Computers in Biology and Medicine 157 (2023) 106765

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Multimodal adversarial representation learning for breast cancer prognosis prediction

Xiuquan Du [a,b,*], Yuefan Zhao [b]

[a] Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei, China
[b] School of Computer Science and Technology, Anhui University, Hefei, China

ARTICLE INFO

Keywords:
Breast cancer prognosis prediction
Adversarial representation learning
Multimodal data fusion
Bilinear convolutional neural network
Ensemble learning

ABSTRACT

With the increasing incidence of breast cancer, accurate prognosis prediction of breast cancer patients is a key issue in current cancer research, and it is also of great significance for patients' psychological rehabilitation and assisting clinical decision-making. Many studies that integrate data from different heterogeneous modalities such as gene expression profile, clinical data, and copy number alteration, have achieved greater success than those with only one modality in prognostic prediction. However, many of these approaches that exist fail to dramatically reduce the modality gap by aligning multimodal distributions. Therefore, it is crucial to develop a method that fully considers a modality-invariant embedding space to effectively integrate multimodal data. In this study, to reduce the modality gap, we propose a multimodal data adversarial representation framework (MDAR) to reduce the modal heterogeneity by translating source modalities into distributions for the target modality. Additionally, we apply reconstruction and classification losses to embedding space to further constrain it. Then, we design a multi-scale bilinear convolutional neural network (MS-B-CNN) for uni-modality to improve the feature expression ability. In addition, the embedding space generates predictions as stacked feature inputs to the extremely randomized trees classifier. With 10-fold cross-validation, our results show that the proposed adversarial representation learning improves prognostic performance. A comparative study of this method and other existing methods on the METABRIC (1980 patients) dataset showed that Matthews correlation coefficient (Mcc) was significantly enhanced by 7.4% in the prognosis prediction of breast cancer patients.

1. Introduction

According to data released by the International Agency for Research on Cancer, breast cancer has surpassed lung cancer to become the world's largest cancer in 2020, accounting for 11.7% of all new cancer patients [1]. According to the American Cancer Society, there are approximately 685,000 cancer deaths and 2.3 million breast cancer cases diagnosed per year [1]. Therefore, accurate prognosis prediction is of great significance for the psychological recovery of breast cancer patients and for guiding clinicians to develop appropriate treatment plans [2]. However, the complexity of breast cancer and the significant differences in its clinical outcomes make it extremely difficult to be predicted and treated [3]. In addition, medical professionals find it difficult to manually interpret multimodal data because of their high dimensionality [4]. Given this situation, there is a pressing need to develop computer-based methods to provide an efficient and accurate prognosis [5,6].

There have been lots of studies to predict the prognosis for breast cancer, which classify women with breast cancer in a good group or a poor prognosis group [7–9]. Recent studies have successfully predicted the prognosis of women with breast cancer patients through the application of statistical and machine-learning methods to single-module data, in particular data on gene expression [10–13]. Such as [14] identified 70 genes prognostic signatures by performing a multivariate analysis of gene expression data. [2] used the support vector machine (SVM) to extract the most important features in gene expression data for predicting breast cancer prognosis. But the above studies are intended to work with a handful of gene expression profile data, still giving way to improvement, although with high performance [15–18]. In particular, next-generation sequencing technologies are generating a great deal of multimodal data. As shown in Fig. 1(a), data from multiple sources are used for prognosis predictions of patients. These data provide widely available information for the prognosis of cancer.

* Corresponding author at: Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei, China.
E-mail address: dxqllp@163.com (X. Du).

https://doi.org/10.1016/j.compbiomed.2023.106765
Received 9 August 2022; Received in revised form 27 February 2023; Accepted 7 March 2023
Available online 15 March 2023

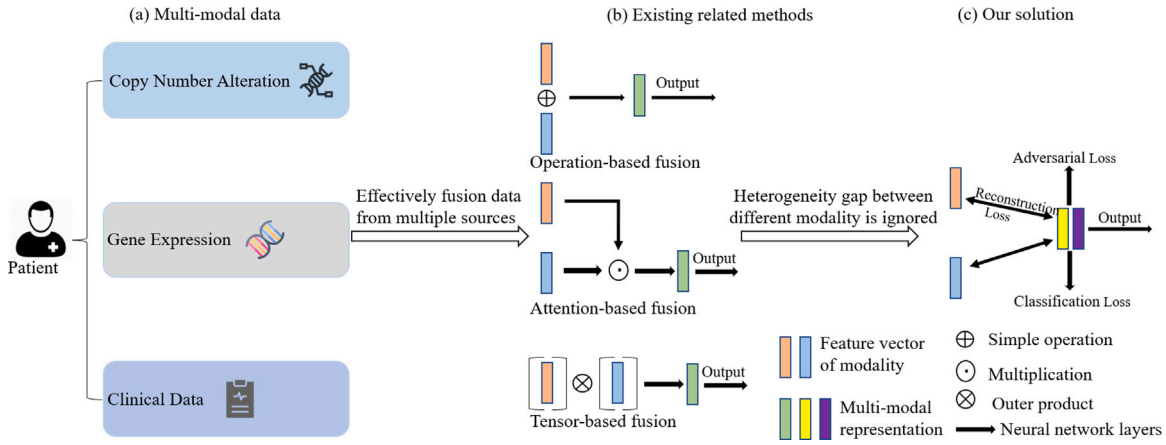

Fig. 1. Illustration of related strategies and challenges of multimodal data fusion.

Apart from the success of the above approaches, researchers started integrating multi-modal data in the field of predicting breast cancer prognosis. For example, [19] identified the signatures of gene expression profiles for predicting low-risk and high-risk in glioblastoma (GBM) patients. In Ref. [20], proposed a multiple kernel machine learning method by integrating histopathological images and multi-omics data for GBM prognosis prediction. [21,22] developed a framework to integrate multi-omics data by denoising autoencoder for cancer prognosis accurate prediction. TAN [23] proposed a Tree-Augmented Bayesian Belief Network (TAN)-based analytics methodology to identify the factors affecting cancer progression. The method consists of four steps: data acquisition and preprocessing, variable selection via Genetic Algorithm, data balancing with synthetic minority over-sampling and random undersampling methods, and finally the development of TAN model to determine the probabilistic inter-conditional dependency structure among breast cancer-related variables along with the posterior survival probabilities. Multi-PEN [24] proposed a deep learning model using multi-omics and multi-modal schemes, namely the Multi-Prognosis Estimation Network. When using Multi-PEN, gene attention layers are employed for each datatype, thereby allowing us to identify prognostic genes. Additionally, recent developments in deep learning, such as residual learning and layer normalization, are utilized. Not surprisingly, good work is obtained while several modalities are taken into account, which are various features that are widely used to predict cancer prognosis. Unfortunately, most of these methods associate various types of data directly in model generation and do not account for the fact that the features of the different modalities may be represented differently [25].

Recent developments in deep learning approaches have shown that models with multiple forms of input data sources perform better than models with a single input data source. In addition, considering that methods using only one source of information often come with limitations such as lack of non-versatility, data noise, and unique, multi-modal learning is offered to resolve these issues by associating relevant information from multiple sources to make final decisions [26]. This fact was confirmed by several studies on the prognosis for breast cancer patients, which are based on multimodal data. As shown in Table 1, our literature survey brief summarizes the feature fusion strategies and fusion methods of the relevant works. For instance, using decision fusion to integrate multimodal data, [7] developed a multimodal deep neural network for the first time. In Ref. [4], proposed an unsupervised encoder for compressing mRNA expression data, clinical data, microRNA expression data, and histopathology whole slide images (WSI) into single feature vectors for patients, a prognosis prediction model was derived from the feature vectors. In Ref. [27], developed a STACKED_RF method, which combines random forest with a stacked integrated framework, to analyze multimodal data. In [8], proposed a gated attention deep learning model stacked with random forest. Based on these results, more accurate results can be achieved with multimodal data. Unfortunately, despite many efforts to integrate multimodal data for cancer prognosis prediction, which is shown in Fig. 1(b). It remains a challenging task. A major problem with multi-modal fusion is the distribution of heterogeneous data from different modalities [28,29], leading to difficulties in extracting additional information through modalities that are essential for an overall interpretation of multi-modal information. Most of the previous work does not put effort into learning the modality-invariant embedding space for various modalities to match multi-modal distributions. On the contrary, a subnetwork to each modality and then proceed with the merger immediately [8,30]. Consequently, the modality deviation as usual heavily affects the effect of the merge.

In all the methods, Generative Adversarial Networks (GANs) [34] have achieved significant advancements, using adversarial training, GANs can map one distribution to another [35]. In light of its unique characteristics, adversarial training is adapted to the translation of the distribution of modalities. The novel and powerful architecture of "Modality to Modality Translation (ARGF)" for sentiment analysis [36] motivated us to develop a multimodal data adversarial representation framework to predict breast cancer prognosis, which is shown in Fig. 1(c). Specifically, similar to the previous works by [37–40], a feature projector performs the main task of representation learning, namely, which generate an invariant representation of modalities for items from different modalities in the common subspace. Its purpose is to confuse a modality classifier that acts as an adversary. The modality classifier tries to distinguish the items according to their modalities and in this way guides the learning of the feature projector. By introducing the modality classifier into the adversarial training, the alignment of the across modalities representation distribution can be better achieved, and the modality invariance can be obtained more effectively. The representation subspace being optimized for cross-modal associates will then result through the convergence of this process, namely when the modality classifier "fails". Furthermore, a classifier is learning to classify the encoded representations into the correct label and preserve the underlying cross-modal semantic structure in data. In this way, it can ensure that the embedding space is discriminative for predicting tasks. Moreover, a decoder is also defined for each modality to avoid the loss of unimodal information. Additionally, considering that genes do not play alone, a feature projector model based on multi-scale bilinear convolutional neural network was proposed. To obtain better feature expression ability, based on B-CNN [41], we perform a multi-scale fusion of features of different convolutional layers. We fuse shallow features and deep features, which are bound to enhance the expression ability of features. That is, to obtain richer hidden features, simulate the joint action of multiple features to a greater extent. Finally,

Table 1
The overall information of METABRIC breast cancer dataset.

Study	Fusion strategy				Fusion details
	Addition	Subspace	Attention	Tensor fusion	
Cheerla and Gevaert [4]	✓	✓			The average of learned uni-modal features, while a margin-based hinge-loss was used to regularize the similarity of learned uni-modal features.
Huang et al. [31]	✓				Concatenation of genomic biomarkers and the learned co-expression features
Arya and Saha [27]	✓				Concatenation of learned uni-modal features
Arya and Saha [8]	✓		✓		Concatenation of learned uni-modal features with Bi-Modal Attention
Guo et al. [9]	✓		✓		Affinity fusion module concatenates attention-based unimodal features
Tong et al. [32]		✓			Shared features of learned uni-modal features, regular-ized by a similarity loss
Wang et al. [33]	✓			✓	Inter-modal features and intra-modal features produced by the bilinear layers

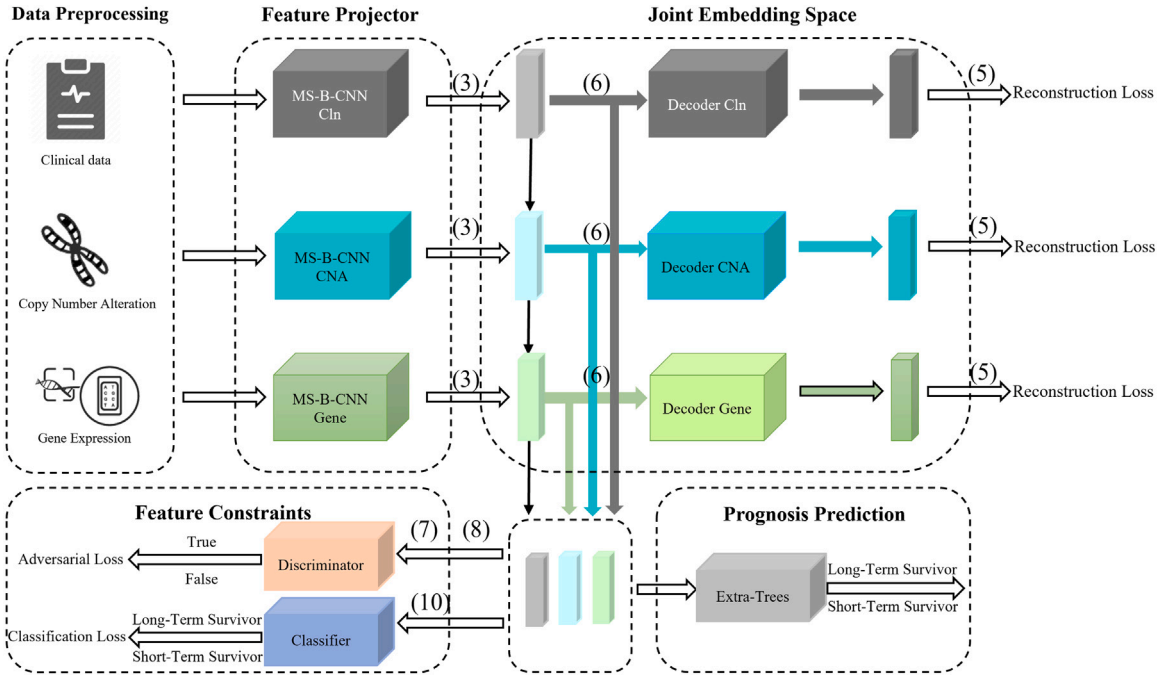

Fig. 2. The overall process of MDAR for the breast cancer prognosis prediction.

considering that the dataset is relatively small and the positive and negative samples are unbalanced, the ensemble learning framework is used to overcome this difficulty. Therefore, we use adversary training for extracting the modality-invariant embedding space and these features feed to the stacked layer, which itself is a machine learning method like the extremely randomized trees. In brief, the main contributions are listed below:

- We propose MDAR for multimodal fusion to address a key problem in multimodal fusion, namely the distribution of heterogeneous data from various modalities, and better fusion of multimodal data.
- We propose a multi-scale bilinear network module with a multi-scale fusion of features of different convolutional layers to enhance the feature expression ability.
- We validate the effectiveness of MDAR on two exposed datasets. The experimental results show that MDAR performs better compared with existing research methods to the best of our knowledge.

The rest of this paper is structured as follows: Section 2 provides the details of our proposed method. Furthermore, the datasets and experimental design are described in Section 3, the results show in Section 4, and some conclusions are drawn in Section 5.

2. Methods

2.1. Framework of our proposed method

In this section, we present the framework of MDAR that is designed to predict the prognosis of breast cancer patients. MDAR aims to distinguish between poor and good samples, each sample was labeled as a good sample if the patient survive more than 5 years, and labeled as a poor sample if the patient did not survive more than 5 years. Fig. 2 shows an overview of MDAR. For the generative model, MS-B-CNN is used to generate a modality-invariant embedding space, the decoder generates a reconstruction representation to maintain the original information of each modality. For the discriminative model, they can be divided into classifier and discriminator. The generative

Table 2
Detailed parameter configuration of MS-B-CNN.

# of convolution layers	3
Filter size of convolution layer	10
Stride sizes	2, 3
Stride size in convolutional layer	2
Padding in convolution layer	Same
Activate function	ReLU
# of hidden layers	1
Mini-batch size	16
Training Epoch	50
Activation function	TANH
Loss function	Binary cross-entropy +L2 regularization

model and discriminative model are trained in an adversarial manner. For the prognosis prediction, we use an ensemble learning approach to classify extracted modality-invariant embedding space features using extreme random trees.

2.2. Multi-scale bilinear convolutional neural network

Multiscale feature fusion of convolutional neural networks (CNN) has important applications in deep learning. For CNN, shallow structures and deep structures learn different semantic features, if these features are fused, it is bound to enhance the expression ability of features. For example, an important feature of ResNet [42] is the fusion feature, in which the "Skip Connection" structure is equivalent to the parallel connection of multiple residual elements to achieve the fusion of different scale features. DenseNet [43] is also enlightened by multi-scale fusion, with a larger span of fusion between different levels and the integration between channels. Inspired by the above two network structures and B-CNN [41], this study by using the output of the last layer of convolution and the output of the convolutional kernel of the previous convolutional block as the inner product, as shown in Fig. 3. Two sub-networks in the form of CNN are input from a single data, each sub-network is passed through a convolutional layer with a certain number of filters or kernels, and the final generated feature map is obtained by combining the features of multiple convolution kernels of different scales. Correspondingly multiplied and then added. We initialize the weights of each layer using the normalizations suggested by [44] and bias is initialized using small numbers. We use the truncated normal distribution to initialize the weights between layers, which is defined by the following formula:

$$W \sim T\left[-\sqrt{\frac{2}{n_i + n_o}}, \sqrt{\frac{2}{n_i + n_o}}\right] \tag{1}$$

Where n_i, n_0 means the number of unit inputs and outputs, respectively. We define cross-entropy loss as the objective function of the MS-B-CNN module in the final output layer. Additionally, to avoid overfitting of the deep learning model, L2 regularization is also added to the network, which is a popularly used regularization technique in deep learning research. The loss function is defined as follows:

$$\begin{aligned} \mathrm{loss}(X, y, \hat{y}) = &- \frac{1}{N}\sum_{i}^{n}[y\log(\hat{y}) + (1-y)\log(1-\hat{y})] \\ &+ \frac{1}{\lambda}\sum_{l=1}^{L}\|W_l\|_2 \end{aligned} \tag{2}$$

Where y is the label, $\hat{y}$ is the prediction result, and λ is a hyperparameter.

Lastly, our MS-B-CNN module has one input layer and three convolutional layers, two pooled layers, one flatten layer, and one fully connected dense layer on each sub-network. For gene expression profile and copy number alteration (CNA), our network utilizes an additional dropout layer of 0.5 dropout [45]. The network hyper-parameters have been optimized by adjusting various values, and the optimal value is selected in the network with the best Area Under Curve (AUC) value. Detailed configurations of the MS-B-CNN parameters are indicated in Table 2. In addition, to deal with the local minimum problem, for example, [46–48] attempts to solve the local minima problem using the momentum method. We choose the adaptive moment estimation (Adam) method to optimize the CNN training, which is still essentially an adaptive constraint on the learning rate, with the advantage that after bias correction, the learning rate has a determined range for each iteration, making the parameters relatively smooth and suitable for high-dimensional data training.

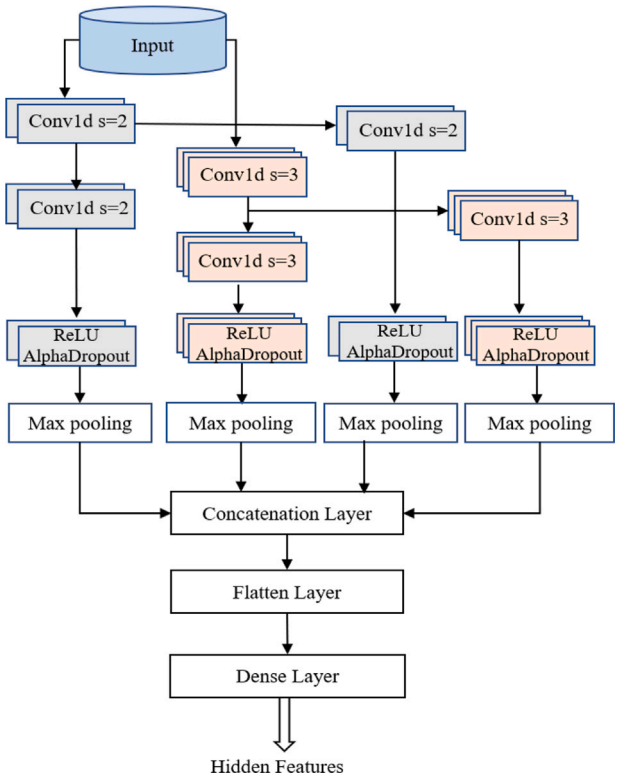

Fig. 3. MS-B-CNN as feature projector for uni-modal.

2.3. Adversarial representation learning for multimodal

Here we propose an efficient approach to translating source modality distributions to target modality distributions to learn a modality invariant embedding space. Statistical properties that differ between modalities could significantly impede cross-modal associations, which is why matching distributions in embedding space are critical.

Under the guidance of game adversarial thinking, GANs [34] can learn features that are difficult to compare with general depth models. For cross-modal feature associations, it requires that the associated features have semantic discrimination and consistency with modalities. Through decoders and classifiers, the associated features have discriminative semantics and are consistent with modalities. Define a decoder for each modality, reconstructing the original features to prevent the loss of single modality information. In addition, a classifier is constructed to classify the feature projector representations to correct categories, ensuring that the modality-invariant space is discriminating against the prediction task.

2.3.1. Generative model

Supposed we have three modalities as input: gene expression profile $x_{exp} \in R^{d_{exp}}$, CNA data $x_{cna} \in R^{d_{cna}}$ and clinical data $x_{cln} \in R^{d_{cln}}$ with d_{exp} being the dimensionality of x_{exp} and so on. Suppose x_{cna} is the

target modality, other modalities are called source modality, and $p(x_{cln})$ represents the original feature distribution of the clinical data modality. In the same way as [49], our transformation distribution for these three modalities is:

$$p_{\theta_{cln}}(x^e_{cln}) = \int_{x_{cln}} q(x^e_{cln}|x_{cln},\theta_{cln})p(x_{cln})dx_{cln}$$
$$p_{\theta_{cna}}(x^e_{cna}) = \int_{x_{cna}} q(x^e_{cna}|x_{cna},\theta_{cna})p(x_{cna})dx_{cna} \quad (3)$$
$$p_{\theta_{exp}}(x^e_{exp}) = \int_{x_{exp}} q(x^e_{exp}|x_{exp},\theta_{exp})p(x_{exp})dx_{exp}$$

Where $q(x^e_{cln}|x_{cln},\theta_{cln})$ is called the feature projector function of the clinical data, θ_{cln} is the parameter, and $p_{\theta_{cln}}(x^e_{cln})$ is the clinical data distribution of the transformation in the learning embedding space limited by θ_{cln}. Our goal is to explicitly map the converted distributions $p_{\theta_{cln}}(x^e_{cln})$ and $p_{\theta_{exp}}(x^e_{exp})$ to $p_{\theta_{cna}}(x^e_{cna})$, by optimizing θ_{cln}, θ_{cna}, θ_{exp}.

The transform distribution method may lose unimodal information necessary to extract complementary information between modalities. To preserve modality-specific information in the learning embedding space, we define the decoder as follows:

$$p_{\theta_{dcln}}(\tilde{x}_{cln}) = \int_{x^e_{cln}} q(\tilde{x}_{cln}|x^e_{cln},\theta_{d_{cln}})p_{\theta_{cln}}(x^e_{cln})dx^e_{cln}$$
$$p_{\theta_{dcna}}(\tilde{x}_{cna}) = \int_{x^e_{cna}} q(\tilde{x}_{cna}|x^e_{cna},\theta_{d_{cna}})p_{\theta_{cna}}(x^e_{cna})dx^e_{cna} \quad (4)$$
$$p_{\theta_{dexp}}(\tilde{x}_{exp}) = \int_{x^e_{exp}} q(\tilde{x}_{exp}|x^e_{exp},\theta_{d_{exp}})p_{\theta_{exp}}(x^e_{exp})dx^e_{exp}$$

Where $q(\tilde{x}_{cln}|x^e_{cln},\theta_{dcln})$ is the decoder of the clinical data, where θ_{dcln} is the parameter and $p_{\theta_{dcln}}(\tilde{x}_{cln})$ is the distribution of the reconstructed representation of the clinical data. Given the feature input $x_f = \{x_{cln}, x_{cna}, x_{exp}\}$ for the encoder, we want the decoder reconstruction output $\tilde{x}_f = \{\tilde{x}_{cln}, \tilde{x}_{cna}, \tilde{x}_{exp}\}$ approximate x_f to minimize information loss. To this end, we define reconstruction losses as:

$$\mathcal{L}_{rl}(\tilde{x}_f, x_f) = \sum_m \|\tilde{x}_m - x_m\|_2, m \in \{cln, cna, exp\} \quad (5)$$

By minimizing $\mathcal{L}_{rl}$, the single-omics information can be preserved for further fusion by encoding representation.

2.3.2. Discriminative model

However, the distribution of different modalities is very complex, matching their properties is extremely difficult with just a simple feature projector network. So, we add constraints to the transformed distribution by using adversarial training. Specifically, a discriminator D is defined to classify $p_{\theta_{cln}}(x^e_{cln})$ and $p_{\theta_{exp}}(x^e_{exp})$ as false, but $p_{\theta_{cna}}(x^e_{cna})$ as true, and the generator (encoder E_{cln}, E_{exp}) attempts to trick discriminator D into classifying $p_{\theta_{cln}}(x^e_{cln})$ and $p_{\theta_{exp}}(x^e_{exp})$ as true. In min–max games, generators and discriminators compete to learn modality-invariant embedding spaces. Here we can divide the loss function into two parts: false adversarial loss $\mathcal{L}_{fal}$ and true adversarial loss $\mathcal{L}_{tal}$, as shown below:

$$\mathcal{L}_{al} = \mathcal{L}_{fal}(x^e_{cln}, x^e_{exp}) + \mathcal{L}_{tal}(x^e_{cna}, x^e_{cln}, x^e_{exp}) \quad (6)$$

E_{cln}, E_{exp} try to deceive D, resulting in a false adversarial loss $\mathcal{L}_{fal}$ while D aimed at determining that the distribution of the target modality is correct and that the distribution of other modalities is incorrect, resulting in a true adversarial loss $\mathcal{L}_{tal}$. Specifically, we define $\mathcal{L}_{fal}$ and $\mathcal{L}_{tal}$ as:

$$\mathcal{L}_{fal} = -w[\log(D(x^e_{cln})) + \log(D(x^e_{exp}))] \quad (7)$$

$$\mathcal{L}_{tal} = -w[\log(1 - D(x^e_{cln})) + \log(1 - D(x^e_{exp})) + \log(D(x^e_{cna}))] \quad (8)$$

where $D(x^e_{cln})$ represents the predicted distribution value of x^e_{cln}, with a range of 0 to 1, and w is the adversarial loss of learning weight. If the discriminator cannot distinguish the target modality (i.e., $D(x^e_{cln}) \approx D(x^e_{exp}) \approx D(x^e_{cna})$) from all modalities, once that is done, the distributions of the different modalities can be successfully mapped within a modality-invariant embedding space. The adversarial training strategy imposes limits on the statistical properties represented by the feature projector. As a further step towards enhancing the discrimination of the learned embedding space regarding the learning task, we define a classifier that takes the feature projector representation of each modality as input. A classifier is defined as follows:

$$\tilde{y}_m = C(x^e_m; \Theta_c), m \in \{cln, cna, exp\} \quad (9)$$

where C represents classifiers, $\tilde{y}_m$ is a predicted label based on feature representation of modality m. For the purpose of minimizing prediction error, we define categorical losses as follows:

$$\mathcal{L}_{cl}(\tilde{y}, y) = \sum_m \|\tilde{y}_m - y\|_2 \quad (10)$$

where y is the label. Using classification loss, feature projector representations can carry label information, and so the embedding space becomes discriminative for prediction tasks.

In summary, the loss function for generating model and discriminative model can be defined as follows:

$$\mathcal{L} = [\lambda\mathcal{L}_{fal} + (1-\lambda)\cdot\mathcal{L}_{rl}] + 0.5\mathcal{L}_{tal} + \mathcal{L}_{cl} \quad (11)$$

where λ is a hyper-parameter that decides the importance of the loss function, the value of which is determined by a network search.

2.3.3. Cross-modal adversarial training procedure

The overall training procedure is presented in Algorithm 1. During a gradient update, first, we update the feature projector and decoder using $\mathcal{L}_{fal}$ and $\mathcal{L}_{rl}$; second, we use the $\mathcal{L}_{tal}$ to update the discriminator to advance its discriminative against false/true distributions; and finally, we update the encoder and classifier with $\mathcal{L}_{cl}$ to enhance the discrimination of the modality-invariant embedding space.

2.4. Stacked-based prediction model

The ensemble learning methods have gained popularity because of their superior prediction performance in practice. In practice, the performance of a learner will depend on the sample size, dimensionality, and bias–variance tradeoff of the model. Thus, with finite sample datasets and prediction problems, it is usually impossible to know a priori which learner performs best. Ensemble learning works by computing the best convex combination of the underlying learners, rather than selecting an algorithm. In this study, our motivation for using an ensemble method comes from the fact that it performs well on unbalanced data and that the dataset available for the current study is unbalanced. At present, ensemble deep learning has been successfully applied in biological sequence studies, such as prediction of antifungal peptides (iAFPs-EnC-GA) [50], prediction of antitubercular peptides (iAtbP-Hyb-EnC) [51], among many others. To take advantage of the idea of ensemble learning [52], we stack modality-invariant embedding spaces to form a stacked representation. Consider the heterogeneity that exists between different data modalities, as well as the powerful ability of adversarial training to match modal distributions. We extract the modality-invariant embedding space features from the generative model from each modality and combine all these features to form the stacked features. To overcome the data imbalance, the feature stacking of modalities is used for the final breast cancer prognosis prediction using Extra-Trees after obtaining the stacked features.

3. Experimental design

3.1. Datasets

For this study, we utilized two separate datasets of breast cancer samples with a total of 3036 samples. From the cBioPortal [53], we

Algorithm 1 Pseudocode of optimizing our MDAR

Require: Multi-modal Datasets $D = \{(X_i = (X_{cln}, X_{cna}, X_{exp}), y_i)\}$
Require: batch size: N, hyperparameter: λ, Initialize modality-specific feature projectors as $E_{cln}, E_{cna}, E_{exp}$ Initialize modality-specific decoders as $D_{cln}, D_{cna}, D_{exp}$
while Not Converged **do**
  Apply $\pounds_{fal}$ and $\pounds_{rl}$ update feature projectors and decoders
  Sample a batch of samples from D
  $loss \leftarrow 0$
  for each sample i in the batch **do**
    Compute decoder as $\tilde{x}_f = Dec_f(E_f(X_f))$, $f \in \{cln, cna, exp\}$
    Compute reconstruction loss by Equation (5)
    Compute fake adversarial loss by Equation (7)
    $loss \leftarrow loss + \pounds_{rl} + \pounds_{fal}$
  end for
  Apply $\pounds_{tal}$ to update the discriminator to learn modality-invariant embedding space
  Sample a batch of samples from D
  $loss \leftarrow 0$
  for each sample i in the batch **do**
    Compute true adversarial loss by Equation (8)
    $loss \leftarrow loss + \pounds_{tal}$
  end for
  Apply $\pounds_{cl}$ to improve the discrimination ability of modality-invariant embedding space
  Sample a batch of samples from D
  $loss \leftarrow 0$
  for each sample i in the batch **do**
    Compute modality-specific feature projectors $E_{cln}, E_{cna}, E_{exp}$
    Compute classification loss $\pounds_{cl}$ using y_i and prediction $\tilde{y}_i$ by Equation (10)
    $loss \leftarrow loss + \pounds_{cl}$
  end for
end while
Output: learned modality-specific $E_{cln}, E_{cna}, E_{exp}$

downloaded the METABRIC dataset. There are three sub-dataset, which are gene expression profile, CNA data, and clinical data. Similar to previous work by Khademi and Nedialkov [54], each sample was classified as either a "good" or "poor" sample. A "good" sample was assigned a label of 1 if the patient survived more than 5 years, while a "poor" sample was assigned a label of 0 if the patient survived less than 5 years. In the METABRIC dataset, the median age of patients at diagnosis was 61 years and the mean survival time was 125.1 months. As shown in Table 3, the dataset is described in general terms. Each patient's clinical profile contains 27 features such as age at diagnosis, size of the tumor, grade, etc.

3.2. Data preprocessing

As for pre-processing the data, according to [7], gene expression profiles and CNA profiles were estimated using weighted nearest neighbors algorithms for missing values, three categories of gene expression features are identified after normalizing the data: under-expressed (−1), over-expressed (1), and baseline (0). For CNA features, the raw data with five discrete values (2, 1, 0, 1, 2) were directly utilized. 25 clinical features have been selected for the final data set of clinical features and normalized in the range of [0,1] by the min–max normalization algorithm. Despite the relatively small number of samples in this dataset, omics datasets usually have tens of thousands of features. Consequently, prognosis prediction of cancer often falls victim to the "curse of dimensionality". The high dimensionality and small sample size of the dataset may lead to bad results for deep learning methods. To alleviate the dimensionality challenges, researchers often apply feature selection or downscaling techniques to raw features [31]. Hence, [7] have used the well-known feature selection algorithm mRMR [55] to perform feature selection on the METABRIC dataset. The AUC value was used in the study as a criterion to evaluate the performance of the features. In the end, 400 genes from gene expression profile data, 200 genes from CNA profile data with the highest AUC value, and 25 clinical features from clinical data are chosen as features for prognosis prediction. The details of the features used in this study are shown in Table 4.

Table 3
The overall information of METABRIC breast cancer dataset.

Cut-off(years)	5
Total population	1980
Long time survivors	1489
Short time survivors	491
Median age at the time of diagnosis	61
Average survival(months)	125.1

Table 4
The properties of the dataset.

Data category	Number	Feature number
Clinical	27	25
Gene expression	24368	400
Copy number	26298	200

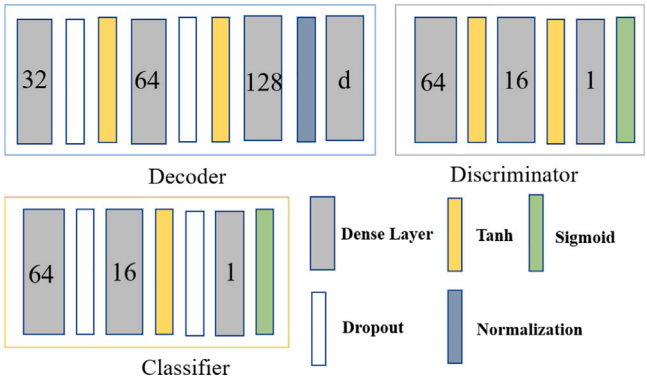

Fig. 4. Diagram of decoder, discriminator, and classifier.

3.3. Experimental details

We using Keras implemented our model on the Nvidia Tesla P100 server. We use the Mean Squared Error as the loss function for adversarial training and Adam [56] as the optimizer. The implementation details of the decoder, discriminator, and classifier are shown in Fig. 4. Specifically, for the METABRIC dataset, the decoder of the generative model uses a feed-forward neural network to reconstruct the modality-invariant embedding space into the original features of each modality in a nonlinear manner. The numbers in the dense layer are the output dimensions, and the dropout rate is 0.3. Where d represents the dimension of the input feature vector for each modality. For the extracted modality-invariant embedding representation, the discriminator is composed of a fully connected layer, and each modality is distinguished by a modality-invariant embedding representation, which is labeled 1 for CNA features and 0 for gene expression and clinical data features. The classifier consists of three fully connected layers. The first layer has 64 hidden units, followed by Dropout and the Activation Layer, and features are fed to the sigmoid layer for label prediction.

3.4. Evaluation metrics

In this section, to fully evaluate our proposed approach, we used a ten-fold cross-validation experiment, similar to previous existing

cancer prognosis prediction studies. In our experiment, the patients were randomly divided into ten groups. Nine of these ten subsets are divided further into training sets (80%) and validation sets (20%), with the remaining subsets serving as test sets. After ten rounds, we can calculate the prediction scores for each subset of the test and combine them into an overall prediction score. We define our results using the confusion matrix generated by the binary classification of the test data by the model. The reason for using the confusion matrix is that it accurately captures the agreement and discrepancies between the true and predicted values. A confusion matrix is a situation analysis table in machine learning that summarizes the predicted results of a classification model, in the form of a matrix that summarizes the records in a dataset according to two criteria: the true value and the value judgement predicted by the classification model. Meanwhile, we use a ROC curve to estimate performance based on a comparison between the false positive rate and the true positive rate when changing the decision threshold. Using the ROC curve, we derive the AUC value as a measure of model effectiveness. We also used the following performance as an assessment: Sensitivity (Sn), Specificity (Sp), Accuracy (Acc), Precision (Pre), and Matthew correlation coefficients (Mcc), they are defined as follows:

$$Sn = \frac{TP}{TP + FN} \tag{12}$$

$$Sp = \frac{TN}{TN + FP} \tag{13}$$

$$Acc = \frac{TP + TN}{TP + TN + FN + FP} \tag{14}$$

$$Pre = \frac{TP}{TP + FP} \tag{15}$$

$$Mcc = \frac{TP * TN - FP * FN}{\sqrt{(TP+FN) * (TP+FP) * (TN+FN) * (TN+FP)}} \tag{16}$$

where TP, FP, TN, and FN stand for true positive, false positive, true negative, and false negative, respectively. With the above calculated indicators, the result of the quantity in the confusion matrix can be converted into a ratio between 0–1. Easy to standardize measurements. The results of all indicators are calculated by ten times cross-validations. In particular, the ten-fold cross-validation scores are generated by concatenating all the tests.

3.5. Other prediction methods for comparison

To verify that MS-B-CNN is superior to other uni-modality methods, we compared the performances of MS-B-CNN-CLN, MS-B-CNN-CNA, and MS-B-CNN-Expr with that of DNN-based approaches [7] such as DNN-CLN, DNN-CNA, and DNN-Expr. Additionally, to further demonstrate that the MS-B-CNN extracts more information and relevant features, we also compare other methods. These methods are as follows: CNN-CLN, CNN-CNA, and CNN-Expr [27].

To validate the effectiveness of the adversarial representation learning in multimodal fusion for breast cancer patient prognosis prediction, we have used the performance metrics mentioned in the previous section for MDAR and other existing methods. The comparative study involves MDNNMD [7], AMND [30], STACKED RF [27], and MAFN [9], and SiGaAtCNN [8]. Likewise, ten-fold cross-validation is used to evaluate the performance.

4. Results

4.1. Validation of the effectiveness of MS-B-CNN

MS-B-CNN was compared to the DNN [7], and CNN [27] models to validate its effectiveness. The AUC values of different models are shown in Fig. 5. Compared with other methods, MS-B-CNN consistently produces better results. In contrast with CNN-CLN, CNN-CNA, and CNN-Expr, the AUC value of MS-B-CNN-CLN, MS-B-CNN-CNA, and MS-B-CNN-Expr is improved by 0.3%, 5.8%, and 3.7%, respectively. The respective Acc, Pre, Sn, and Mcc for all comparable models were also calculated. As shown in Table 5, we can state that MS-B-CNN-CNA achieves 0.3%, and 1.2% improvement in Acc values as compared to DNN-CNA, and CNN-CNA. MS-B-CNN-Expr achieves 5.1%, and 0.4% improvement in Acc values as compared to DNN-Expr, and CNN-Expr. Similarly, each of the other three performance indicators has improved to varying degrees. According to these results, the MS-B-CNN module is an effective tool for predicting breast cancer prognosis.

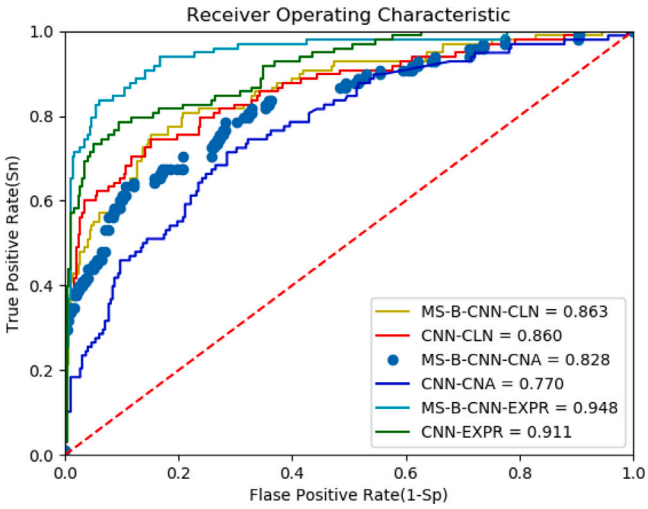

Fig. 5. The ROC curves and AUC values of MS-B-CNN and CNN in uni-modal.

Table 5
Acc, Pre, Sn, Mcc, and F1 predictive performance metrics of MS-B-CNN and other models.

Methods	Acc	Pre	Sn	Mcc	F1
DNN-CLN	0.802	0.620	0.314	0.366	0.413
CNN-CLN	0.806	0.658	0.349	0.376	0.462
MS-B-CNN-CLN	0.806	0.683	0.392	0.355	0.510
DNN-CNA	0.757	0.258	0.113	0.070	0.155
CNN-CNA	0.748	0.409	0.202	0.119	0.267
MS-B-CNN-CNA	0.760	0.437	0.274	0.139	0.340
DNN-Expr	0.759	0.320	0.135	0.086	0.200
CNN-Expr	0.806	0.596	0.430	0.382	0.505
MS-B-CNN-Expr	0.810	0.626	0.451	0.442	0.530

4.2. Validation of the effectiveness of the cross-modality adversarial representation

We benchmarked SimpleConcat and Bi-Attention [8] performance to validate the effectiveness of adversarial representation learning. Specifically, we use a simple three-layer CNN for obtaining intermediate features of the uni-modal data. We demonstrate in Table 6 that the Acc value of the proposed method is 2.3%, 10.5% better than the SimpleConcat and BiAttention models in the METABRIC dataset. Moreover, adversarial representation learning produces corresponding improvements in other indicators as well. In the results presented here, we demonstrate that adversarial representation learning to match distributions prior to feature fusion is indeed useful and effective.

4.3. Validation of the effectiveness of multimodal data

The purpose of this experiment is to demonstrate the significance of fusing multi-modal data and the effectiveness of adversarial training in prognosis, the MDAR models are employed to process different kinds of

Table 6
Comparison of Acc, Pre, Sn, and Mcc between different fusion strategies.

Methods	Acc	Pre	Sn	Mcc	F1
SimpleConcat	0.902	0.841	0.747	0.730	0.840
Bi-Attention	0.820	0.723	0.446	0.467	0.560
Ours	0.925	0.911	0.812	0.794	0.860

Table 7
Comparison of Acc, Pre, Sn, Mcc between different modality combinations.

Methods	Acc	Pre	Sn	Mcc
Gene-CNA	0.925	0.911	0.812	0.794
Gene-CLN	0.925	0.901	0.835	0.800
CNA-CLN	0.830	0.785	0.566	0.511
Gene-CNA-CLN	0.930	0.902	0.828	0.836

data combinations (gene expression profile data, clinical data, and CNA data). Further, it can be used to investigate the effect of gene expression profile data, CNAs, and clinical data on prognosis prediction in breast cancer. The following four comparative experiments were designed:

- MDAR with Gene and CNA data.
  The MDAR model was trained using gene expression and CNA data as inputs, namely Gene_CNA. We extracted the features from gene expression profile and CNA by MS-B-CNN module and pass adversarial representation learning to obtain modality-invariant embedding space. Finally, we use the ensemble model to get the results.
- MDAR with Gene and Clinical data.
  We chose to use Gene expression profile data and Clinical data as inputs to the MDAR model in this experiment, namely Gene-CLN.
- MDAR with CLN and CNA data.
  The MDAR model was trained using Clinical data and CNA data as inputs, namely CLN-CNA.
- MDAR with all three types of data.
  We employ all multi-modal data as input to the MDAR model, namely MDAR (Gene-CNA-CLN).

Table 7 presents the results of the comparative experiments. From Table 7, we observe that using as many modalities as possible performs better than using data from both modalities. For example, the Acc value of Gene-CNA-CLN is 93.0%, which is superior to Gene-CNA, Gene-CLN, and CNA-CLN models by 0.5%, 0.5%, and 10%, respectively. In addition, the metric of the Pre in Gene-CLN and CNA-CLN models are 90.1% and 78.5%, which are inferior to the Gene-CLN model. All comparison results confirm the benefit of integrating better modalities of data and features by adversarial training in prognosis prediction.

4.4. Comparison with other prediction methods

To verify MDAR, we compared its results with other deep learning-based methods, including MDNNMD [7], AMND [30], STACKED_RF [27], MAFN [9], and SiGaAtCNN [8]. The METABRIC dataset was used for the experiments. According to Fig. 6, MDAR achieves the best performance of all deep learning methods and obtains an Acc increase of 10.4%, 8.1%, 2.8%, 4%, and 1.8% compared with MDNNMD, AMND, STACKED_RF, MAFN, and SiGaAtCNN. Furthermore, the Pre, Sn, and Mcc of different methods were also analyzed. The Pre value of MDAR in the METABRIC dataset is 90.2%, this is an improvement over other methods. These results further demonstrate MDAR's ability to predict breast cancer patient prognosis. The results regarding the other metrics are also consistent. According to these findings, the MDAR approach shows a dramatic enhancement in breast cancer prognosis.

We also compared MDAR's performance with that of three common machine learning classification methods, including LR [57], RF [10], and SVM [2]. The METABRIC dataset was used for the experiments.

Table 8
Comparison of Acc, Pre, Sn, Mcc between SVM, RF, LR, and MDAR.

Methods	Acc	Pre	Sn	Mcc
MDAR	0.930	0.902	0.828	0.836
SVM	0.805	0.708	0.365	0.407
RF	0.791	0.766	0.226	0.337
LR	0.760	0.549	0.183	0.209

Table 9
Comparison of Acc, Pre, Sn, Mcc between STACKED RF, SVM, RF, LR, and MDAR on TCGA-BRCA.

Methods	Acc	Pre	Sn	Mcc
SVM	0.758	0.481	0.509	0.336
RF	0.794	0.630	0.278	0.319
LR	0.728	0.390	0.306	0.177
MDNNMD	0.896	0.854	0.792	0.728
STACKED RF	0.917	0.831	0.804	0.764
MDAR	0.916	0.882	0.845	0.764

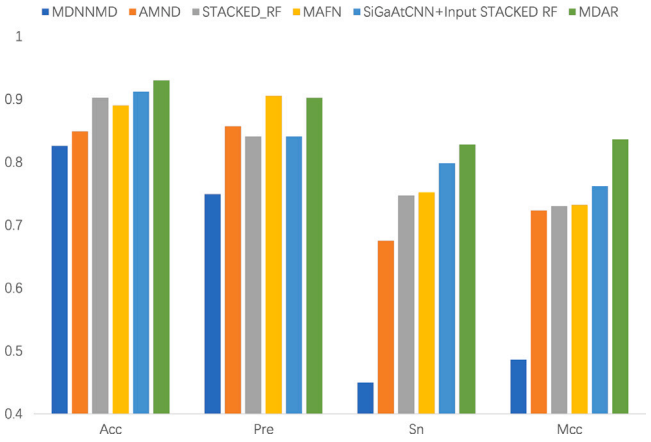

Fig. 6. Comparison of Acc, Pre, Sn, and Mcc of MDAR and existing deep learning-based methods on METABRIC Dataset.

Experimental results in Table 8 demonstrate that MDAR provided better performance than machine learning classification methods. Such as, the Acc value of MDAR on the METABRIC dataset is larger than SVM, RF, and LR by 12.5%, 13.9%, and 17%, respectively. From Table 8, This is clear that the method based on deep learning works better than the traditional method. Overall, MDAR works better than other available deep learning methods, as well as machine learning methods.

4.5. Validation

To further verify the effectiveness of our method, we use an additional breast cancer dataset. There is a new dataset called TCGA-BRCA [58], which contains 1080 breast cancer patients and is used to validate the effectiveness of the proposed model. Cancer-related data, including gene expression profile, clinical details, and copy number alteration, are combined in this dataset. TCGA-BRCA dataset had pre-processed similarly with METRBRIC. There are 830 samples in class 0, which is a short-time survivor, and 250 samples in class 1, which is a long-time survivor. In cross-validation, the TCGA-BRCA dataset also includes three parts: training, validation, and test. The performance indicator for the proposed model and other available methods are provided in Table 9. For the TCGA-BRCA dataset, the MDAR framework is better than other methods. The Acc value of our model is 15.8%, 12.2%, and 18.8% higher than those of non-deep learning methods such as SVM, RF, and LR. And comparable performance to deep learning method, such as MDNNMD and STACKED RF.

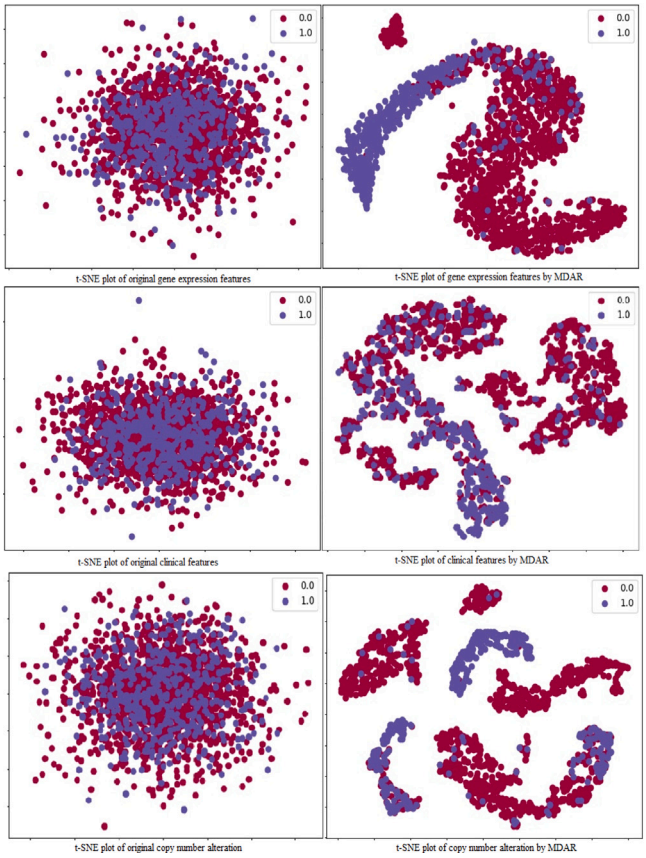

Fig. 7. Visualization of original gene expression profile, and modality-invariant embedding space features extracted by MDAR.

4.6. Visualization of the features

We use t-SNE [59] to visualize the original gene expression profile features as well as the modality-invariant embedding spaces learned from MDAR. As shown in Fig. 7, on the left are the features of the gene expression profile, and on the right are the features obtained from the approach we proposed. The red dot represents the long-term survivors, and the blue dot represents short-term survivors. The original gene expression features have an obvious overlap, while the features extracted from adversarial training show a tendency to separate, which implies that long-term survivors and short-term survivors can be better separated through adversarial training and additional reconstruction and classification loss. The results show that it is possible to transform original features into meaningful representations by adversarial training, and MDAR can produce better modality-invariant embedding space having greater discriminant power to distinguish between long-term and short-term survivors.

4.7. Statistical significance test

A system developed for breast cancer survival prediction is accepted if the obtained results prove its significance over various statistical significance tests like t-test, chi-square test, ANOVA test, etc. To show that the results of MDAR are statistically significant, we performed t-tests using our proposed model and performance evaluation metrics of MDNNMD [7], STACKED RF [27], and SiGaAtCNN [8] models. We executed these models on the METABRIC dataset 10 times and recorded the AUC and Acc values. Later on, we performed a t-test on the recorded measures with scipy library functions, *stats.ttest_ind*, and *scipy.stats.f_oneway*, respectively. For MDNNMD [7], STACKED RF [27], and SiGaAtCNN [27] models, the t-values of t-test for AUC and Acc are 26.21, 19.12, and 23.00, 18.12 and 21.25, 15.98, respectively, and $p < .001$ for all methods. These t-values and p-values imply that our proposed model has statistical significance, and it will be useful for breast cancer survival estimation tasks.

5. Discussion and conclusion

The most common cancer in the world is breast cancer, the disease contributes significantly to the rising mortality rate among cancer patients. The complexity of genes and the diversity of clinical outcomes make it difficult for clinicians to choose the right treatment for different breast cancer patients. As a consequence, a system for predicting breast cancer outcomes needs to be developed. Based on this, we present a deep learning model based on adversarial training to learn a modality-invariant embedding space for a more accurate prediction of breast cancer prognosis using multimodal data. Findings indicate that fusion of features based on better modalities outperforms fewer modalities-based methods for predicting outcomes. Furthermore, our proposal MS-B-CNN module extracts critical information more efficiently than others already existing uni-modal architectures. Additionally, adversarial learning within multimodal data can be significantly narrowed the modality gap. Extensive experiments have shown that by using the modality-invariant embedding space feature as input of the ensemble method, MDAR is comparable to existing methods. The key to this work is to introduce adversarial training into prognostic prediction methods for cancer patients, by aligning the features of different modalities to achieve better performance. Finally, the proposed method provides a new strategy for the prognosis prediction of related diseases.

For the prognosis prediction task, we used gene expression, copy number alteration, and clinical data as inputs. As a next step, we might consider adding modalities such as histopathological images of cancerous tissues, miRNA expression, and gene methylation data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the Provincial Natural Science Research Program of Higher Education Institutions of Anhui province under Grant (KJ2020A0035) and supported by Hefei Municipal Natural Science Foundation (2022009). And then, the authors acknowledge the High-performance Computing Platform of Anhui University for providing computing resources.

References

[1] Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, Freddie Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: Cancer J. Clin. 71 (3) (2021) 209–249.

[2] Xiaoyi Xu, Ya Zhang, Liang Zou, Minghui Wang, Ao Li, A gene signature for breast cancer prognosis using support vector machine, in: 2012 5th International Conference on Biomedical Engineering and Informatics, IEEE, 2012, pp. 928–931.

[3] Leslie R Martin, Summer L Williams, Kelly B Haskard, M Robin DiMatteo, The challenge of patient adherence, Therapeutics Clin. Risk Manag. 1 (3) (2005) 189.

[4] Anika Cheerla, Olivier Gevaert, Deep learning with multimodal representation for pancancer prognosis prediction, Bioinformatics 35 (14) (2019) i446–i454.

[5] Fátima Cardoso, S Kyriakides, S Ohno, F Penault-Llorca, P Poortmans, IT Rubio, S Zackrisson, E Senkus, Early breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up, Ann. Oncol. 30 (8) (2019) 1194–1220.

[6] Jingyang Zhou, Weiwei Cao, Lan Wang, Zezheng Pan, Ying Fu, Application of artificial intelligence in the diagnosis and prognostic prediction of ovarian cancer, Comput. Biol. Med. (2022) 105608.

[7] Dongdong Sun, Minghui Wang, Ao Li, A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data, IEEE/ACM Trans. Comput. Biol. Bioinform. 16 (3) (2018) 841–850.
[8] Nikhilanand Arya, Sriparna Saha, Multi-modal advanced deep learning architectures for breast cancer survival prediction, Knowl.-Based Syst. 221 (2021) 106965.
[9] Weizhou Guo, Wenbin Liang, Qingchun Deng, Xianchun Zou, A multimodal affinity fusion network for predicting the survival of breast cancer patients, Front. Genet. (2021) 1323.
[10] Cuong Nguyen, Yong Wang, Ha Nam Nguyen, Random Forest Classifier Combined with Feature Selection for Breast Cancer Diagnosis and Prognostic, Scientific Research Publishing, 2013.
[11] Yixin Wang, Jan GM Klijn, Yi Zhang, Anieta M Sieuwerts, Maxime P Look, Fei Yang, Dmitri Talantov, Mieke Timmermans, Marion E Meijer-van Gelder, Jack Yu, et al., Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer, Lancet 365 (9460) (2005) 671–679.
[12] Mohammadreza Momenzadeh, Mohammadreza Sehhati, Hossein Rabbani, Using hidden Markov model to predict recurrence of breast cancer based on sequential patterns in gene expression profiles, J. Biomed. Inform. 111 (2020) 103570.
[13] Eskezeia Yihunie Dessie, Jan-Gowth Chang, Ya-Sian Chang, A nine-gene signature identification and prognostic risk prediction for patients with lung adenocarcinoma using novel machine learning approach, Comput. Biol. Med. 145 (2022) 105493.
[14] Marc J Van De Vijver, Yudong D He, Laura J Van't Veer, Hongyue Dai, Augustinus AM Hart, Dorien W Voskuil, George J Schreiber, Johannes L Peterse, Chris Roberts, Matthew J Marton, et al., A gene-expression signature as a predictor of survival in breast cancer, N. Engl. J. Med. 347 (25) (2002) 1999–2009.
[15] Yijun Sun, Steve Goodison, Jian Li, Li Liu, William Farmerie, Improved breast cancer prognosis through the combination of clinical and genetic markers, Bioinformatics 23 (1) (2007) 30–37.
[16] Olivier Gevaert, Frank De Smet, Dirk Timmerman, Yves Moreau, Bart De Moor, Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks, Bioinformatics 22 (14) (2006) e184–e190.
[17] Yanfeng Wang, Chuanqian Zhu, Yan Wang, Junwei Sun, Dan Ling, Lidong Wang, Survival risk prediction model for ESCC based on relief feature selection and CNN, Comput. Biol. Med. 145 (2022) 105460.
[18] Min Yang, Huandong Yang, Lei Ji, Xuan Hu, Geng Tian, Bing Wang, Jialiang Yang, A multi-omics machine learning framework in predicting the survival of colorectal cancer patients, Comput. Biol. Med. 146 (2022) 105516.
[19] Josie Hayes, Helene Thygesen, Charlotte Tumilson, Alastair Droop, Marjorie Boissinot, Thomas A Hughes, David Westhead, Jane E Alder, Lisa Shaw, Susan C Short, et al., Prediction of clinical outcome in glioblastoma using a biologically relevant nine-microrna signature, Mol. Oncol. 9 (3) (2015) 704–714.
[20] Ya Zhang, Ao Li, Chen Peng, Minghui Wang, Improve glioblastoma multiforme prognosis prediction by using feature selection and multiple kernel learning, IEEE/ACM Trans. Comput. Biol. Bioinform. 13 (5) (2016) 825–835.
[21] Li Tong, Jonathan Mitchel, Kevin Chatlin, May D. Wang, Deep learning based feature-level integration of multi-omics data for breast cancer patients survival analysis, BMC Med. Inform. Decis. Mak. 20 (1) (2020) 1–12.
[22] Hua Chai, Xiang Zhou, Zhongyue Zhang, Jiahua Rao, Huiying Zhao, Yuedong Yang, Integrating multi-omics data through deep learning for accurate cancer prognosis prediction, Comput. Biol. Med. 134 (2021) 104481.
[23] Asli Z Dag, Zumrut Akcam, Eyyub Kibis, Serhat Simsek, Dursun Delen, A probabilistic data analytics methodology based on Bayesian Belief network for predicting and understanding breast cancer survival, Knowl.-Based Syst. 242 (2022) 108407.
[24] Sanghyuk Roy Choi, Minhyeok Lee, Estimating the prognosis of low-grade glioma with gene attention using multi-omics and multi-modal schemes, Biology 11 (10) (2022) 1462.
[25] Ashley G. Rivenbark, Siobhan M. O'Connor, William B. Coleman, Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine, Am. J. Pathol. 183 (4) (2013) 1113–1124.
[26] P.C. Stone, S. Lund, Predicting prognosis in patients with advanced cancer, Ann. Oncol. 18 (6) (2007) 971–976.
[27] Nikhilanand Arya, Sriparna Saha, Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model, IEEE/ACM Trans. Comput. Biol. Bioinform. (2020).
[28] Tadas Baltrušaitis, Chaitanya Ahuja, Louis-Philippe Morency, Multimodal machine learning: A survey and taxonomy, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2) (2018) 423–443.
[29] Lauren Houston, Ping Yu, Allison Martin, Yasmine Probst, Heterogeneity in clinical research data quality monitoring: a national survey, J. Biomed. Inform. 108 (2020) 103491.
[30] Hongling Chen, Mingyan Gao, Ying Zhang, Wenbin Liang, Xianchun Zou, Attention-based multi-NMF deep neural network with multimodality data for breast cancer prognosis model, BioMed. Res. Int. 2019 (2019).
[31] Zhi Huang, Xiaohui Zhan, Shunian Xiang, et al., SALMON: survival analysis learning with multi-omics neural networks on breast cancer, Front. Genet. 10 (2019) 166.
[32] Li Tong, Hang Wu, May D. Wang, Integrating multi-omics data by learning modality invariant representations for improved prediction of overall survival of cancer, Methods 189 (2021) 74–85.
[33] Zhiqin Wang, Ruiqing Li, Minghui Wang, Ao Li, GPDBN: deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction, Bioinformatics 37 (18) (2021) 2963–2970.
[34] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[35] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, Brendan Frey, Adversarial autoencoders, 2015, arXiv preprint arXiv:1511.05644.
[36] Sijie Mai, Haifeng Hu, Songlong Xing, Modality to modality translation: An adversarial representation learning and graph fusion network for multimodal fusion, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 01, 2020, pp. 164–172.
[37] Deepanway Ghosal, Md Shad Akhtar, Dushyant Chauhan, Soujanya Poria, Asif Ekbal, Pushpak Bhattacharyya, Contextual inter-modal attention for multi-modal sentiment analysis, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 3454–3466.
[38] Chao Li, Cheng Deng, Ning Li, Wei Liu, Xinbo Gao, Dacheng Tao, Self-supervised adversarial hashing networks for cross-modal retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4242–4251.
[39] Bokun Wang, Yang Yang, Xing Xu, Alan Hanjalic, Heng Tao Shen, Adversarial cross-modal retrieval, in: Proceedings of the 25th ACM International Conference on Multimedia, 2017, pp. 154–162.
[40] Yuxin Peng, Jinwei Qi, CM-GANs: Cross-modal generative adversarial networks for common representation learning, ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) 15 (1) (2019) 1–24.
[41] Tsung-Yu Lin, Aruni RoyChowdhury, Subhransu Maji, Bilinear convolutional neural networks for fine-grained visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 40 (6) (2017) 1309–1322.
[42] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[43] Gao Huang, Shichen Liu, Laurens Van der Maaten, Kilian Q Weinberger, Condensenet: An efficient densenet using learned group convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2752–2761.
[44] Xavier Glorot, Yoshua Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, in: JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.
[45] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[46] Ashfaq Ahmad, Shahid Akbar, Salman Khan, Maqsood Hayat, Farman Ali, Aftab Ahmed, Muhammad Tahir, Deep-AntiFP: prediction of antifungal peptides using distanct multi-informative features incorporating with deep neural networks, Chemometr. Intell. Lab. Syst. 208 (2021) 104214.
[47] Shahid Akbar, Salman Khan, Farman Ali, Maqsood Hayat, Muhammad Qasim, Sarah Gul, iHBP-DeepPSSM: Identifying hormone binding proteins using PsePSSM based evolutionary features and deep learning approach, Chemometr. Intell. Lab. Syst. 204 (2020) 104103.
[48] Shahid Akbar, Maqsood Hayat, Muhammad Tahir, Salman Khan, Fawaz Khaled Alarfaj, cACP-DeepGram: classification of anticancer peptides via deep neural network and skip-gram-based word embedding model, Artif. Intell. Med. 131 (2022) 102349.
[49] Qingchao Chen, Yang Liu, Structure-aware feature fusion for unsupervised domain adaptation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 07, 2020, pp. 10567–10574.
[50] Ashfaq Ahmad, Shahid Akbar, Muhammad Tahir, Maqsood Hayat, Farman Ali, iAFPs-EnC-GA: Identifying antifungal peptides using sequential and evolutionary descriptors based multi-information fusion and ensemble learning approach, Chemometr. Intell. Lab. Syst. 222 (2022) 104516.
[51] Shahid Akbar, Ashfaq Ahmad, Maqsood Hayat, Ateeq Ur Rehman, Salman Khan, Farman Ali, iAtbP-Hyb-EnC: Prediction of antitubercular peptides via heterogeneous feature representation and genetic algorithm based ensemble learning model, Comput. Biol. Med. 137 (2021) 104778.
[52] Rakesh Chandra Joshi, Rashmi Mishra, Puneet Gandhi, Vinay Kumar Pathak, Radim Burget, Malay Kishore Dutta, Ensemble based machine learning approach for prediction of glioma and multi-grade classification, Comput. Biol. Med. 137 (2021) 104829.

[53] Jianjiong Gao, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S Onur Sumer, Yichao Sun, Anders Jacobsen, Rileen Sinha, Erik Larsson, et al., Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, Sci. Signal. 6 (269) (2013) pl1.

[54] Mahmoud Khademi, Nedialko S. Nedialkov, Probabilistic graphical models and deep belief networks for prognosis of breast cancer, in: 2015 IEEE 14th International Conference on Machine Learning and Applications, ICMLA, IEEE, 2015, pp. 727–732.

[55] Hanchuan Peng, Fuhui Long, Chris Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.

[56] Diederik P. Kingma, Jimmy Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

[57] Miles F Jefferson, Neil Pendleton, Sam B Lucas, Michael A Horan, Comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with nonsmall cell lung carcinoma, Cancer: Interdiscip. Int. J. Am. Cancer Soc. 79 (7) (1997) 1338–1342.

[58] Katarzyna Tomczak, Patrycja Czerwińska, Maciej Wiznerowicz, The cancer genome atlas (TCGA): an immeasurable source of knowledge, Contemp. Oncol. 19 (1A) (2015) A68.

[59] Laurens Van der Maaten, Geoffrey Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).