAR-100 dataset. Therefore, the model was used as a feature extractor and then efficiently pre-trained for multi-class BC classification. Consequently, the extracted feature vector was used as input to the PSO model's feature selection, thereby enabling significant feature selection. Thus, the outputs of the 6B-Net selection vector and ResNet-50 selection vector were sequentially combined to improve BC classification. However, the introduced approach required analyzing the integration of different feature selection methods to further enhance classification accuracy.

Yiping Zhou et al. [19] introduced a resolution-adaptive network for classifying breast cancer from histopathological images. Specifically, a new diagnostic technique was emerging using image processing and was widely accepted among pathologists. This research presented Anomaly Detection with SVM (ADSVM) and the Resolution Adaptive Network (RANet) for anomaly detection and adaptive image resolution. In the ADSVM method, a semi-supervised approach separates misleading and undesirable trained connections to improve the RANet model's training performance. However, to enhance classification performance and computational efficiency, a feature-extraction approach was required to reproduce the RANet model's features.

Mengyun Qiao et al. [20] developed a classification of BC on MRI-US by separating the modality-specific features using MUM-Net. This method effectively leverages paired multi-modality information to enhance breast tumor classification performance by providing complementary information from MRI and US images. The method currently focuses on joint predictions across dual modalities; however, once data for a single modality are available, this could limit its applicability in certain scenarios.

From the overall analysis, the existing approach has faced challenges such as a lack of accurate identification of breast tumors due to the localization of lesions and tumor prediction. Thus, in this research, the ensemble learning approach uses RNN, CNN, and LSTM networks, combined to make predictions using a soft voting mechanism.

3. PROPOSED METHODOLOGY

Specifically, 2D and 3D convolutional layers are separately applied to extract features from different modalities. Then, a discrimination-adaptation module is designed to learn modality-agnostic features via adversarial learning with three discriminators. Thereafter, a feature fusion module is designed to integrate modality-agnostic features and is made more compact using AM with NNS. Lastly, 3 different layers are set up for the fused features to obtain the joint prediction of expression levels. Figure 1 shows the workflow of the proposed model.

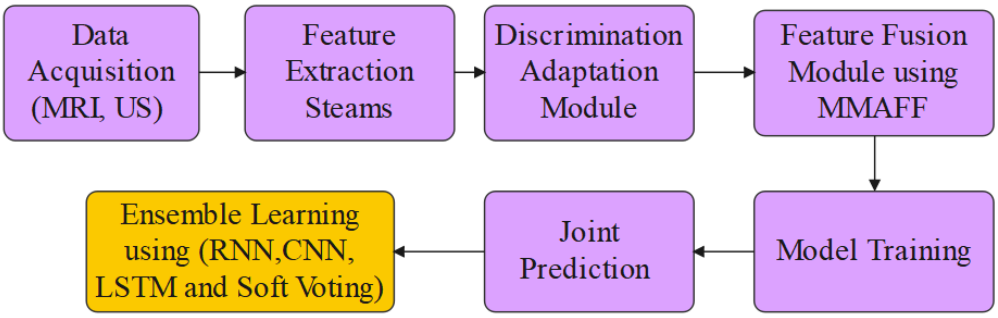

Figure 1. Workflow of the proposed model

3.1. Dataset Acquisition

This approach enrolled 502 patients who received breast US and MRI at the Cancer Center between 2017 and 2020. Institutional review board approval was acquired; however, informed consent was waived for all patients. Among the 502 patients, both MRI and ultrasound images were available for each case. In particular, the dataset comprises both benign and malignant tumors, with balanced representation. All images were reviewed and labeled by skilled radiologists based on pathological confirmation, and clinical indicators were thereby used as

https://doi.org/10.31436/iiumej.v27i2.3387

auxiliary labels [21]. For the US examination, all images were acquired by a 7-15 MHz transducer using both the Aixplorer and the Resona 5S systems. A standardized imaging protocol was used: starting with the major axis, 12 2D images were captured with a 180-degree clockwise rotation between each image at equal intervals.

The multi-angle images could provide more comprehensive architectural data than conventional single-view US images. MR imaging in the prone position was performed on multiple MR scanners with breast coils, including a 3-T MR, a 1.5-T MR, and a 3-T MR. The contrast agent was injected intravenously through a compression nozzle at an injection rate of 1.5-3 mL/s, followed by a 15 mL flush with 0.9% NaCl at a similar injection rate. All patients underwent DCE-MRI, which included one pre-contrast image and four post-contrast images for each patient. Clinical features, including LNM, histological grade, and Ki-67, were collected from the medical system for all patients. The dataset of 502 cases was divided into training, validation, and testing sets in a 70:15:15 ratio, yielding 351 training samples, 75 validation samples, and 76 testing samples. All reported performance results are obtained by comparing the predicted outputs with the ground-truth labels of the independent test set.

3.2. Feature extraction based on multi-stream

A multi-stream feature extractor is used to extract feature embeddings from both modalities. After data acquisition, feature extraction is performed using 12 US images and 5 MRI sequences, each with a different convolutional kernel. Features are extracted from the left and right breasts using an MRI scanner, then concatenated and passed through the network's feature discrimination-adaptation and fusion steps to reduce complexity. The streams for each modality are separated in the shallow convolutional layers. The US data uses a ResNet-50V2 network, which adopts the well-known residual neural network design and has been shown to achieve superior performance. The MRI feature extractor uses a 3D ResNet-50V2 architecture with 3 × 3 × 3 convolutional kernels, and the features are subsequently reduced to 512 dimensions in the last convolutional layer. ResNet-50V2, which uses many layers, skip connections, and pre-trained weights, and has a high-accuracy profile, is a solid choice for feature extraction in breast tumor classification.

3.3. Discrimination-Adaptation Module

The discrimination–adaptation module is formulated to explicitly distinguish Modality-Specific Features (MSF) and Modality-Agnostic Features (MAF). Therefore, two discriminators are incorporated to classify the extracted features as belonging to the MRI or US modality. Additionally, during adversarial training, the feature extractor learns to confuse these discriminators, which facilitates the network to provide modality-agnostic representations while retaining modality-specific characteristics for complementary learning. This module is adopted by including a dual-modality discrimination and a single-adaptation module, with embedded feature loss and adversarial training. The two losses, i.e., L_M and L_U are imposed originally through the discrimination modules of the two modalities to achieve feature disentanglement: on the one hand, to separate the modality-specific features; on the other hand, to project the Modality-Specific features (MSF) to other patients' Modality-Agnostic features (MAF) for the same patient. The prediction error is formulated by Eqs. (1) and (2):

$$\mathrm{L_M} = \mathbb{E}_\mathrm{M}\left[\left\|\mathrm{g}_{\mathrm{w_m}}(\mathrm{F_M^s}) - \mathrm{F_M^\alpha}\right\|\right] \quad (1)$$

https://doi.org/10.31436/iiumej.v27i2.3387

$$\mathrm{L_U} = \mathbb{E}_{\mathrm{U}}\left[\left\|\mathrm{g}_{\mathrm{w_U}}(\mathrm{F_U^s}) - \mathrm{F_U^{\alpha}}\right\|\right] \quad (2)$$

where $g\omega M$, $g\omega U$ signify discriminators for MRI and US, respectively. In the generation, the MSF is uncorrelated with MAF to mislead the prediction. Alternatively, the feature extractor learns MSF from MAF by adjusting the prediction loss. At that time, MSF had 3 fully connected layers, and to eliminate the MAF, which is expressed by Eq. (3);

$$\mathrm{L_{mo}} = \mathbb{E}_{\mathrm{m}}\left[-\log\left(\mathrm{p}\left(\mathrm{y_m}|\mathrm{F_m^a}, \mathrm{g_{w_a}}\right)\right)\right] \quad (3)$$

where g_{w_a} characterizes the modality adaptation constraints. $p\left(y_m|F_m^a, g_{w_a}\right)$ refers to the predicted probability of F_m^a which belongs to the ground-truth modality y_m which is a one-hot-encoded label equivalent to $\{0, 1\}$. This module confirms that the features from MAF and MSF are autonomous and encompass less unnecessary data.

3.4. Feature Fusion Module

The MMAFF framework fuses modality-agnostic features by modeling both spatial relationships and temporal dependencies. Specifically, this operation combines Spatial Attention Graph Convolution (SAGC), which helps capture spatial correlations, and Temporal Adaptive Feature Fusion (TAFF), which dynamically aggregates temporal features, thereby ensuring robust fusion across heterogeneous modalities. Thus, the modality-agnostic features obtained from the discrimination–adaptation module are embedded into the MMAFF framework for structured spatial and temporal fusion.

3.4.1. Multi-Modality Adaptive Feature Fusion Graph Convolutional Network

The Global Average Grouping (GAP) and fully connected layers are used to convert the outputs from the previous layers, yielding layer-by-layer predictions for distinct streams. MMAFF comprises SAGC and TAFF. SAGC is used to obtain data from spatial dimensions, while TAFF is used to extract temporal features, along with the initial features.

3.4.2. SAGC Module

This approach adopts a dynamic setting for creating a spatial attention map convolution. The model simultaneously feeds initial information into 2 equivalent divisions, σ, φ and ψ that are managed through 1×1 convolution operation and a Temporal Pooling (TP) block. This feature map is added through the adjacency matrix A_p for attaining optimal A_{cwt}. This module performs attentional feature fusion to replace the actual residual integration used to combine spatial and temporal-scale data, as expressed in Eq. (4).

$$A_{cwt} = \beta Q(X_{in}) + A_p \quad (4)$$

β refers to learnable constraints; A_p refers to the shared topology at $p-th$ channel; Q denotes a topological relationship of the definite channel, which is demarcated as shown in Eq. (5);

$$Q(X_i) = \sigma\left(TP\left(\phi(X_{in})\right) - TP\left(\psi(X_{in})\right)\right) \quad (5)$$

https://doi.org/10.31436/iiumej.v27i2.3387

This approach attains the channel-wise model A_{cwt}; Then, the initial features of 1×1 convolution multiply the outcomes with A_{cwt} to combine the spatial-dimension data, which is mentioned in Eq. (6):

$$X_{out} = A_{cwt} \otimes \left(\theta(X_{in})\right) \tag{6}$$

θ refers to 1 × 1 convolution; the operation of matrix multiplication is referred to as $\otimes$.

3.4.3. TAFF Module

The multi-scale temporal adaptive feature fusion comprises two components: Temporal Adaptive (TA) and Temporal Feature Fusion (TFF). TA adjusts the convolution kernel size and dilation rate across several layers using a robust attention mechanism. This part is enhanced by the multi-scale temporal convolution, which comprises 4 branches that use 1 × 1 convolution to reduce the channel dimension. This module exploits the formulation, which is stated as Eq. (7):

$$t = \frac{abs(log\, C_l,2)+b}{gamma} \tag{7}$$

C_l refers to the output channel dimension at $l-th$ network layer; $gamma$ and b stated as mapping-function constraints, varied to 2 and 1, respectively. The four branches at various scales are obtained by the aggregation function. This module does not introduce additional branches; consequently, no modification to the constraint or computational complexity. In another part of this module, an attention feature fusion combines contextual data of various dimensions with the channel. This module exploits AFF to combine features of various branches and exploits dual-input branches X and X_1. But this module focuses on the initial feature data and the temporal dimension to enable efficient feature fusion. The combination of initial and temporal features addresses the context aggregation issue, enhancing modeling efficiency, as stated in Eq. (8).

$$X' = X \otimes M(\cdot) + X_1 \otimes \left(1 - M(\cdot)\right) \tag{8}$$

The input residual connection is referred to as X; X_1 refers to the aggregated output obtained by multi-scale convolution. The formulation for $M(\cdot)$ is stated in Eq. (9);

$$M(\cdot) = (Sigmoid\left(L(X \uplus X_1) \oplus G(X \uplus X_1)\right) \tag{9}$$

$L(\cdot)$ & $G(\cdot)$ refer to the local and global channel context, correspondingly. In the attention module, the data from $L(\cdot)$ are combined with $G(\cdot)$ data The primary fusion is accomplished on X and . After applying the sigmoid, the output is displayed from 0 to 1. This approach takes the weighted average and subtracts the fusion weights 1, which is referred to as soft selection [22]. The network model calculates its weights during training.

3.5. Joint Prediction

The MRI, US, and combined MAF are summarized as a final set and processed into 3 distinct layers for joint prediction of indicators. The final task loss integrates 3 losses, which are expressed by Eqs. (10) to (13);

$$\mathrm{L_{cls}} = \mathrm{L_{t_1}} + \mathrm{L_{t_2}} + \mathrm{L_{t_3}} \tag{10}$$

$$L_{t_1} = \mathbb{E}_m \left[-log \left(p \left(y_{t_1} | [F^a, F_M, F_U], f_{\theta_{t_1}} \right) \right) \right] \tag{11}$$

$$L_{t_2} = \mathbb{E}_m \left[-log \left(p \left(y_{t_2} | [F^a, F_M, F_U], f_{\theta_{t_2}} \right) \right) \right] \tag{12}$$

$$L_{t_3} = \mathbb{E}_m \left[-log \left(p \left(y_{t_3} | [F^a, F_M, F_U], f_{\theta_{t_3}} \right) \right) \right] \tag{13}$$

F_M and F_U signifies the MAF and MSF sets in MRI and US. t_1, t_2 and t_3 denote the classification task of LNM, HG, and Ki-67.

3.6. Model Training

The total loss involves the feature loss, including , L_U and L_C for decoupling the features, task loss, including , L_{t_2}, and for predicting the three indicators. The min-max is implemented in the network training process and used by Eqs. (14) and (15);

$$L_f = L_{cls} + L_c - \lambda_1 L_{mo} - \lambda_2 (L_M + L_U) \tag{14}$$

$$L_g = -\lambda_1 L_{mo} - \lambda_2 (L_M + L_U) \tag{15}$$

where λ_1 and λ_2 are fixed at 0.2. There are 2 steps involved: first, minimize, which increases L_{mo} and the classification losses, L_{fU}, then increase L_g in various epochs. The alternative learning process is expressed by Eqs. (16) and (17):

$$f_\theta = arg \min_\theta L_f \tag{16}$$

$$g_{w_a}, g_{w_M}, g_{w_U} = arg \max_{w_a, w_M, W_U} L_g \tag{17}$$

where, g_{ω_N} signifies the overall network constraints.

4. CLASSIFICATION

Ensemble learning comprises the parallel training of multiple classifiers to improve performance in breast tumor analysis. Deep learning (DL) techniques are implemented to combine the training experiences of CNN, RNN, and LSTM for further processing. Here, three classifiers are considered in the ensemble model to produce efficient results. A proposed RNN model for BC classification consists of an input layer, 5 hidden layers, and 5 dropout layers [23]. The input layer contains an equal number of features for neurons, ReLU activation functions, and uniform kernel initialization. LSTMs differ from RNNs in how they compute the hidden state, making them more effective for sequential data such as patient history or time-series data related to tumor growth.

The LSTM [23] network is an improved version of the RNN that can learn dependencies for sequence prediction. The LSTM model incorporates enhancements to the RNN design to implement a suitable backpropagation (BP) training method. The LSTM model addresses the vanishing gradient problem prevalent in standard RNNs. This issue arises when gradients diminish over time due to repeated weight updates, but an LSTM mitigates it by propagating errors through its recurrent connections. CNN is a feedforward neural network composed of six layers: an input layer, a convolutional layer, a non-linear layer, a pooling layer, a fully

https://doi.org/10.31436/iiumej.v27i2.3387

connected layer, and an output layer. To train a CNN, the mapping function is determined using the feedforward operation, and the loss function is optimized using propagation techniques, particularly the gradient descent algorithm. The main aim of DL is to achieve multiple levels of feature classification. CNNs are popular in image classification due to their ability to detect patterns and edges. A CNN model trained on image data captures spatial features, making it particularly effective for image classification tasks. The final class label is determined by the class with the highest total probability as expressed in Eq. (18):

$$\hat{y}_i = arg\,max_k \sum_{j=1}^{m} w_j p_{i,k}^{j} \tag{18}$$

where, m represents the classifier count, weight of j^{th} classifier is stated as w_j; $p_{i,}^{j}$ denoted the prediction probability of j^{th} classifier for passing i^{th} sample to k^{th} classes.

5. Experimental Results and Discussion

In this research, the MRI and US images were regularized to [0, 1] and resized to 5 × 256 × 256 × 48 and 12 × 256 × 256, respectively. This study implemented the ensemble classifiers with PyTorch and a graphics processing unit (NVIDIA Tesla V100 32 GB). Assessment parameters, such as accuracy, precision, sensitivity, Area Under Curve (AUC), and specificity, are used to estimate model performance. The AUC is evaluated from the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate against the false positive rate at various classification thresholds. Specifically, the AUC is the area under the ROC curve, which measures the probability that the classifier ranks a randomly selected positive sample higher than a randomly selected negative sample. The mathematical formula is given in Equation. (19-22),

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{19}$$

$$Precision = \frac{TP}{TP+FP} \tag{20}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{21}$$

$$Specificity = \frac{TN}{TN+FP} \tag{22}$$

where, TP, TN denotes True Positive and True Negative; FP and FN specifies False Positive and False Negative correspondingly.

5.1. Performance Analysis

The performance of the proposed method is evaluated on the collected MRI-US dataset. ResNet-50-v2 is used to extract essential features, while MMAFF processes these features to analyze temporal relationships in the data. The proposed approach achieved better performance across all indicators, including LNM, HG, and Ki-67 levels. Table 1 and Figure 2 represent the performance analysis of clinical indicators; Table 2 and Figure 3 represent the performance analysis of classification on the MRI-US dataset.

Table 1. Performance analysis of clinical indicators

Methods	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC (%)
Lymph Node Metasis	82.5	82.7	81.9	81.8	85.7
Histological Grade	85.7	85.4	84.8	84.6	87.3
Ki-67 level	89.6	89.7	88.6	88.9	91.2

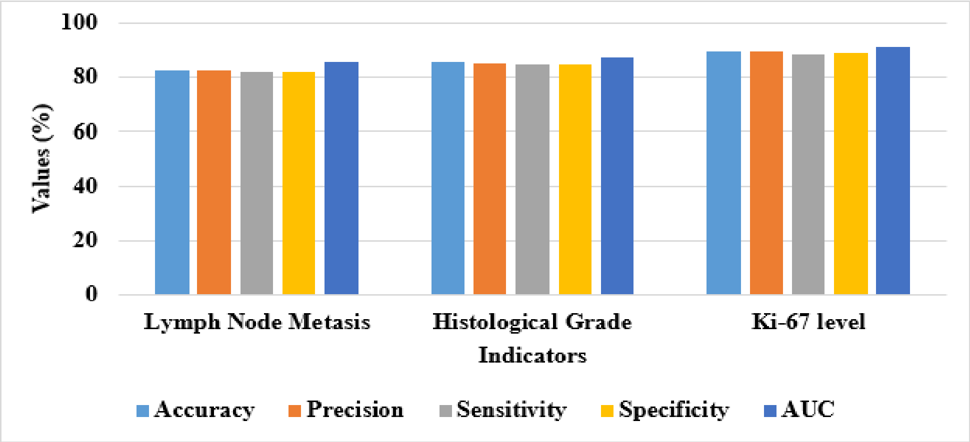

Figure 2. Graphical illustration of the clinical indicator results.

Table 2. Performance analysis of the classification on the MRI-US dataset

Methods	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC (%)
CNN	96.85	83.75	96.23	92.75	84.15
LSTM	97.15	83.96	97.45	93.25	85.26
RNN	98.52	84.16	98.25	94.63	86.12
Ensemble Learning	99.25	85.45	99.45	95.99	87.45

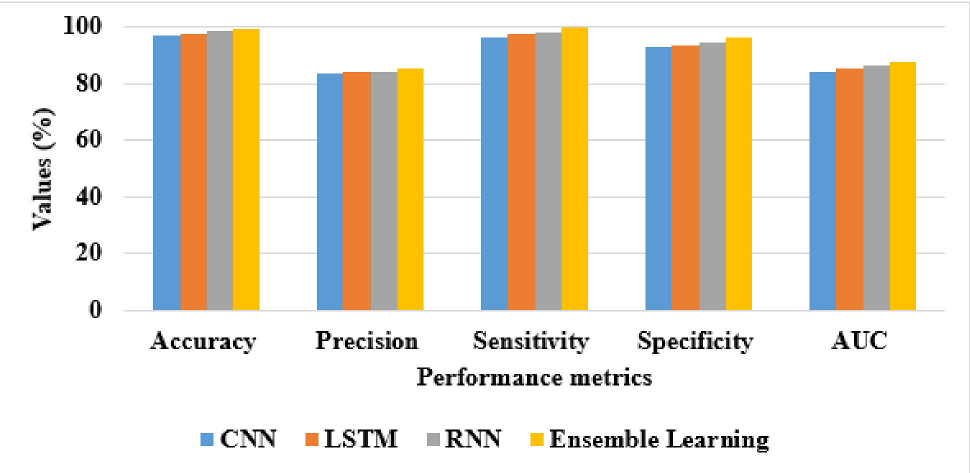

Figure 3. Graphical illustration of classification results.

From Table 2, it is observed that existing methods employing feature selection techniques, such as CNN, LSTM, and RNN, were evaluated. The ensemble learning method achieves a high accuracy of 99.25%, a precision of 97.62%, a recall of 96.02%, and an F1-score of 96.79%.

5.2. Comparative Analysis

The proposed model is compared with existing approaches, including BPBRW-HKH-ABO [16], ADSVM [19], and MUM-Net [20], as shown in Table 3 and Figure 4. Accuracy, precision, sensitivity, and specificity are used as evaluation metrics to assess model performance.

Table 3. Comparison of the proposed model

Methods	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC (%)
BPBRW-HKH-ABO [16]	99.06	N/A	99.09	95.78	N/A
RANet [19]	82.4	82.9	82.1	82.3	84.1
MUM-Net [20]	85.4	85.6	84.9	85.4	87.8
Proposed Ensemble Learning Method	99.25	85.45	99.45	95.99	87.45

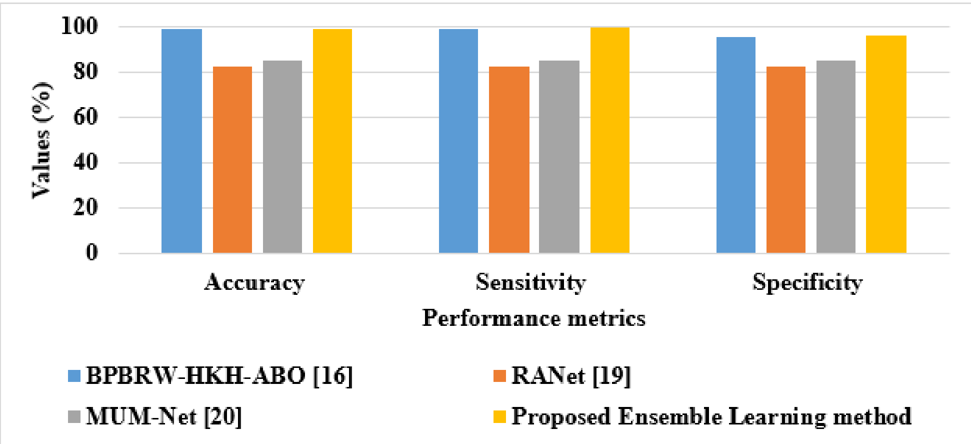

Figure 4. Graphical illustration of comparative analysis results

From Table 3 and Figure 4, it is observed that, compared with various existing models, the proposed approach achieved accuracies of 82.5%, 85.7%, and 89.6% for LNM, HG, and Ki-67, respectively.

5.3. Discussion

The proposed ensemble learning method is introduced to enhance patient survival in BC and thereby enable early and accurate diagnosis. In particular, the proposed model uses a feature fusion module to integrate modality-agnostic features. Thereby, it is necessary to distinguish between two different evaluation levels in this research. Specifically, the accuracy of 89.6% indicates that the clinical indicator prediction (Ki-67 level) incorporating the fused MMAFF features is accurate, as demonstrated in Table 1. Further, the accuracy of 99.25% pertains to the final breast tumor classification, which uses a soft-voting ensemble of CNN, RNN, and LSTM on the MRI-US dataset, as illustrated in Table 2. Hence, these values signify different tasks and are therefore not contradictory. The ensemble model outperforms individual CNN, RNN, and LSTM models because each model captures complementary characteristics: CNNs assist in extracting spatial image features, RNNs capture sequential patterns, and LSTMs handle long-term dependencies. Thus, soft voting combines these heterogeneous predictions, thereby reducing model bias and improving generalization. Despite strong performance, the research has limitations: the dataset comprises 502 cases collected from a single medical center, which is a moderate sample size for deep learning–based medical image analysis and may limit generalizability. Thus, future work should validate the framework on multi-center datasets and include additional imaging modalities.

6. Conclusion

In the proposed approach, an ensemble deep learning method is proposed for breast tumor classification from MRI and US. In this research, the ensemble learning approach uses RNN, CNN, and LSTM networks, combined via soft voting to make predictions. First, the features are extracted from 2D and 3D convolutional layers. Then, a feature fusion module is proposed that uses MMAFF to maximize feature compactness. The MRI-US dataset comprises 502 cases and is used to assess the developed ensemble learning approach. Compared with existing techniques such as Deep Convolutional Neural Networks (DCNN) and Resolution Adaptive Network (RANet), the proposed ensemble learning technique achieves 89.6% accuracy for clinical indicator prediction and 99.25% accuracy for final tumor classification on the MRI-US dataset. In future work, novel optimization techniques can be used to solve high-dimensional optimization problems and consider more classes.

References

[1] Zou Y, Zhang J, Huang S, Liu B. (2022) Breast cancer histopathological image classification using attention high-order deep network. International Journal of Imaging Systems and Technology, 32(1):266-279.

[2] Liu M, He Y, Wu M, Zeng C. (2022) Breast histopathological image classification method based on autoencoder and siamese framework. Information, 13(3):107.

[3] Ahmad N, Asghar S, Gillani SA. (2022). Transfer learning-assisted multi-resolution breast cancer histopathological images classification. The Visual Computer, 38(8):2751-2770.

[4] Alqahtani Y, Mandawkar U, Sharma A, Hasan MNS, Kulkarni MH, Sugumar R. (2022) Breast cancer pathological image classification based on the multiscale CNN squeeze model. Computational Intelligence and Neuroscience, 2022(1):7075408.

[5] Zhou Y, Zhang C, Gao S. (2022) Breast cancer classification from histopathological images using resolution adaptive network. IEEE Access, 10:35977-35991.

[6] Michael E, Ma H, Li H, Qi S. (2022) An optimized framework for breast cancer classification using machine learning. BioMed Research International, 2022(1):8482022.

[7] Wakili MA, Shehu HA, Sharif MH, Sharif MHU, Umar A, Kusetogullari H, Ince IF, Uyaver S. (2022) Classification of breast cancer histopathological images using DenseNet and transfer learning. Computational Intelligence and Neuroscience, 2022(1):8904768

[8] Balkenende L, Teuwen J, Mann RM. (2022) Application of deep learning in breast cancer imaging. In Seminars in Nuclear Medicine. WB Saunders, 52(5):584-596

[9] Liu M, Hu L, Tang Y, Wang C, He Y, Zeng C, Lin K, He Z, Huo W. (2022) A deep learning method for breast cancer classification in the pathology images. IEEE Journal of Biomedical and Health Informatics, 26(10):5025-5032.

[10] Rakha EA, Tse GM, Quinn CM. (2023) An update on the pathological classification of breast cancer. Histopathology, 82(1):5-16.

[11] Yu D, Lin J, Cao T, Chen Y, Li M, Zhang X. (2023) SECS: An effective CNN joint construction strategy for breast cancer histopathological image classification. Journal of King Saud University-Computer and Information Sciences, 35(2):810-820.

[12] Jabeen K, Khan MA, Balili J, Alhaisoni M, Almujally NA, Alrashidi H, Tariq U, Cha, JH. (2023) BC2NetRF: breast cancer classification from mammogram images using enhanced deep learning features and equilibrium-jaya controlled regula falsi-based features selection. Diagnostics, 13(7):1238.

[13] Shankar K, Dutta AK, Kumar S, Joshi GP, Doo IC. (2022) Chaotic sparrow search algorithm with deep transfer learning enabled breast cancer classification on histopathological images. Cancers, 14(11):2770.

[14] Hamedani-KarAzmoudehFar F, Tavakkoli-Moghaddam R, Tajally AR, Aria SS. (2023) Breast cancer classification by a new approach to assessing deep neural network-based uncertainty quantification methods. Biomedical Signal Processing and Control, 79:104057.

[15] Rashmi R, Prasad K, Udupa CBK. (2023) Region-based feature enhancement using channel-wise attention for classification of breast histopathological images. Neural Computing and Applications, 35(8):5839-5854.

[16] Ahmad N, Asghar S, Gillani SA. (2022) Transfer learning-assisted multi-resolution breast cancer histopathological images classification. The Visual Computer, 38(8):2751-2770.

[17] Ukwuoma CC, Hossain MA, Jackson JK, Nneji GU, Monday HN, Qin Z. (2022) Multi-classification of breast cancer lesions in histopathological images using DEEP_Pachi: Multiple self-attention head. Diagnostics, 12(5):1152.

[18] Clement D, Agu E, Suleiman MA, Obayemi J, Adeshina S, Soboyejo W. (2022) Multi-class breast cancer histopathological image classification using multi-scale pooled image feature representation (MPIFR) and one-versus-one support vector machines. Applied Sciences, 13(1):156.

[19] Umer MJ, Sharif M, Kadry S, Alharbi A. (2022) Multi-class classification of breast cancer using 6B-Net with deep feature fusion and selection method. Journal of Personalized Medicine, 12(5):683.

[20] Zhou Y, Zhang C, Gao S. (2022) Breast cancer classification from histopathological images using resolution adaptive network. IEEE Access, 10:35977-35991.

[21] Qiao M, Liu C, Li Z, Zhou J, Xiao Q, Zhou S, Chang C, Gu Y, Guo Y, Wang Y. (2022) Breast tumor classification based on MRI-US images by disentangling modality features. IEEE Journal of Biomedical and Health Informatics, 26(7):3059-3067.

[22] Akyol K, Uçar E, Atila Ü, Uçar M. (2024) An ensemble approach for classification of tympanic membrane conditions using soft voting classifier. Multimedia Tools and Applications, 1-22.

[23] Saleh H, Abd-El Ghany SF, Alyami H, Alosaimi W. (2022) Predicting breast cancer based on optimized deep learning approach. Computational Intelligence and Neuroscience, 2022(1):1820777.