BMC Medical Imaging

RESEARCH

Open Access

A deep learning approach for early prediction of breast cancer neoadjuvant chemotherapy response on multistage bimodal ultrasound images

Jiang Xie[1†], Jinzhu Wei[2†], Huachan Shi[1], Zhe Lin[1], Jinsong Lu[3*], Xueqing Zhang[4*] and Caifeng Wan[3,5*]

Abstract

Neoadjuvant chemotherapy (NAC) is a systemic and systematic chemotherapy regimen for breast cancer patients before surgery. However, NAC is not effective for everyone, and the process is excruciating. Therefore, accurate early prediction of the efficacy of NAC is essential for the clinical diagnosis and treatment of patients. In this study, a novel convolutional neural network model with bimodal layer-wise feature fusion module (BLFFM) and temporal hybrid attention module (THAM) is proposed, which uses multistage bimodal ultrasound images as input for early prediction of the efficacy of neoadjuvant chemotherapy in locally advanced breast cancer (LABC) patients. The BLFFM can effectively mine the highly complex correlation and complementary feature information between gray-scale ultrasound (GUS) and color Doppler blood flow imaging (CDFI). The THAM is able to focus on key features of lesion progression before and after one cycle of NAC. The GUS and CDFI videos of 101 patients collected from cooperative medical institutions were preprocessed to obtain 3000 sets of multistage bimodal ultrasound image combinations for experiments. The experimental results show that the proposed model is effective and outperforms the compared models. The code will be published on the https://github.com/jinzhuwei/BLTA-CNN.

Keywords Deep learning, Multistage bimodal ultrasound images, Breast cancer, Neoadjuvant chemotherapy

Introduction

Breast cancer is the most common cancer in women, seriously threatening women's physical and mental health worldwide [1, 2]. Neoadjuvant chemotherapy (NAC) is the standard treatment for breast cancer patients [3]. It not only suitable for locally advanced breast cancer and high-risk operable patients at risk of breast cancer recurrence or metastasis but also for understanding tumor sensitivity to chemotherapeutic agents and developing more rational chemotherapy regimens [4–6]. However, not all patients have a good outcome after NAC, with approximately 10–35% of patients not responding significantly after NAC [7, 8]. Patients who fail to achieve results after several courses of NAC not only suffer irreversible physical and psychological damage but also

†Jiang Xie and Jinzhu Wei contributed equally to this work.

*Correspondence:
Jinsong Lu
lujinsongdoctor@163.com
Xueqing Zhang
yuqing79@sina.com
Caifeng Wan
wancaifengky@sina.com
[1] School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China
[2] School of Medicine, Shanghai University, Shanghai 200444, China
[3] Department of Ultrasound, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China
[4] Department of Pathology, Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China
[5] Department of Breast Surgery, School of Medicine, Renji Hospital, Shanghai Jiao Tong University, Shanghai 200030, China

miss out on the best opportunity for surgical treatment. Accurate evaluation of chemotherapy response in the early stage of NAC treatment will help doctors to adjust the treatment plan in time and significantly improve the possibility of pathologic complete response (pCR) [9]. Therefore, it is essential to develop a method that can accurately predict the efficacy of NAC for breast cancer at an early stage.

At present, the pathological examination is the gold standard for evaluating the efficacy of NAC for breast cancer, however, its invasive and late nature limits its use in the early assessment of NAC efficacy. As the development of imaging continues, imaging modalities such as ultrasound and magnetic resonance of the breast are often performed in parallel with NAC, allowing assessment of the biological properties of the oncology and predicting the response to NAC in breast cancer, in addition to measuring tumor size [10]. Ultrasound has become the preferred screening method for breast disease because it is non-radioactive, economical and reproducible [11–15]. With the continuous development of newer ultrasound imaging technologies, different ultrasound imaging techniques play their respective roles in assessing the efficacy of NAC, providing an imaging basis for selecting treatment options and prognosis of breast cancer patients.

However, ultrasonography is an examination that is highly dependent on the operating and diagnosing physicians, primarily because of its low imaging resolution, low imaging quality due to scattered noise and artifacts, and insufficient detection of tissue details [16], which can affect the objectivity and accuracy of diagnosis to a certain extent [17]. With the rapid development of deep learning, convolutional neural network (CNN) based models are used as feature extractors to automatically extract more abstract and higher-level features to predict the pathological response of NAC. Nowadays, several studies have applied CNNs to breast cancer NAC efficacy prediction.

Some studies [18–20] designed a dual-branch neural network based on ultrasound images to early predict NAC efficacy. These models aimed to simultaneously use pre-neoadjuvant chemotherapy (pre-NAC) and after the first NAC (NAC1) ultrasound images. Predictions were made by calculating the similarity between the ultrasound images of pre-NAC and NAC1. Gu et al. [21] expanded the approach proposed in [20] and developed a deep learning radionics pipeline using cascading models constructed at different stages of NAC treatment. The cascade consists of two Siamese networks. The first network predicted efficacy through ultrasound images pre-NAC and after the second NAC (NAC2). The second Siamese network aimed to predict the outcome from ultrasound images pre-NAC and fourth NAC (NAC4).

In addition, Adoui et al. [22] proposed a multi-input deep learning architecture for predicting NAC responses in breast cancer based on MRI images for the first time, using a parallel CNN architecture to explore changes in lesions in MRI images pre-NAC and NAC1 simultaneously, ultimately enabling the classification of patient pathological responses with considerable accuracy. The study also developed a single-input model that used MRI images from pre-NAC or NAC1 as input and found that the predictive effect of using multistage data was better than that of single-stage data, demonstrating the great potential of multistage data early in chemotherapy in the field of NAC efficacy prediction. The performance of deep learning models can be improved by fusing MRI image features from different modalities [23–25]. Based on this, Joo et al. [26] proposed a multimodal deep learning model that combines MRI images and clinical information to predict whether a breast cancer patient achieves complete pathological remission using pre-chemotherapy MRI-T1 and MRI-T2 images and the patient's clinical information based on an improved 3D-ResNet50 architecture.

These studies results showed that the multimodal deep learning model using the fusion of clinical details and MRI images achieved better prediction performance than the deep learning model without fusion. The results of these studies suggest that a computer-aided approach that fuses multimodal information can help to improve the early prediction of NAC responses in breast cancer.

In clinical practice, doctors evaluate the efficacy of neoadjuvant chemotherapy in both of grayscale ultrasound (GUS) and color Doppler flow imaging (CDFI). GUS evaluates the efficacy by comparing changes in tumor size, echoes of tumors and tumor marginal tissue before and after chemotherapy, while CDFI can observe blood flow in the mass and adjacent tissues, which effectively reflects the changes in breast cancer blood vessels. However, existing work faces two issues: Firstly, effectively fusing features from multimodal data. Previous researches [23–26] typically treated different modalities as inputs to branch networks and performed concatenation operations along the channel dimension to merge high-level feature information between different modalities. However, this approach has not been effective in exploring the highly complex relationships and complementary feature information between different modalities, lacking an effective means of feature sharing for multi-modal data. (2) Effectively extracting temporal information from data at different chemotherapy stages. the dual-branch convolutional neural network is constructed for neoadjuvant chemotherapy efficacy prediction, with each branch taking input from the imaging data at each chemotherapy stage. Only a simple concatenation operation is

performed on the features of each stage's data just before the network's fully connected layers, and this approach does not fully utilize the temporal relationships between data from different chemotherapy stages, leading to the loss of tumor change characteristics during the chemotherapy process.

In order to solve the above problems, we propose a deep learning method for early prediction of NAC for breast cancer based on multistage bimodal ultrasound images. This model is experimental on grayscale ultrasound and color Doppler flow imaging before and after the first stage of chemotherapy. The convolutional neural network model consists of two key modules: the bimodal layer-wise feature fusion and the temporal hybrid attention. The Bimodal Layer-wise Feature Fusion Module (BLFFM) learns richer high-level features from the two modalities of ultrasound images at the same stage. Then, the generated feature maps at different stages are input into the Temporal Hybrid Attention Module (THAM), which can capture key features related to the nature and lesion changes of breast tumors before and after NAC, and finally perform further prediction tasks. The model achieves excellent computer-aided diagnostic performance on the data we have collected. The three main contributions of this paper are as follows:

1) This paper proposed the first convolutional neural network model based on bimodal layer-wise feature fusion and temporal hybrid attention module (BLTA-CNN) for early efficacy prediction of NAC for breast cancer based on multistage bimodal ultrasound images.
2) A new Bimodal Layer-wise Feature Fusion (BLFFM) is designed. It can effectively mine the highly complex correlations between different modal data and the complementary feature information between data and is an effective method for sharing features of bimodal data.
3) The Temporal Attention (TA) module is introduced based on Convolutional Block Attention Module (CBAM) [27] to form the Temporal Hybrid Attention Module (THAM). It not only learns important features of the images, but also correlations between features of lesion changes during chemotherapy and reinforces the network's ability to understand key features of breast tumor progression.

Methods

Figure 1 shows the overall framework of the BLTA-CNN. It uses ResNet50 as the backbone network and grayscale ultrasound (GUS) and color Doppler flow imaging (CDFI) images data from pre-neoadjuvant chemotherapy (pre-NAC) and the first NAC (NAC1) as input to the four-branch network, and share the weight of the backbone of these branches. The BLFFM is designed for both modality data streams at the same stage, by which the model learns the unique information of a single modality and the complementary information between

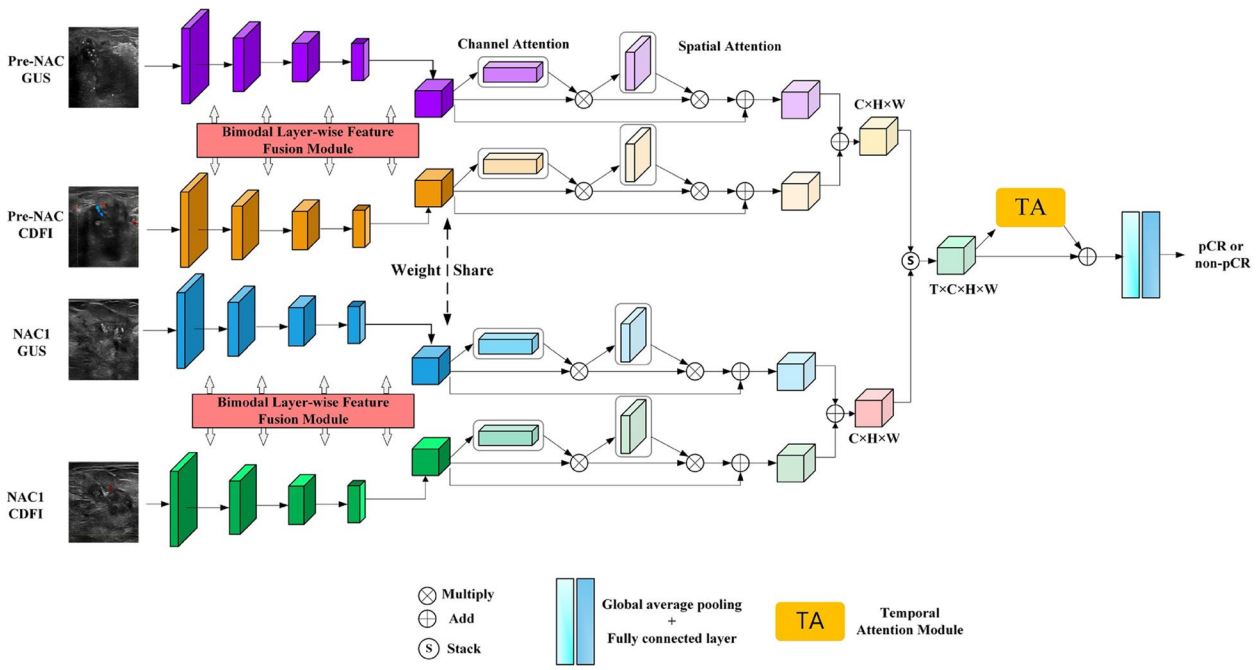


Fig. 1 The overall flowchart of the proposed BLTA-CNN framework

different modalities. In addition, the model incorporates the THAM, which is able to not only focus on the important features of cancer, but also capture the key features related to the change of breast cancer lesions before and after NAC1. Finally, the Global Average Pooling Layer and a fully connected layer complete the mapping from high-to-low-dimensional features, thereby enabling early breast cancer NAC efficacy prediction.

Bimodal Layer-wise feature fusion module

The Layer-wise Feature Fusion Module (LFFM) is a typical network layer-level feature fusion approach. In the traditional level-wise fusion strategy, a single modal image is used as a single input to a single network. The independent feature representations learned by each network are fused into each layer of the network. Finally, the fusion results are fed back to the decision layer to obtain the final prediction results. It can effectively integrate and fully use bimodal images. Its dense connections between network layers can capture the complex relationships between modalities, entirely using more abstract, complex, and complementary information to enhance training for better performance [28–30]. Inspired by HyperDenseNet [28], the BLFFM connects the outputs of the corresponding layers from different modal data streams so that the different modal data in each layer are interrelated and the inputs of each layer of the same modal data stream correspond to the outputs of all previous layers, facilitating the flow of information. The advantage is that by connecting the layers of different modalities, the complex relationships between the different modalities can be captured. At the same time, complementary information is learned, and a significantly more enriched feature representation can be produced.

Figure 2 shows the implementation details of the BLFFM. Firstly, the model receives grayscale ultrasound (GUS) and color Doppler blood flow imaging (CDFI) as inputs. Initially, separate convolution operations are performed on different modalities to obtain their respective initial feature maps, namely F_g^0 and F_c^0. Subsequently, the i^{th} layer feature map of one modality is obtained by convolution and nonlinear activation function operation on the feature map of the $(i-1)^{th}$ layer of this modality path and the fusion feature map of the $(i-1)^{th}$ layer that is obtained by BLFFM. The $(i-1)^{th}$ layer fusion feature map is calculated by combining all the previous feature maps of this modality path with the feature maps of the $(i-1)^{th}$ of another modality path. For example, for the GUS modality, the second layer feature map F_g^2 is obtained by calculating the feature maps F_g^1 and the first

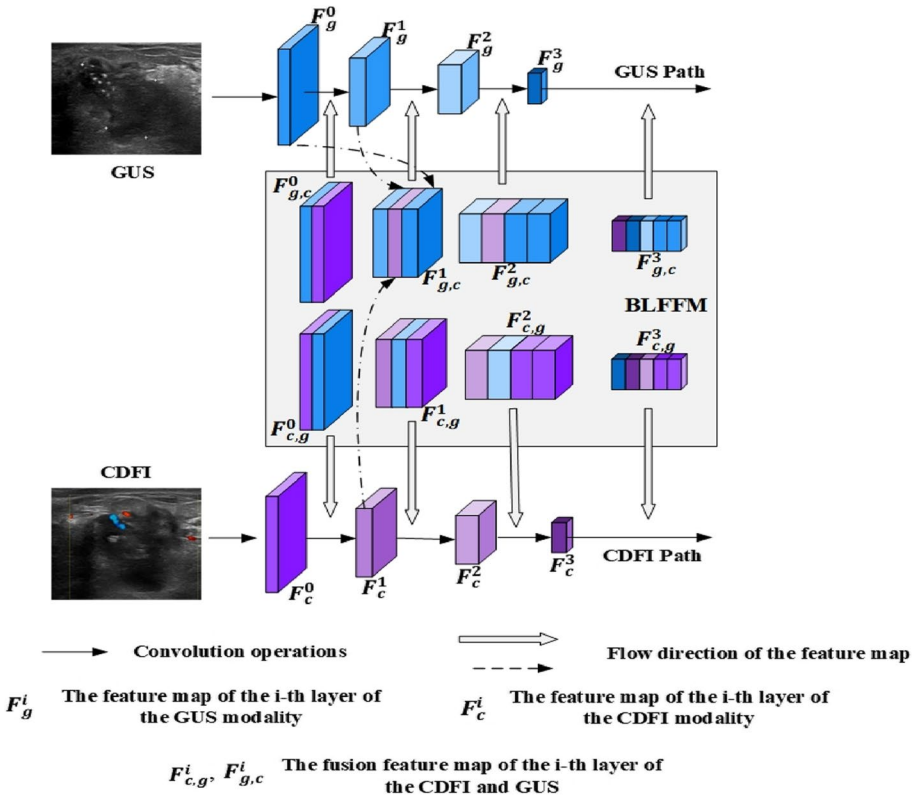


Fig. 2 An illustration of the BLFFM

layer fusion feature map $F_{g,c}^{1}$, where the fusion feature map $F_{g,c}^{1}$ is obtained by calculating the feature maps F_{g}^{0} and F_{g}^{1} of the GUS path, along with the feature map F_{c}^{1} of the CDFI path. This BLFFM is calculated as follows:

$$F_{g,c}^{i-1} = Conv\left(\left[F_{g}^{i-1}, F_{c}^{i-1}, F_{g}^{i-2}, \ldots, F_{g}^{0}\right]\right) \tag{1}$$

$$F_{c,g}^{i-1} = Conv\left(\left[F_{c}^{i-1}, F_{g}^{i-1}, F_{c}^{i-2}, \ldots, F_{c}^{0}\right]\right) \tag{2}$$

where $F_{g,c}^{i-1}$ and $F_{c,g}^{i-1}$ respectively represent fused feature maps obtained through BLFFM,*Conv* represents dimensionality reduction achieved by a 1×1 convolutional operation, and [⋯] denotes the concatenation operation of the feature maps. Finally, $F_{g,c}^{i-1}$ is combined with the $(i-1)^{th}$ layer feature maps F_{g}^{i-1} of the GUS path to compute the i-th layer feature map F_{g}^{i}, the same goes for F_{c}^{i}. The feature maps F_{g}^{i} and F_{c}^{i} can be computed as:

$$F_{g}^{i} = H_{g}^{i}\left(F_{g,c}^{i-1} + F_{g}^{i-1}\right) \tag{3}$$

$$F_{c}^{i} = H_{c}^{i}\left(F_{c,g}^{i-1} + F_{c}^{i-1}\right) \tag{4}$$

where H_{g}^{i} and H_{c}^{i} represent the mapping function consisting of convolution operation and nonlinear activation function.

Temporal hybrid attention module

This study further improved the network structure to fully use the temporal relationship between pre-NAC and NAC1 stages and improve the model's prediction accuracy. This paper proposes a Temporal Hybrid Attention Module (THAM), which introduces temporal information based on Convolutional Block Attention Module (CBAM) [27]. It can not only capture the key features of breast tumor related attributes in the feature map, reduce the influence of non-tumor tissue information in the ultrasound image, but also enhance the learning ability of the network model for the key feature changes of breast tumor lesions before and after NAC.

The THAM comprises the CBAM and the TA modules. The CBAM attention mechanism is a more comprehensive feature attention method that combines the channel domain and spatial domain attention. The target area is enhanced by adding a spatial attention module based on the channel attention module. It is assumed that F_{g}^{p}, F_{c}^{p}, F_{g}^{n}, $andF_{c}^{n}$ represent the output feature maps of GUS and CDFI for pre-neoadjuvant chemotherapy (pre-NAC) and the first NAC (NAC1), respectively, where g and c illustrate the GUS and CDFI modal data, respectively, and p, n represent the pre-NAC and NAC1 data, after the CBAM attention module produces the corresponding feature maps $F_{g}^{p'}$, $F_{c}^{p'}$, $F_{g}^{n'}$, $F_{c}^{n'}$. The feature maps can be computed as:

$$F_{g}^{p'} = M_{sa}\left(M_{ca}\left(F_{g}^{p}\right) \otimes F_{g}^{p}\right) \otimes \left(M_{ca}\left(F_{g}^{p}\right) \otimes F_{g}^{p}\right) + F_{g}^{p} \tag{5}$$

$$F_{c}^{p'} = M_{sa}\left(M_{ca}\left(F_{c}^{p}\right) \otimes F_{c}^{p}\right) \otimes \left(M_{ca}\left(F_{c}^{p}\right) \otimes F_{c}^{p}\right) + F_{c}^{p} \tag{6}$$

$$F_{g}^{n'} = M_{sa}\left(M_{ca}\left(F_{g}^{n}\right) \otimes F_{g}^{n}\right) \otimes \left(M_{ca}\left(F_{g}^{n}\right) \otimes F_{g}^{n}\right) + F_{g}^{n} \tag{7}$$

$$F_{c}^{n'} = M_{sa}\left(M_{ca}\left(F_{c}^{n}\right) \otimes F_{c}^{n}\right) \otimes \left(M_{ca}\left(F_{c}^{n}\right) \otimes F_{c}^{n}\right) + F_{c}^{n} \tag{8}$$

where M_{ca} represents the channel attention weighting factor, M_{sa} represents the spatial attention weighting factor, and $\otimes$ denotes the element-by-element multiplication. The bimodal fusion feature maps $F_{g-c}^{p'}$, $F_{g-c}^{n'}$ of the two chemotherapy stages were obtained by fusion of the different modal data in the same stage. The feature maps can be computed as:

$$F_{g-c}^{p'} = F_{g}^{p'} + F_{c}^{p'} \tag{9}$$

$$F_{g-c}^{n'} = F_{g}^{n'} + F_{c}^{n'} \tag{10}$$

After obtaining the features from the CBAM between each stage, feature maps $F_{g-c}^{p'}$, $F_{g-c}^{n'}$ are spliced together to obtain the CBAM feature sequence that can be represented by X, which can be computed as:

$$X = stack\left(F_{g-c}^{p'}, F_{g-c}^{n'}\right) \tag{11}$$

where stack represents the stitching of feature maps $F_{g-c}^{p'}, F_{g-c}^{n'}$ by temporal dimension.

The CBAM can only focus on the key features of the tumor, but cannot mine the characteristic information of tumor lesion changes, so to model the long-distance dependencies in the image sequences before and after NAC for breast cancer, we developed a temporal attention module on top of the CBAM. As illustrated in Fig. 3, the module is to estimate the salience and relevance of all regions in the breast cancer NAC image sequence through the time regardless of their distance.

The input features $X \in R^{T\times C\times H\times W}$ is first converted into two feature spaces $q(X)$ and $k(X)$ by two sets of $1 \times 1 \times 1$ convolutions, where C, H, and W are its channel, height, and width, $q(X) = W_{q}X$ and $k(X) = W_{k}X$ (W_{q} and W_{k} are trainable weight matrices) respectively. Subsequently, we reshape both $q(X)$ and $k(X) \in R^{M\times C}$, where $M = T \times H \times W$, to calculate the attention map of any pairs of regions through time dimension. The attention map $F_{T}(X)$ is given as follows:

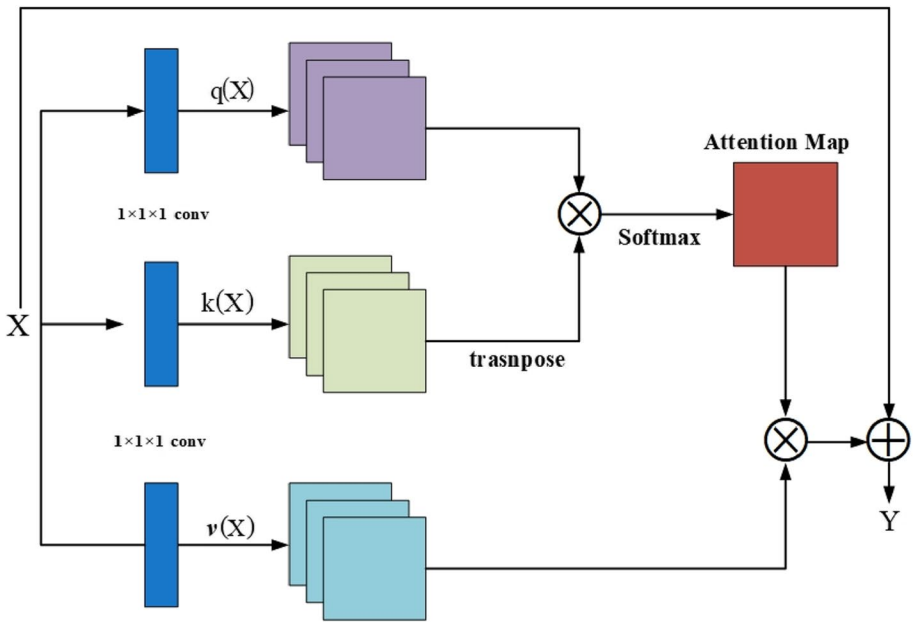


Fig. 3 Temporal attention module

$$F_T(X)_{j,i} = \frac{exp(s_{ij})}{\sum_{i=1}^{M} exp(s_{ij})} \tag{12}$$

where $s_{ij} = q(X)k(X)^{\intercal}$. $F_T(X)_{j,i}$ demonstrates how much X_i correlate with X_j. T represents the temporal dimension or the length of the sequence, i and j denote the indexes of the two regions or feature locations involved in the attention calculation, respectively. The output feature map of the temporal attention is $Y = \left(Y_1^T, Y_2^T, \cdots, Y_j^T, \cdots, Y_M^T\right)$, where Y_j^T represents the feature vectors at different positions in the sequence. Y_j^T is given as follows:

$$Y_j^T = \sum_{i=1}^{M} F_T(X)_{j,i} v(X_i) + X_j \tag{13}$$

$v(X) = W_v X$ (W_v is a learnable matrix) and X_j is added back to keep more information. With such a design, the hybrid attention module can not only focus on important tumor tissue features, but also mine potential features of tumor lesion progression before and after NAC1.

Experimenter and results

Datasets and preprocessing

The data for this study were collected from the partner hospitals, with a collection of ultrasound videos (including GUS and CDFI) and their pathology data from 101 patients with locally advanced breast cancer. The dataset is a tracked ultrasound image dataset collected from 2015 to 2020. All patients completed a four-stage course of NAC, and the post-chemotherapy pathological findings were confirmed by pathological histology. In other words, each patient has video data of GUS and CDFI for pre-NAC, NAC1, NAC2, and NAC4, as well as a final pathological response report. In this work, only the ultrasound video data from the pre-NAC and NAC1 were used for the study, with the report of pathological response as the gold standard. This is a retrospective clinical study and has been ethically approved by the ethics committee of the partner hospital.

The data collected in this study were collected using the Esaote MyLab^{TMT}wice ultrasound device with the LA332 probe. And the device is capable of acquiring grayscale ultrasound video and Color Doppler blood flow video. To input it into the neural network, the ultrasound video of each patient was pre-processed before starting the training process, which consisted of four steps, as shown in Fig. 4. The first step is to cut the video at a fixed frame interval (the length of the video may vary from patient to patient) to form M^i ultrasound images (i represents the i-th patient). The second step is to select N^i high-quality images of breast cancer by removing some images that contain artifacts, blurring, and non-diseased tissue. The process was carried out by two specialist radiologists (with 5 and 10 years of breast ultrasound experience, respectively) who read the breast ultrasound images independently without knowledge of the patient's disease information and reached a consensus through discussion to ensure the correctness and reproducibility of the dataset. The third step is to remove additional information

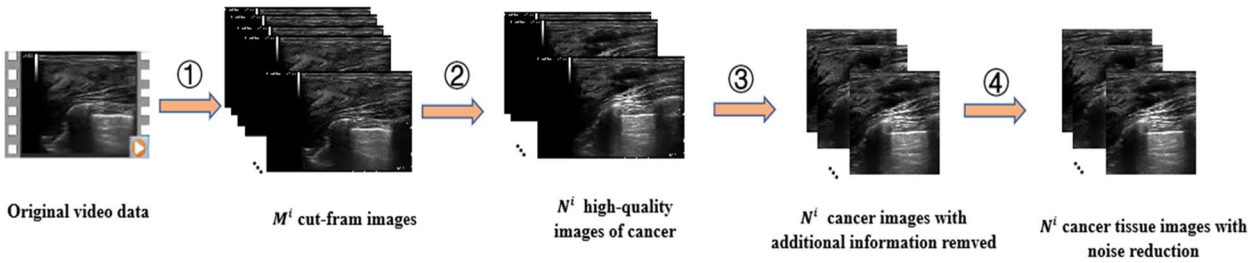


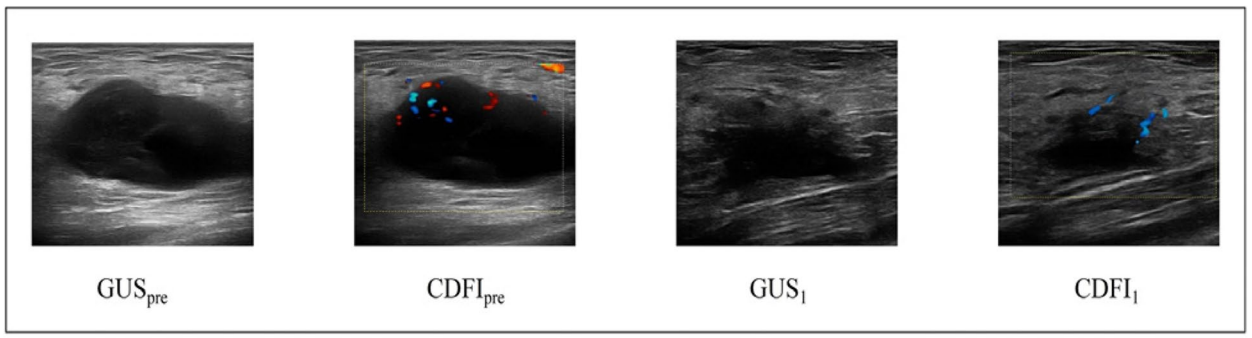
Fig. 4 Data pre-processing flow chart (example for patient *i*)

from the images, such as the model of the instrument, scan or imaging time, and patient information, and to retain the complete ultrasound image of the breast cancer tissue. Finally, all ultrasound images are resized to 256×256 before being fed into the deep neural network.

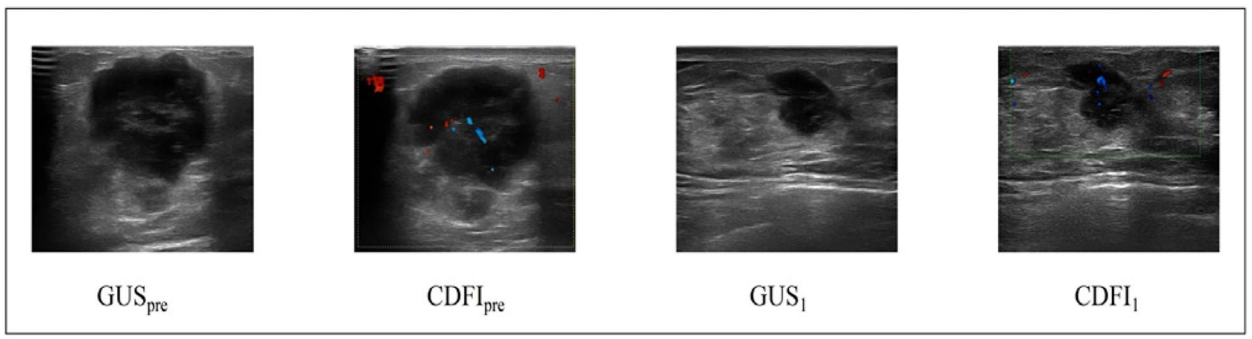
By performing the above four pre-processing steps on the raw ultrasound video of GUS and CDFI of the patient's pre-NAC and NAC1 chemotherapy phases, resulting in $4N$ images (GUS_{pre}, $CDFI_{pre}$, GUS_1, $CDFI_1$). The four images in each group need to ensure that the lesion cross-section position is basically the same, which helps the model to capture the correct and rich key features of the lesion change. These four images correspond to the four-branch input of the BLTA-CNN model. For example, for the 100th patient, 10 images ($N^{100}=10$) are selected from their original GUS and CDFI videos at each of the two stages, and 40 images are chosen to form 10 sets of image combinations, noted as $S_1^{100}, S_2^{100}, \cdots, S_{10}^{100}$, each contains 4 images, such as:

$$s_k^{100} = \left\{ (GUS_{pre})_k^{100}, (CDFI_{pre})_k^{100}, (GUS_1)_k^{100}, (CDFI_1)_k^{100} \right\}, k \in [1,10].$$

Each branch of BLTA-CNN model input an image respectively, namely GUS_{pre}, $CDFI_{pre}$, GUS_1, or $CDFI_1$. Figure 5 shows the combination of images for each data set, with each group labeled with the patient's final pathological response result (pCR or non-pCR).



(A) Combination of Images with pathological finding of pCR



(B) Combination of Images with pathological finding of non-pCR

Fig. 5 Schematic of the image combination samples input to the network

The original video data were preprocessed with the above data to obtain 3000 sets of NAC ultrasound image combinations, each containing GUS and CDFI image data of pre-NAC and NAC1, and the corresponding pathological findings of NAC response. There were 930 sets of ultrasound images (37 patients) with pathological results of pCR and 2070 sets of ultrasound images (64 patients) with non-pCR pathological results.

Experimental evaluation

In our experiments, we divided the dataset based on patient-level. Specifically, we assigned a unique identifier to each patient, then randomly sorted the list of patient identifiers and cross-divided the dataset fivefold based on the list, so that all images of the same patient are in the training or test set separately, rather than being distributed across different sets. The classification accuracy (Acc), sensitivity (Sen), specificity (Spec), Positive Predictive Value (PPV), Negative Predictive Value (NPV), and F1-score are used as evaluation metrics computed as follows:

$$\begin{cases} Acc = \frac{TP+TN}{TP+TN+FP+FN} \\ Sen = \frac{TP}{TP+FN} \\ Spec = \frac{TN}{TN+FP} \\ PPV = \frac{TP}{TP+FP} \\ NPV = \frac{TN}{TN+FN} \\ F1-score = \frac{2TP}{2TP+FP+FN} \end{cases} \tag{14}$$

where TP is the number of true positive, TN is the number of true negative, FP is the number of false positive, and FN is the number of false negative.

Experiments design and implementation details

A series of experiments are conducted on the dataset to evaluate the performance and significance of the proposed BLTA-CNN.

1) Experiment 1: This experiment utilized only the ResNet50 model to evaluate the effectiveness of combining grayscale ultrasound (GUS) and color Doppler flow imaging (CDFI) data from pre-neoadjuvant chemotherapy (pre-NAC) and the first phase of chemotherapy (NAC1).
2) Experiment 2: Ablation experiments on two modules of BLTA-CNN to validate the effectiveness of the layer-wise feature fusion module and the temporal hybrid attention module proposed in this paper to achieve the early prediction task of NAC efficacy.
3) Experiment 3: This experiment compares the method proposed in this paper with mainstream deep learning classification models to evaluate the performance of the present model.

The details on the network training were as follows: Adam [31] with adaptive learning was used as the optimizer, the initial learning rate was set to 0.001, the weight decay factor was 0.1, the number of training iterations was set to 500, the learning rate was reduced to 1/10 of the previous rate every 150 iterations, and the loss function uses the cross-entropy loss function used in the classification task. The experiments were performed on a Dell T640 tower server deep learning workstation with two NVIDIA GeForce RTX 2080Ti discrete graphics cards and two Intel Xeon Silver 4110 CPUs with 64 GB RAM. All models involved in the experiments are based on PyTorch 1.9.0 implementation.

Results

1) Results of Experiment 1: Table 1 shows the results of the different modalities of ultrasound imaging data for different stages of chemotherapy on ResNet50. Specifically, using GUS_{pre}, $CDFI_{pre}$, GUS_1 and $CDFI_1$, respectively, and their combination of $GUS_{pre}+CDFI_{pre}$, GUS_1+CDFI_1, $GUS_{pre}+GUS_1$ and $GUS_{pre}+GUS_1+CDFI_{pre}+CDFI_1$ ($GUS_{pre+1}+CDFI_{pre+1}$) to predict the efficacy. In the

Table 1 Classification results of the different modalities of ultrasound imaging data for different stages of chemotherapy (unit: %)

Data	Acc	Sen	Spec	PPV	NPV	F1-score
GUS_{pre}	77.47 ± 0.63	**72.72± 4.28**	79.57 ± 1.62	61.18 ± 0.95	86.89 ± 1.62	66.38 ± 1.71
$CDFI_{pre}$	77.07 ± 1.48	63.37 ± 5.04	83.13 ± 1.74	62.45 ± 2.36	83.74 ± 1.86	62.82 ± 2.97
GUS_1	78.27 ± 1.63	71.63 ± 2.13	81.20 ± 1.52	62.79 ± 2.52	86.61 ± 1.05	66.92 ± 2.30
$CDFI_1$	77.87 ± 1.40	62.07 ± 2.26	**84.86± 1.56**	64.51 ± 2.67	83,49 ± 0.90	63.24 ± 2.16
$GUS_{pre}+CDFI_{pre}$	79.07 ± 0.59	71.09 ± 4.09	82.60 ± 1.64	64.43 ± 1.22	86.64 ± 1.40	67.51 ± 1.67
$GUS_{pre}+GUS_1$	78.83 ± 0.61	71.85 ± 6.12	81.92 ± 2.61	63.88 ± 1.65	86.93 ± 2.22	67.45 ± 2.11
GUS_1+CDFI_1	79.27 ± 0.33	71.74 ± 3.02	82.60 ± 1.47	64.63 ± 1.08	86.88 ± 1.02	67.95 ± 0.94
$GUS_{pre+1}+CDFI_{pre+1}$	**79.70± 0.67**	72.39 ± 2.94	82.93 ± 1.91	**65.33± 1.82**	**87.19± 0.97**	**68.61± 0.89**

GUS_{pre+1}+$CDFI_{pre+1}$ combination, the model could achieve the optimal results in accuracy, positive predictive value, negative predictive value, and F1-score index, and the sensitivity and specificity were close to the optimal results. Meanwhile, the experimental results of GUS-pre, $CDFI_{pre}$, and GUS_{pre}+$CDFI_{pre}$, and GUS_1, $CDFI_1$, and GUS_1+$CDFI_1$ showed that the prediction effect of bimodal data was better than that of single modality in the same chemotherapy stage. It also can be further observed that the results of GUS_{pre}+GUS_1 are better than those of GUS_{pre}, and GUS_1, which indicated the effectiveness of multistage data was better than that of single chemotherapy stage and confirmed the effectiveness of bimodal data of different chemotherapy stages for NAC efficacy prediction.

2) Results of Experiment 2: Table 2 shows the results of different feature fusion methods on the GUS_{pre+1}+$CDFI_{pre+1}$ image dataset. This experiment compared the proposed Bimodal Layer-wise Feature Fusion Module (BLFFM) with other common feature fusion methods. It validated the complementary information of different modes of data streams and the effect of feature fusion between different abstraction layers of the same data stream on the prediction performance of the network.

In Table 2, Sum and Concat are the two commonly used feature fusion methods. Sum superimposes values on the feature map element by element while keeping the number of channels constant. At the same time, Concat performs a merge operation on the number of channels. Neural discriminative dimensionality reduction (NDDR) [32] is a feature fusion that can automatically learn each abstraction layer from different data streams. Specifically, features with the same spatial resolution in a single-branch network are cascaded by channel, and a convolution operation dimensionally reduces the features with a 1×1 convolution kernel. Finally, the fused features were fed into the next layer of the network. BLFFM is a layer-wise feature fusion approach proposed in this paper. BLFFM1 and BLFFM2 represent two implementations to maintain the same spatial resolution of the feature maps output from different convolutional layers, respectively, where BLFFM1 is a downsampling method of nearest neighbor interpolation and BLFFM2 is to maintain the same spatial resolution by clipping the edge information. The results showed that the BLFFM method achieved the best performance. Among them, the accuracy, specificity, positive predictive value, and F1-score of BLFFM2 reached 83.93 ± 1.22%, 74.62 ± 7.44%, and 73.98 ± 1.4%, respectively. BLMFF1 had the best sensitivity (76.20 ± 4.64%) and negative predictive value (89.14 ± 1.75%). In addition, BLFFM2 improved accuracy by 2.70% (81.23% vs. 83.93%), specificity by 4.04% (83.89% vs. 87.93%), and F1-score by 3.04% (70.94% vs.73.98%) compared to the Sum fusion method, which ranked third overall. It shows that the proposed method can effectively integrate and fully use the rich feature information between multimodal and single-modal data, thereby improving the model's prediction performance.

Table 3 shows the results of the different attention mechanisms on the GUS_{pre+1}+$CDFI_{pre+1}$ image dataset. It is clear from Table 2 that the layer-wise feature fusion module is effective; therefore, the results of the experiments in Table 3 all use ResNet50 with the introduction of the bimodal layer-wise feature fusion module BLFFM2 as the baseline model to verify the effectiveness of the Temporal Hybrid Attention Module (THAM).

The influence of channeled, spatial attentional mechanisms was first explored, including the five classical attentional mechanisms of Squeeze-and-Excitation (SE) [33], Bottleneck Attention Module (BAM) [34], Dual Attention Network (DANet) [35], Coordinate Attention (CA) [36], and Convolutional Block Attention Module (CBAM) [27]. The results showed that among the five channel and space attention mechanisms, the CBAM attention mechanism achieved the best F1-score of 79.07 ± 1.68%, which was 5.09% (79.07% vs. 73.98%), 3.68% (79.07% vs. 75.39%), 1.69% (79.07% vs. 77.38), 4.27% (79.07% vs.74.80%) and 2.35% (79.07% vs. 76.72%) better than no attention mechanism (baseline), SE, BAM, DANet and CA, respectively.

Table 2 Classification results of the different fusion methods (unit: %)

Fusion methods	Acc	Sen	Spec	PPV	NPV	F1-score
Sum	81.23 ± 1.74	75.22 ± 8.95	83.89 ± 5.10	68.22 ± 5.31	88.71 ± 3.06	70.94 ± 2.95
Concat	81.00 ± 1.42	73.70 ± 2.52	84.23 ± 1.66	67.45 ± 2.45	87.87 ± 1.04	70.41 ± 2.06
NDDR	71.09 ± 8.75	**86.83± 4.83**	71.34 ± 5.02	**87.41± 3.05**	70.63 ± 2.62	71.09 ± 8.75
BLFFM1	82.90 ± 1.37	76.20 ± 4.64	85.87 ± 2.43	70.63 ± 2.90	**89.14± 1.75**	73.18 ± 2.22
BLFFM2	**83.93± 1.22**	74.89 ± 8.21	**87.93± 4.77**	74.62 ± 7.44	89.00 ± 2.62	**73.98± 1.84**

Table 3 Classification results of the different attentional module (unit: %)

Attentional	Acc	Sen	Spec	PPV	NPV	F1-score
Baseline	83.93 ± 1.22	74.89 ± 8.21	87.93 ± 4.77	74.62 ± 7.44	89.00 ± 2.62	73.98 ± 1.84
+ SE	85.27 ± 0.17	73.80 ± 4.63	90.34 ± 2.01	77.42 ± 2.71	88.69 ± 1.54	75.39 ± 1.24
+ BAM	85.17 ± 0.46	83.15 ± 6.79	86.06 ± 3.03	72.80 ± 2.56	92.21 ± 2.90	77.38 ± 1.52
+ DANet	86.13 ± 0.48	67.61 ± 7.27	**94.33± 3.41**	**85.13± 5.94**	86.95 ± 2.32	74.80 ± 1.93
+ CA	86.20 ± 0.39	74.79 ± 7.92	91.25 ± 3.45	79.85 ± 4.72	89.30 ± 2.79	76.72 ± 1.96
+ CBAM	87.20 ± 0.24	79.46 ± 7.81	90.63 ± 3.57	79.87 ± 6.01	91.07 ± 2.66	79.07 ± 1.68
+ THAM	**88.53± 0.28**	**83.48± 3.82**	90.77 ± 1.59	80.15 ± 2.24	**92.60± 1.44**	**81.67± 0.90**

In this study, the THAM module was based on the CBAM attention mechanism. The addition of the Temporal Attention (TA) module achieved optimal prediction results in terms of accuracy, sensitivity, negative predictive value, and F1-score, with a 2.60% increase in F1-score (81.67% vs. 79.07%) compared to the next best performing CBAM attention module. It shows that the traditional attention mechanism method is extended to the temporal dimension to adapt to the characteristics of tracking image data, which can effectively extract multitype information, including global space, channel, and temporal, and enhance the feature representation to improve the performance of the network. These results suggest that focusing on the time sequence information of ultrasound imaging data is helpful for the prediction of NAC response.

In addition, to investigate the impact of different combinations of modules on the model's prediction of NAC effectiveness, we used ResNet50 as the baseline and conducted experiments with different combinations of BLFFM, CBAM, and TA modules. The experimental results are shown in Table 4. We clearly observed that using all three modules, BLFFM, CBAM and TA, simultaneously yielded better predictive performance compared to other combinations.

3) Results of Experiment 3: This compares the performance of BLTA-CNN with the deep learning classification model of the mainstream. These include EfficientNet [37], DenseNet [38], ShuffleNetV2 [39], Xception [40], MobileNetV2 [41], InceptionV4 and Inception_ResNetV2 [42], ResNeXt [43], these models represent various current mainstream CNN classification models. Figure 6 shows that the accuracy, sensitivity, positive predictive value, negative predictive value, and F1-score of the proposed BLTA-CNN model are 88.53%, 83.48%, 80.15%, 92.60%, and 81.6%, respectively, which are better than other mainstream classification models. The effectiveness of the proposed BLFFM and THAM for early prediction of NAC efficacy using GUS and CDFI data from pre-NAC and NAC1 was confirmed. Moreover, it can be observed from the box plot that the proposed BLTA-CNN model demonstrates relatively consistent performance across each fold of the dataset, indicating its stability across the different fold datasets, and also suggests that the BLTA-CNN model exhibits a high level of robustness. Therefore, it can identify NAC patients with different efficacy more accurately and consistently, and can provide a meaningful reference for clinical auxiliary diagnosis and personalized treatment scheme [44, 45].

Discussion

The early prediction of neoadjuvant chemotherapy (NAC) efficacy is critical for the improvement and personalized of treatment in breast cancer patients. At

Table 4 Classification results of the different module combinations (unit: %)

Methods	Acc	Sen	Spec	PPV	NPV	F1-score
ResNet50	79.70 ± 0.67	72.39 ± 2.94	82.93 ± 1.91	65.33 ± 1.82	87.19 ± 0.97	68.61 ± 0.89
+ BLFFM	83.93 ± 1.22	74.89 ± 4.63	87.93 ± 4.77	74.62 ± 7.44	89.00 ± 2.62	73.98 ± 1.84
+ CBAM	83.27 ± 0.59	75.43 ± 7.23	85.81 ± 4.21	71.28 ± 5.86	89.07 ± 3.22	73.70 ± 2.01
+ TA	83.01 ± 0.33	76.55 ± 6.81	84.07 ± 3.21	70.61 ± 4.96	87.72 ± 2.98	71.21 ± 2.34
+ BLFFM + CBAM	87.20 ± 0.24	79.46 ± 7.81	90.63 ± 3.57	79.87 ± 6.01	91.07 ± 2.66	79.07 ± 1.68
+ BLFFM + TA	85.26 ± 1.19	78.91 ± 5.97	87.07 ± 2.19	77.9 ± 5.12	88.72 ± 3.05	77.58 ± 2.13
+ CBAM + TA	86.32 ± 0.93	80.98 ± 4.98	85.77 ± 3.65	71.21 ± 4.61	90.79 ± 2.35	75.39 ± 2.26
+ BLFFM + CBAM + TA	**88.53± 0.28**	**83.48± 3.82**	**90.77± 1.59**	**80.15± 2.24**	**92.60± 1.44**	**81.67± 0.90**

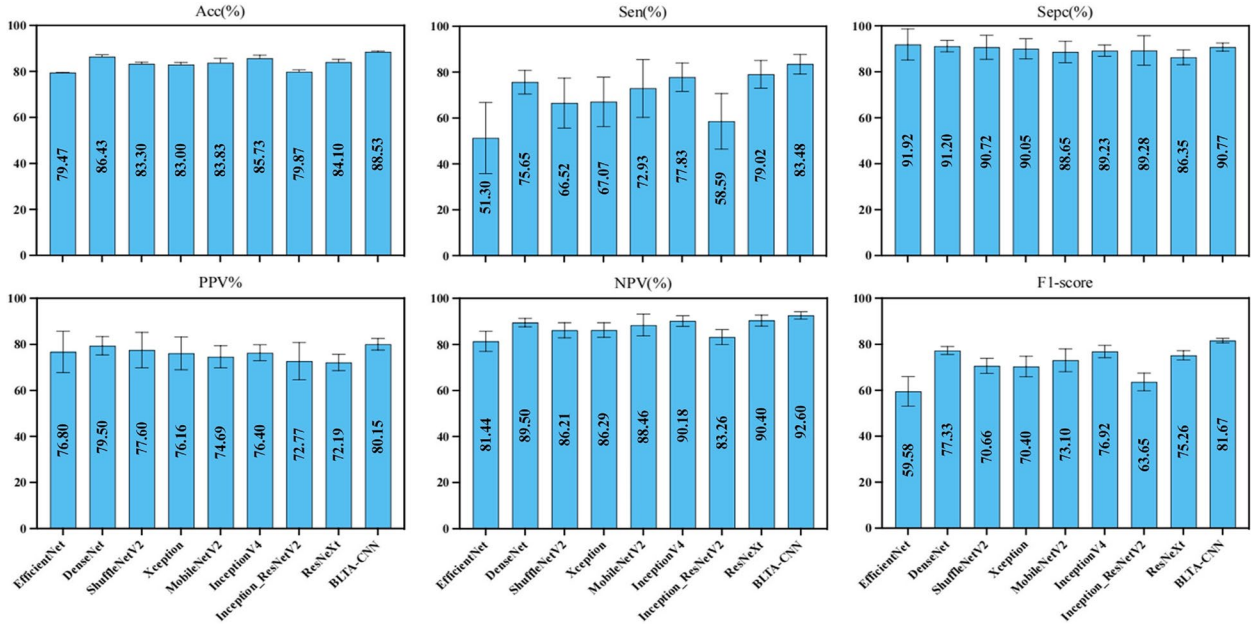


Fig. 6 Comparison of the classification performance of BLTA-CNN with other mainstream models

different stages of NAC, gray-scale ultrasound can record the changes in tumor size in real time in different sections. However, the initial response is not always a change in tumor size, and the density of tumor cells may also change, so the use of gray-scale ultrasound alone to measure the maximum diameter of the mass may not be a comprehensive evaluation of the effect of NAC [46]. Due to the rapid growth of the breast cancer, the pressure on blood vessels and resistance index (RI) will be increased [47]. After NAC, some sensitive tumor tissues are denaturized and necrotic, the internal blood vessels are significantly atrophied, and the volume is reduced. The pressure on blood vessels is also reduced, and RI decreases accordingly. Color Doppler flow imaging (CDFI) can predict the response of breast cancer to NAC by detecting the distribution of tumor vessels and the pattern of vascular blood supply. In this study, experiments were performed using ResNet-50 in different data combinations. Table 1 results also demonstrate the great potential of multistage bimodal ultrasound data combinations for early predict the efficacy of NAC in breast cancer.

Then, we propose the BLFFM, which can effectively explore the highly complex correlations between different modal data and the complementary feature information between the data, and Table 2 shows that it is an effective method for sharing the features of bimodal data. Finally, the Temporal Attention (TA) module was further introduced on top of Convolutional Block Attention Module (CBAM) in order to enable the model to focus not only on the important features of each modal image itself at each stage and suppress unnecessary regional responses, but also on the key information about the important feature changes before and after NAC. Table 3 shows that the proposed temporal attention module can enhance the model's ability to learn key features of lesion changes in different stages of breast tumors. Thus, the proposed deep learning method BLTA-CNN for multistage bimodal ultrasound images is a data-driven deep learning model. The model can accurately predict treatment outcomes at an early stage of chemotherapy, enabling doctors to adjust chemotherapy regimens in a timely manner. For example, for patients predicted to have poor chemotherapy efficacy, doctors can quickly modify the combination of chemotherapy drugs or switch to other treatment approaches (such as targeted therapy or immunotherapy), thereby maximizing therapeutic effectiveness and increasing the likelihood of achieving pathological complete response (pCR). This precise predictive capability significantly enhances the scientific and practical value of treatment decisions. In addition, the model's predictions can reduce unnecessary chemotherapy cycles, sparing patients from the physical side effects and psychological stress caused by ineffective treatments. Earlier efficacy evaluation allows patients to choose more appropriate treatment plans in a timely manner, thereby improving their quality of life. Furthermore, the model's predictions can provide patients with clear treatment

expectations, strengthening their understanding and trust in the treatment process. This personalized and precise approach to therapy not only significantly enhances patients' overall treatment experience but also contributes to the optimization of healthcare resource allocation and utilization.

While this study has shown promising results in the early efficacy prediction of neoadjuvant chemotherapy for breast cancer, there are several limitations to our research. Firstly, in this study, we considered the data before and after NAC to be of equal importance, and future work will consider the design of multiple loss functions to assign feature weights to different modalities at different stages, and explore the effect of multistage bimodal ultrasound images on NAC efficacy prediction in a deeper way. Secondly, it is currently focused on the multimodal image hierarchy in this study. Metabolomics is a technique that allows early disease detection, including tumors, by using body fluids and tissues to detect changes in small metabolic molecules. It is non-invasive, convenient, and easy to implement [48]. It has been shown that changes in small metabolic molecules can also predict the efficacy and prognosis of chemotherapy for tumors [49–52]. In the future, we will further consider combining metabolomics with medical imaging to implement a deep learning model based on multistage cross-modal early efficacy prediction of NAC for breast cancer. In addition, the proposed method is not fully automated, as it still requires manual image cropping from ultrasound videos for model training and testing. In the future, the development of a model for video-level data could be considered to comprehensively extract patient lesion features.

Conclusions

In summary, a novel BLTA-CNN model for predicting the efficacy of neoadjuvant chemotherapy (NAC) for breast cancer based on multistage and bimodal ultrasound images was proposed, which provides corresponding solutions to the current challenges of deep learning in this research field. The Bimodal Layer-wise Feature Fusion (BLFFM) connects the features between different data flow layer pairs and the features between different layers of the same data flow to achieve an efficient multimodal data feature-sharing mode. We further introduce the Temporal Hybrid Attention Module (THAM), it not only learns important features of the images, but also correlations between features of lesion changes during chemotherapy and reinforces the network's ability to understand key features of breast tumor progression. The rationality and effectiveness of the BLTA-CNN is verified by experiments on ultrasound image datasets of grayscale ultrasound (GUS) and color Doppler flow imaging (CDFI) data from pre-neoadjuvant chemotherapy (pre-NAC) and the first NAC (NAC1). It performs optimally in comparative experiments with eight leading deep learning classification models. It suggests the potential for early prediction of NAC outcomes based on multistage bimodal ultrasound images. To our knowledge, this is the first study to combine deep learning with multistage bimodal ultrasound imaging for early prediction the efficacy of Neoadjuvant chemotherapy in patients.

Acknowledgements
Not applicable.

Authors' contributions
Conception and design: JX, JZW and HCS. Collection and assembly of data: JSL, XQZ and CFW. Verification of the underlying data: JX, JZW and ZL. Development of methodology: JX, JZW and HCS. Data analysis and interpretation: JX, JZW, JSL, XQZ, CFW and ZL. Writing original draft: JX and JZW. Revised the final manuscript CFW, JSL and XQZ. All authors contributed to the article and approved the submitted version.

Funding
This research was supported by National Natural Science Foundation of China (no. 61873156, no.81801697, and no. 82103695), and Shanghai Jiao Tong University Medical interdisciplinary research Foundation [no. YG2023QNB09]. The authors are very grateful to the anonymous reviewers for their valuable comments and suggestions for improving the quality of the paper.

Data availability
The data analyzed in this study is subject to the following licenses/restrictions: Due to the privacy of patients, the related data cannot be available for public access. Requests to access these datasets should be directed to Caifeng Wan, wancaifengky@sina.com.

Declarations

Ethics approval and consent to participate
The studies involving human participants were reviewed and approved by Shanghai Jiao Tong University School of Medicine, Ren Ji Hospital Ethics Committee. Written informed consent was waived in this study.

Consent for publication
All authors agree to publication.

Competing interests
The authors declare no competing interests.

Received: 11 August 2024 Accepted: 19 December 2024
Published online: 23 January 2025

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. Ca-Cancer J Clin. 2020;70:7–30. https://doi.org/10.3322/caac.21590.
2. Sung H, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Ca-Cancer J Clin. 2021;71:209–49. https://doi.org/10.3322/caac.21492.
3. Gradishar WJ, et al. Breast Cancer, Version 3.2020. J Natl Compr Canc Ne. 2020;18:452–78. https://doi.org/10.6004/jnccn.2020.0016.
4. Kaufmann M, et al. Recommendations from an international consensus conference on the current status and future of neoadjuvant systemic therapy in primary breast cancer. Ann Surg Oncol. 2012;19:1508–16. https://doi.org/10.1245/s10434-011-2108-2.

5. Korde LA et al (2021) Neoadjuvant chemotherapy, endocrine therapy, and targeted therapy for breast cancer: ASCO Guideline. J Clin Oncol 39:1485-+. https://doi.org/10.1200/JCO.20.03399
6. Masood S (2016) Neoadjuvant chemotherapy in breast cancers.Women's health 12:480–491. https://doi.org/10.1177/1745505716677139
7. Chen X, et al. Gene expression patterns in human liver cancers. Mol Biol Cell. 2002;13:1929–39. https://doi.org/10.1091/mbc.02-02-0023.
8. Segal E, Friedman N, Kaminski N, Regev A, Koller D. From signatures to models: understanding cancer using microarrays. Nat Genet. 2005;37:S38–45. https://doi.org/10.1038/ng1561.
9. Liu SV, Melstrom L, Yao K, Russell CA, Sener SF. Neoadjuvant therapy for breast cancer. J Surg Oncol. 2010;101:283–91. https://doi.org/10.1002/jso.21446.
10. Hayashi M, Yamamoto Y, Iwase H (2020) Clinical imaging for the prediction of neoadjuvant chemotherapy response in breast cancer. Chin Clin Oncol 9:31. https://doi.org/10.21037/cco-20-15
11. Stavros AT,Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA (1995) Solid breast nodules: use of sonography to distinguish between benign and malignant lesions.Radiology 196:123–134. https://doi.org/10.1148/radiology.196.1.7784555
12. Hu YZ, et al. Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model. Med Phys. 2019;46:215–28. https://doi.org/10.1002/mp.13268.
13. Moon WK, Lee YW, Ke HH, Lee SH, Huang CS, Chang RF (2020) Computer -aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks.Comput Meth Prog Bio 190:105361. https://doi.org/10.1016/j.cmpb.2020.105361
14. Zhang EL, Seiler S, Chen ML, Lu WG, Gu XJ. BIRADS features-oriented semi-supervised deep learning for breast ultrasound computer-aided diagnosis. Phys Med Biol. 2020;65: 125005. https://doi.org/10.1088/1361-6560/ab7e7d.
15. Pi Y, et al. Automated diagnosis of multi -plane breast ultrasonography images using deep neural networks. Neurocomputing. 2020;403:371–82. https://doi.org/10.1016/j.neucom.2020.04.123.
16. Wu T, Sultan LR, Tian JW, Cary TW, Sehgal CM. Machine learning for diagnostic ultrasound of triple-negative breast cancer,". Breast Cancer Res Tr. 2019;173:365–73. https://doi.org/10.1007/s10549-018-4984-7.
17. Cheng HD, Shan J, Ju W, Guo YH, Zhang L. Automated breast cancer detection and classification using ultrasound images: A survey. Pattern Recogn. 2010;43:299–317. https://doi.org/10.1016/j.patcog.2009.05.012.
18. Xie J, et al. Dual-branch convolutional neural network based on ultrasound imaging in the early prediction of neoadjuvant chemotherapy response in patients with locally advanced breast cancer. Front Oncol. 2022;12: 812463. https://doi.org/10.3389/fonc.2022.812463.
19. Tong T, et al. Dual-input Transformer: An end-to-end model for preoperative assessment of pathological complete response to neoadjuvant chemotherapy in breast cancer ultrasonography. IEEE J Biomed Health. 2022;27:251–62. https://doi.org/10.1109/JBHI.2022.3216031.
20. Byra M, Dobruch-Sobczak K, Klimonda Z, Piotrzkowska-Wroblewska H, Litniewski J. Early prediction of response to neoadjuvant chemotherapy in breast cancer sonography using siamese convolutional neural networks. IEEE J Biomed Health. 2021;25:797–805. https://doi.org/10.1109/JBHI.2020.3008040.
21. Gu JH et al (2022) Deep learning radiomics of ultrasonography can predict response to neoadjuvant chemotherapy in breast cancer at an early stage of treatment: a prospective study.Eur Radiol 32:2099–2109. https://doi.org/10.1007/s00330-021-08293-y
22. El Adoui M, Drisis S, Benjelloun M. Multi-input deep learning architecture for predicting breast tumor response to chemotherapy using quantitative MR images. Int J Compu Assist Radiol Surg. 2020;15:1491–500. https://doi.org/10.1007/s11548-020-02209-9.
23. Xi IL, et al. Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging. Clin Cancer Res. 2020;26:1944–52. https://doi.org/10.1158/1078-0432.CCR-19-0374.
24. Le MH, et al. Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks. Phys Med Biol. 2017;62:6497–514. https://doi.org/10.1088/1361-6560/aa7731.
25. Nie D, et al. Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. Sci Rep. 2019;9:1103. https://doi.org/10.1038/s41598-018-37387-9.
26. Joo S et al (2021) Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer.Sci Rep 11:18800. https://doi.org/10.1038/s41598-021-98408-8
27. Woo S, Park J, Lee JY, Kweon IS (2018) CBAM: convolutional block attention module. In: The Proceedings of the European conference on computer vision, pp 3–19. https://doi.org/10.1007/978-3-030-01234-2_1
28. Dolz J, Gopinath K, Yuan J, Lombaert H, Desrosiers C, Ben Ayed I. HyperDense-Net: A hyper-densely connected cnn for multi-modal image segmentation. IEEE Trans Med Imaging. 2019;38:1116–26. https://doi.org/10.1109/TMI.2018.2878669.
29. Dolz J, Desrosiers C, Ben Ayed I (2018) IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet. https://doi.org/10.48550/arXiv.1811.08305
30. Chen LL, Wu Y, Dsouza AM, Abidin AZ, Wismuller A, Xu CL, (2018) MRI tumor segmentation with densely connected 3D cnn. In: The Proceedings of the Conference on medical imaging - image processing. https://doi.org/10.1117/12.2293394
31. Kingma DP, Ba J (2015) Adam: A Method for Stochastic Optimization. In: The Proceedings of the Conference for learning representations. https://doi.org/10.48550/arXiv.1412.6980
32. Gao Y, Ma JY, Zhao MB, Liu W, Yuille AL, Soc IC (2019) NDDR-CNN: Layer-wise feature fusing in multi-task cnns by neural discriminative dimensionality reduction.In: The Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3200–3209. https://doi.org/10.1109/CVPR.2019.00332
33. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: The Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141. https://doi.org/10.1109/CVPR.2018.00745
34. Park J, Woo S, Lee JY, Kweon IS (2018) BAM: Bottleneck Attention Module. https://doi.org/10.48550/arXiv.1807.06514
35. Fu J et al (2019) Dual attention network for scene segmentation. In: The Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3141–3149. https://doi.org/10.1109/CVPR.2019.00326
36. Hou QB, Zhou DQ, Feng JS (2021) Coordinate attention for efficient mobile network design. In: The Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 13708–13717. https://doi.org/10.1109/CVPR46437.2021.01350
37. Tan MX, Le QV (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. In: The Proceedings of the International conference on machine learning, pp 10691–10700. https://doi.org/10.48550/arXiv.1905.11946
38. Huang G, Liu Z, L. van der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: The Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2261–2269. https://doi.org/10.1109/CVPR.2017.243
39. Ma NN, Zhang XY, Zheng HT, Sun J (2018) ShuffleNet V2: practical guidelines for efficient cnn architecture design. In: The Proceedings of the European conference on computer vision, pp 122–138. https://doi.org/10.1007/978-3-030-01264-9_8
40. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: The Proceedings of the IEEE conference on computer vision and pattern recognition, pp, 1800–1807. https://doi.org/10.1109/CVPR.2017.195
41. Sandler M, Howard A, Zhu ML, Zhmoginov A, Chen LC (2018) MobileNetV2: Inverted residuals and linear bottlenecks. In: The Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4510–4520. https://doi.org/10.1109/CVPR.2018.00474
42. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: The Proceedings of the AAAI Conference on Artificial Intelligence, pp. 4278–4284. https://doi.org/10.48550/arXiv.1602.07261
43. Xie SN, Girshick R, Dollar P, Tu ZW, He KM (2017) Aggregated residual transformations for deep neural networks. In: The Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5987–5995. https://doi.org/10.1109/CVPR.2017.634
44. Wu BT, Sun XW, Hu LJ, Wang YZ (2019) Learning with unsure data for medical image diagnosis. In: The Proceedings of the IEEE conference on computer vision and pattern recognition, pp 10589–10598. https://doi.org/10.1109/ICCV.2019.01069
45. Faes L, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a

feasibility study. Lancet Digit Health. 2019;1:E232–42. https://doi.org/10.1016/S2589-7500(19)30108-6.
46. Keune JD, Jeffe DB, Schootman M, Hoffman A, Gillanders WE, Aft RL. Accuracy of ultrasonography and mammography in predicting pathologic response after neoadjuvant chemotherapy for breast cancer. Am J Surg. 2010;199:477–84. https://doi.org/10.1016/j.amjsurg.2009.03.012.
47. Shia WC, Huang YL, Wu HK, Chen DR. Using flow characteristics in three-dimensional power doppler ultrasound imaging to predict complete responses in patients undergoing neoadjuvant chemotherapy. J Ultrasound Med. 2017;36:887–900. https://doi.org/10.7863/ultra.16.02078.
48. Bartkowiak K, et al. Circulating cellular communication network factor 1 protein as a sensitive liquid biopsy marker for early detection of breast cancer. Clin Chem. 2022;68:344–53. https://doi.org/10.1093/clinchem/hvab153.
49. Wei SW, et al. Metabolomics approach for predicting response to neoadjuvant chemotherapy for breast cancer. Mol Oncol. 2013;7:297–307. https://doi.org/10.1016/j.molonc.2012.10.003.
50. Vignoli A, et al. Effect of estrogen receptor status on circulatory immune and metabolomics profiles of HER2-positive breast cancer patients enrolled for neoadjuvant targeted chemotherapy. Cancers. 2020;12:314. https://doi.org/10.3390/metabo13020296.
51. Zidi O, et al. Fecal metabolic profiling of breast cancer patients during neoadjuvant chemotherapy reveals potential biomarkers. Molecules. 2021;26:2266. https://doi.org/10.3390/molecules26082266.
52. Debik J, et al. Assessing treatment response and prognosis by serum and tissue metabolomics in breast cancer patients. J Proteome Res. 2019;18:3649–60. https://doi.org/10.1021/acs.jproteome.9b00316.

Publisher's Note