

RESEARCH

Open Access



RNA-RBP interactions recognition using multi-label learning and feature attention allocation

Huirui Han¹, Bandeh Ali Talpur², Wei Liu^{1*}, Limei Wang^{1*}, Bilal Ahmed³, Nadia Sarhan⁴ and Emad Mahrous Awwad⁵

Abstract

In this study, we present a sophisticated multi-label deep learning framework for the prediction of RNA-RBP (RNA-binding protein) interactions, a critical aspect in understanding RNA functionality modulation and its implications in disease pathogenesis. Our approach leverages machine learning to develop a rapid and cost-efficient predictive model for these interactions. The proposed model captures the complex characteristics of RNA and recognizes corresponding RBPs through its dual-module architecture. The first module employs convolutional neural networks (CNNs) for intricate feature extraction from RNA sequences, enabling the model to discern nuanced patterns and attributes. The second module is a multi-view multi-label classification system incorporating a feature attention mechanism. This mechanism is designed to intricately analyze and distinguish between common and unique deep features derived from the diverse RNA characteristics. To evaluate the model's efficacy, extensive experiments were conducted on a comprehensive RNA-RBP interaction dataset. The results emphasize substantial improvements in the model's ability to predict RNA-RBP interactions compared to existing methodologies. This advancement emphasizes the model's potential in contributing to the understanding of RNA-mediated biological processes and disease etiology.

Keywords RNA-RBP interactions prediction, Multi-view deep learning, Multi-label learning

Introduction

Gene expression within individuals results in different phenotypes due to strict regulation, encompassing various levels of control such as gene-level regulation through histone modifications and methylation, transcriptional regulation, RNA-binding proteins (RBPs) interactions, post-transcriptional regulation, and translation regulation [1]. Imbalances in various regulatory factors can lead to changes in downstream gene expression. RBPs, through specific binding to target RNA, directly or indirectly modulate RNA functionality, including mediating RNA maturation, selective splicing, transport, localization, and translation, thereby forming a complex RNA-RBP regulatory network. Mutations and abnormal expression of RBPs can impact various steps in RNA processing, consequently altering

*Correspondence:

Wei Liu

liuweil@hainmc.edu.cn

Limei Wang

wanglm@hainmc.edu.cn

¹ College of Biomedical Information and Engineering, Hainan Engineering Research Center for Health Big Data, Hainan Medical University, Haikou, China

² School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

³ Department of Structural Engineering, Faculty of Civil Engineering, Doctoral School, Akademicka 2, Silesian University of Technology, 44-100 Gliwice, Poland

⁴ Department of Quantitative Analysis, College of Business Administration, King Saud University, Riyadh, Saudi Arabia

⁵ Department of Electrical Engineering, College of Engineering, King Saud University, P.O. Box 800, 11421 Riyadh, Saudi Arabia

the functions of target RNAs. Therefore, genetic alterations and abnormal expression of RBPs are closely related to the occurrence of many complex diseases [2]. Additionally, research indicates a significant reduction in the quantity of RBPs in certain cancer patients [3]. Studies have revealed a substantial number of genetic variations located in RBPs and their regulated target RNAs. These variations may result in abnormal regulation of RBPs [4]. Therefore, a thorough investigation into the interaction mechanisms between RNA and RBPs, the construction of an RNA-centric RBP interaction network, and the identification of genetic variations affecting RNA-RBP interaction relationships can contribute to a deeper analysis of the etiology of certain diseases. Furthermore, this may lead to the discovery of improved therapeutic approaches for treating or alleviating the suffering caused by these diseases [5].

In situations with abundant experimental validation data, it is necessary to establish a rapid and low-cost predictive model for the RNA-RBP interactions. However, it is impractical to perform binding tests for every pair of RNA and RBP in medical scenarios. Since the highly developed landscape of big data and sequencing technologies, numerous algorithms that utilize machine learning models to identify RBP binding sites from RNA sequences have been developed [6]. Rpicool [7] extracts motif information and repetitive patterns from experimentally validated RNA-RBP regulatory data. It pairs the motif sequences of RBPs with the sequences of RNAs, and determines the RNA-RBP interaction based on the pairing conditions. Additionally, research suggests that extracting both sequence and structural information from RNA to construct predictive models for RBP-RNA interaction relationships is an efficient approach [8].

To fully leverage the diverse features of RNA, this paper proposes the construction of a multi-label deep learning model that integrates multiple RNA features to predict the interaction between RNA and RBP. The overall framework of our model is illustrated in Fig. 1. The model comprises a feature extraction module and a multi-view multi-label classification module. The feature extraction module employs three convolutional neural networks to independently extract deep features from the original matrix of RNA sequence, the original matrix of amino acids sequence, and the original histogram matrix of dipeptides. In the multi-view multi-label learning module, the deep private features and common feature are extracted by a fully connected layer with a ReLU activation function separately, and combine them based on the feature dimension to generate the synergistic features, which can express the semantic information of the view data. Following this, a similarity matrix is computed by comparing the original features with the common features. This matrix is then multiplied with the synergistic features to generate attention weights. By fusing the attention weight matrix with the synergistic features matrix through a shortcut branch, a feature matrix capable of the semantics of multi-view RNA-RBPs data is effectively captured. Finally, the feature matrix is fed into a fully connected layer to produce the output for the multi-label learning task that predicts multiple RBPs for a given RNA. The experimental results demonstrate the proposed model achieves some improvements in prediction of the interaction between RNA and RBP by the multi-label deep learning model leveraging the latent relationships among multi-view RNA features.

The main contributions of this paper are summarized as follows:

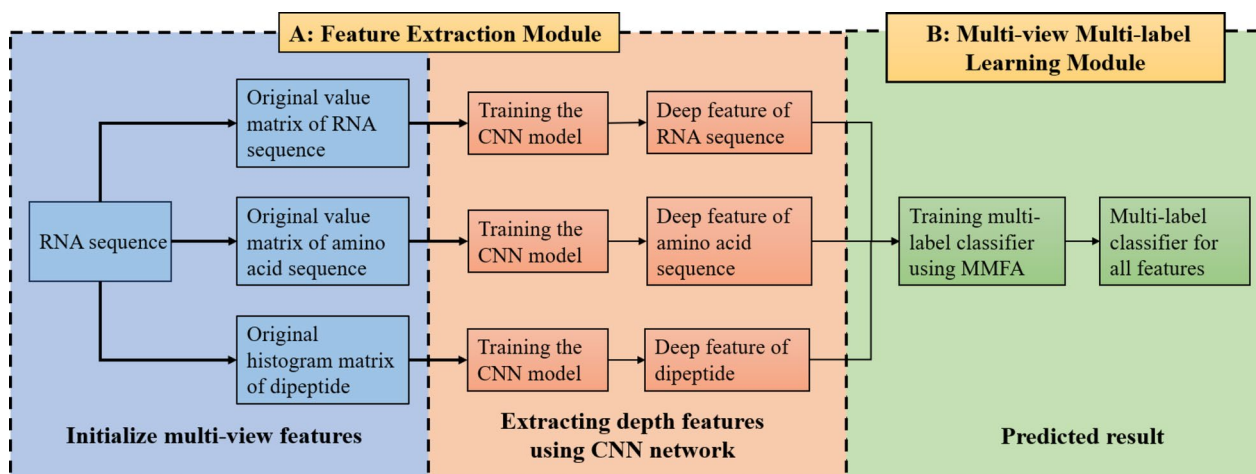


Fig. 1 Overall framework of our model. The model comprises two main modules. Part A is a feature extraction module, and Part B is a multi-view multi-label classification module

- (1) Based on the RNA sequence information, the amino acid features are acquired, and the dipeptide composition is obtained using the dipeptide composition representation method, thus representing the multi-view features of RNA.
- (2) Within a deep feature learning framework, the multi-view features deep features are automatically acquired.
- (3) Considering the varying importance of the captured deep views features, the model with view feature attention assignment is used to extract the common and private deep features from different views to predict the interaction relation of a pair of RNA and RBP.

The article is organized as follows. Related work is introduced in Related work section, and the working principle of the proposed method is shown in Methodology section. Comparative test and results in Experiments and Results section show the advantages of our method in predicting RNA-RBP interactions. Finally, we make some discussions of the work in Conclusion Section.

Related work

Techniques of predicting RNA-RBP interactions

The rapid development of sequencing technology has greatly increased the sequence information of proteins and RNA, and research has also shown that extracting sequence information of proteins and RNA to predict their interactions is an efficient method [9]. LncPNet [10] was introduced to predict potential lncRNA-Protein interactions by embedding the lncRNA-Protein heterogeneous network, and it achieved superior prediction performance. Suresh et al. proposed the RPI-Pred model using support vector machines to predict the interactions between RNA and RBP [11]. This approach involves learning through the interaction of sequence and structural information. The results indicate that the secondary structure information of RNA significantly influences the prediction of RNA-RBP interactions. Leveraging the structures of RBPs and the secondary structures of RNAs, RNACommender [12] constructs a recommendation system to suggest target RNAs to a particular RBP. Experimental validation demonstrates that the model can identify the majority of target RNAs. Additionally, in the RNACommender dataset, at least 74.7% of RNAs are found to bind with at least two proteins. Therefore, to efficiently explore RNA-RBP interactions, the predictive model needs to recommend several RBPs simultaneously. Due to the capability of multi-label deep learning algorithms to simultaneously identify multiple labels for a given sample, these algorithms are suitable for predicting RNA-RBP interactions. In the case of a single RNA

binding to multiple RBPs, iDeepM [13] proposed by Pan et al. employs a deep learning model to extract deep features from RNA sequence information, and adopts a multi-label classification model to predict multiple RBPs that a given RNA may potentially bind to. Li et al. constructed a deep neural network framework named RDense [14], incorporating paired probability features obtained from RNA secondary structure as input, and leveraging a combination of Bidirectional Long Short-Term Memory networks (Bi-LSTM) and DenseNet (Dense Convolutional Network) to learn RBP binding preference.

Nevertheless, the mentioned algorithms have some limitations. Although RNA sequence and secondary structure information are effective for classification, incorporating amino acid information of RNA can represent a more comprehensive RNA feature, contributing to improved predictive accuracy. Additionally, different features of RNA play varying roles in predicting RBPs, and quantifying the importance of RNA features can further enhance the model's effectiveness.

Multi-view multi-label learning methods

The purpose of multi-label learning is to find relevant labels for a given sample as accurately as possible. Therefore, the output of a multi-label learning model may include a set of one or more labels [15]. Two strategies are employed to address multi-view multi-label problems. The first strategy involves constructing a multi-label classifier for each view data and combining the predictive results from all classifiers to obtain the final prediction [16]. For example, Huang et al. [17] utilized multi-label category attributes to make predictions in each view and then combined all predicted results, considering the weight of view contribution for final learning. However, this approach overlooked the diversity of specific information. To address this issue, Zhao et al. [18] proposed a method based on a single hidden layer feedforward neural network without iteration. This algorithm not only utilized the Hilbert-Schmidt Independence Criterion (HSIC) to thoroughly investigate the consistency and diversity of the data, but also considered label correlation and the significance of different viewpoints [19–21].

The second strategy is to integrate all views into a unified view and apply a multi-label learning approach to make predictions. For example, Zhao et al. [22] designed an approach, which firstly discovered the shared spaces of all views by subspace learning and handled the label missing problem by the kernel extreme learning machine. However, this sequential approach may not facilitate sufficient information exchange, potentially resulting sub-optimal models. To fully consider the shared and specific information among views, the multi-view multi-label

learning with view feature attention allocation (MMFA) [23] was proposed.

From these, it is evident that MMFA has already demonstrated a certain level of accuracy in multi-label classification tasks. Therefore, this paper adopts it to extract the common and private deep features from different views to predict the interaction relation of a pair of RNA and RBP.

Methods

Initially, the raw text data extracted from dataset is transformed to the representation of RNA sequence, the representation of amino acid sequence and the representation of dipeptide component by encoding techniques. Subsequently, distinct deep features are generated by three CNN models, and these deep features then serve as the input data for the downstream MMFA classifier. Ultimately, the MMFA produces the final classification results.

The RNA-RBP interactions data

The data used in this study is sourced from the AURA website [24]. To investigate the issue of RBP binding correlation, a subset of data related to RNA and RBP was selected, comprising 67 RBPs, 72,226 RNA sequences, and 550,386 binding sites between RNA and RBPs. The final distribution of RBPs is described in the Fig. 2. The counts of interacting-RNA or the 14 RBPs are all below 1000, with AGO4 having the fewest interacting RNA sequences, only 400 in total. On the contrary, AGO1 has the highest count of interacting RNA sequences, totaling

31,964. It is evident from the counts of interacting RNA that there is a significant imbalance in the distribution of RBPs.

Additionally, a new label "negative" was introduced to represent the negative correlation category, including 31,964 RNA sequences without binding site information as samples in this category. Therefore, a total of 68 labels were considered.

The representation of the RNA sequence

RNA sequences are composed of four naturally occurring bases—A (adenine), G (guanine), C (cytosine), and U (uracil)—arranged in a specific order [25]. RNA sequences possess both sequential and spatial structures, and until now numerous RNA sequences have been discovered through advanced sequencing technologies, but exploring the spatial structure of RNA sequences incurs significant costs [26]. Artificial intelligence methods commonly utilize the sequential structure of RNA, such as the arrangement information of bases in RNA sequences, to characterize the RNA sequence. This is primarily achieved using encoding methods to represent the RNA sequence. In this study, the one-hot encoding method is employed, transforming the textual sequence of RNA into a numerical matrix, serving as input for machine learning models. Through one-hot encoding, an RNA sequence of length n is converted into a blank matrix of size $4 \times n$. Using the vectors $(1, 0, 0, 0)^T$, $(0, 1, 0, 0)^T$, $(0, 0, 1, 0)^T$ and $(0, 0, 0, 1)^T$ to represent the occurrence of bases A, C, G, and U in the RNA sequence

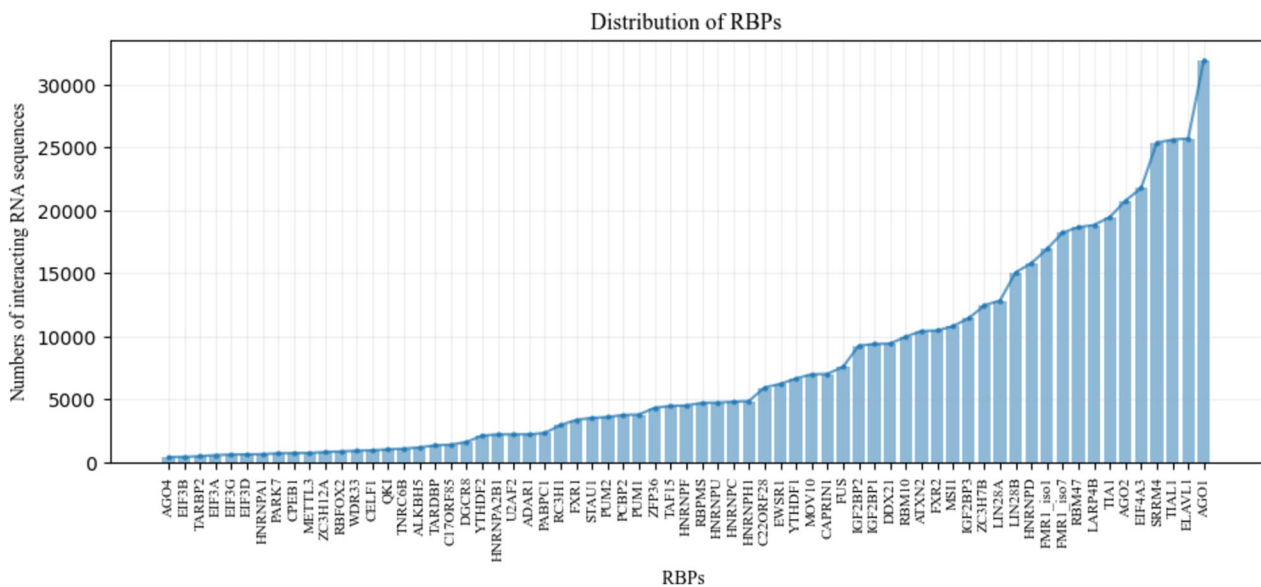


Fig. 2 Distribution of RBPs. It is evident from the numbers of interacting RNA sequences that there is a significant imbalance in the distribution of RBPs

respectively, then the matrix is filled based on the positions of bases in the RNA sequence. Since the lengths of RNA sequences in the dataset are various, the length of the RNA sequence is fixed to n_m . If the length of the provided RNA sequence is less than n_m , bases are padded with the vector representation of base B, which is $(0.25, 0.25, 0.25, 0.25)^T$. Consequently, through the one-hot encoding technique, each RNA is transformed into a two-dimensional matrix of size $4 \times n_m$. The process of representing RNA sequence feature is shown in Fig. 3A.

The representation of the amino acid sequence

RNA sequences providing a partial representation of RNA characteristics but lacking sufficient information. Amino acid sequences represent a high-dimensional expression form of RNA sequences [27].

In theory, any RNA sequence can be transcribed into mRNA, aligning the start and stop positions of the RNA sequence with the coding region, and then translated to form the corresponding amino acid sequence. An amino acid sequence consists of 20 different amino acids, each encoded by a set of three bases in the RNA sequence. This implies that each amino acid in the amino acid sequence

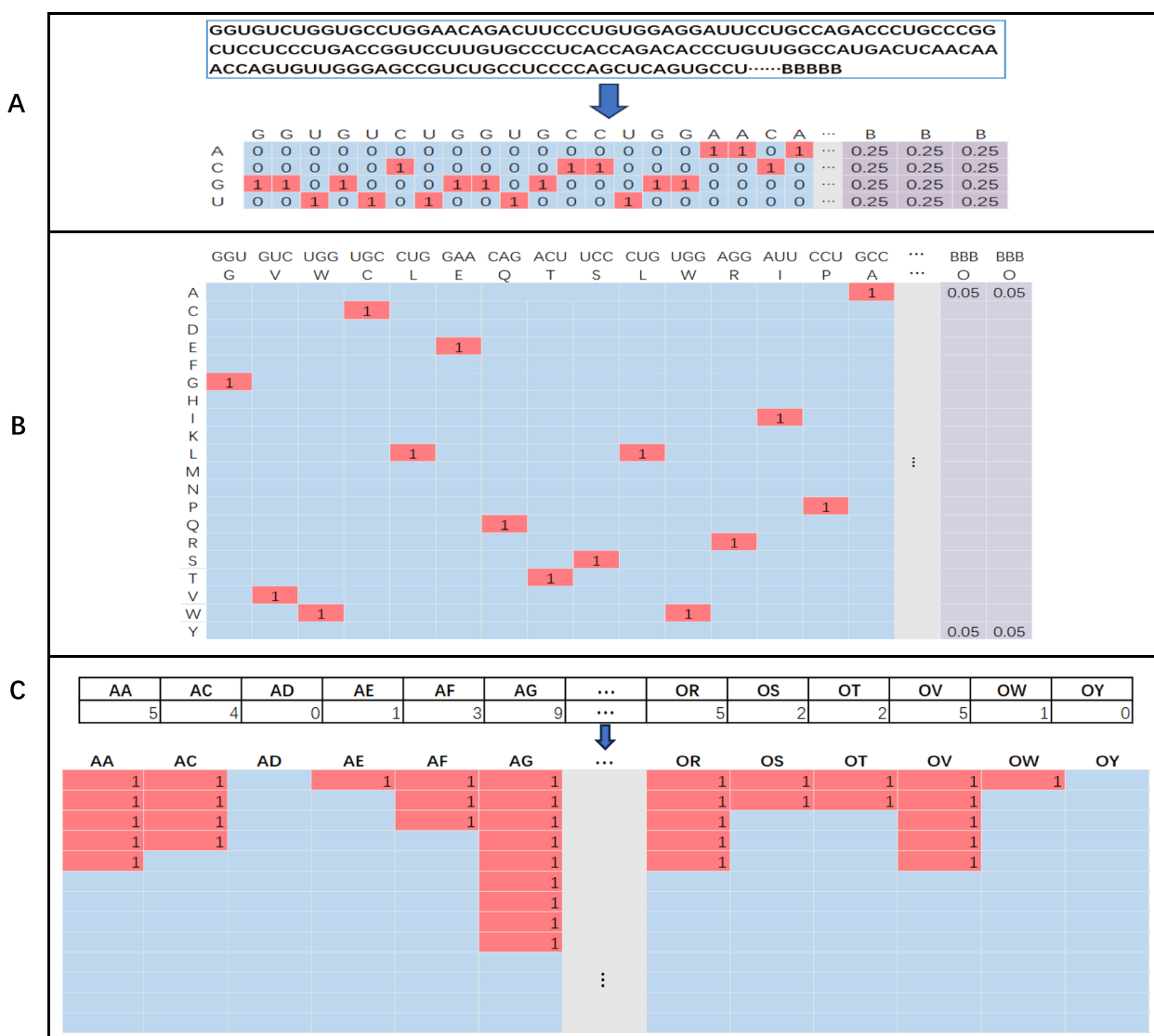


Fig. 3 The representations of multi-view RNA features. **A** The process of one hot encoding for RNA sequences. Through one-hot encoding, an RNA sequence of length n is converted into a blank matrix of size $4 \times n$. **B** The matrix of an amino acid sequence with size $20 \times n$, which indicates the occurrence of amino acids in the amino acid sequence. **C** The process of generating the dipeptide features. Here, the occurrence of dipeptides in an amino acid sequence is shown in the table, and the dipeptide histogram is depicted in the following matrix

carries contextual information from the RNA sequence. Therefore, the amino acid sequence can be viewed as a high-dimensional product of the RNA sequence, providing richer information. While there are 4^3 possible combinations of selecting three bases from the four, "UGA," "UAG," and "UAA" are stop codons that cannot be translated into amino acids. Therefore, there are a total of 61 corresponding relationships bases and translated amino acids. Translating RNA sequences into amino acid sequences is a unidirectional and unique process. However, due to the fact that one amino acid can correspond to multiple base combinations, employing conventional translation methods results in amino acid sequences that cannot be reverted to the original RNA sequence. This leads to consequences of information loss and misinterpretation. To address this issue, three codon-based translation approaches are employed in this study for translating RNA sequences into amino acid sequences: (1) translating the RNA sequence from the beginning; (2) skipping the first base of the RNA sequence and starting the translation; (3) skipping the first and second bases of the RNA sequence and commencing the translation. Through codon-based translation approaches, an RNA sequence of length n is translated into 3 amino acid sequences with length $1/3n$. By concatenating these three amino acid sequences, an amino acid sequence of length n is obtained. Subsequently, one-hot encoding is applied to convert it into a blank matrix of size $20 \times n$.

A 20-dimensional standard orthogonal basis vector is utilized to represent the occurrence of amino acids in the amino acid sequence. For example, the order of amino acid arrangement is "ACDEFGHIKLMNPQRSTVWY". The matrix is then filled according to the positions of amino acids, which is shown in Fig. 3B. For instance, in the vector representation of amino acid A, the first column is set to 1; for amino acid C, the second column is set to 1; for amino acid G, the sixth column is set to 1; for amino acid I, the eighth column is set to 1, and so forth. For segments that cannot be translated, such as those containing stop codons in the RNA sequence and bases B used for padding, represented by the letter O, all values in the 20-dimensional vector are set to 0.05.

The representation of the dipeptide component

Both RNA sequences and amino acid sequence focus on extracting sequence information, but it is also necessary to consider the composition of RNA. Dipeptides [28, 29] are a method of studying the structure of amino acid sequences, which focuses on the combination of any two amino acids. By considering the role of hydrogen bonds in the secondary structure of proteins, the g-gap dipeptide composition [30] not only describes the correlation between two amino acids in the amino acid sequence,

but also considers the possibility of two amino acids that are farther apart in the sequence being adjacent in three-dimensional space. Therefore, utilizing the g-gap dipeptide composition to generate g-gap dipeptide features can provide a more comprehensive description of the composition information in both RNA sequences and amino acid sequences.

A dipeptide is formed by the linkage two amino acids through a peptide bond. Due to the spatial structure of the side chain and backbone of amino acids, dipeptides are sensitive to the arrangement of the left and right amino acids. Therefore, different arrangements can result in distinct dipeptide structures [31, 32]. This sensitivity imparts significant meaning to dipeptides in describing the structure and functionality of amino acid sequences. There are 400 combinations of 0-gap dipeptides for the 20 amino acids. By analyzing the occurrence frequency of 0-gap dipeptide combinations, we can gain insights into the distribution of different dipeptides in amino acid sequences, thereby inferring the composition and arrangement characteristics of RNA sequences. The converted amino acid sequence includes the padding amino acid O, resulting in a total of 21×21 dipeptides. After removing the meaningless dipeptide "OO", there are ultimately 440 dipeptides. As shown in Fig. 3C, dipeptides where both amino acids are alanine are represented as "AA" with the number 5 indicating their occurrence in the sample sequence. Converting the one-dimensional vector of dipeptides into a 440×30 dipeptide histogram facilitates the extraction of more robust deep features by deep learning models. The dipeptide histogram is depicted in Fig. 3C, where "AA" appears 5 times, signifying that the top five elements in columns the vector representing "AA" are set to 1, and the rest are set to 0.

The generation of multi-view deep features

By employing biological methods, the representation of the RNA sequence, the representation of the amino acid sequence and the representation of dipeptide component were obtained. Based on these, three distinct deep convolutional networks were constructed to individually extract deep features. Convolutional Neural Network (CNN) [33–35], a common deep learning architecture, possesses advantages in convolutional computations and deep structures. The core components of a CNN are convolutional layers and pooling layers. The convolutional layers exhibit the property of "weight sharing", which not only reduces the number of parameters but also mitigates the risk of overfitting associated with an excessive number of parameters, thereby effectively utilizing local information from input features. The pooling layers serve to decrease

the spatial dimensions of the input data, reducing both computational requirements and memory consumption of the model.

Since the max length of the RNA sequence n_m is set to 2700, each RNA is transformed into a two-dimensional matrix of size 4×2700 by the one-hot encoding technique. To obtain contextual information from the beginning and end of the RNA sequence, 5 additional base B were added to both the start and end of the RNA sequence. Consequently, the RNA sequence is represented by a two-dimensional matrix of size 4×2710 . 101 convolutional kernels of size 10 were used to perform convolutional calculations on the RNA sequence matrix by the stride of 1, and then 101 feature maps with size 1×2701 were obtained. To reduce computational complexity and eliminate noise, 101 feature maps were processed by a maxpooling layer with a size of 3, resulting in 101 deep feature vectors of size 1×900 . After flattening 101 deep feature vectors, a deep feature map with size 1×90900 was obtained. Then it was calculated by a dropout layer with rate of 0.5 and a fully connected layer to generate the intermedia outputs with size of 1×202 . At last, a dropout layer with rate of 0.5 and a fully connected layer with 68 hidden units were used to gain the final outputs with size of 1×68 . Figure 4A shows the CNN model of extracting deep features for RNA sequences.

According to codon-based translation approaches, each RNA sequence can be transformed to amino acid sequence, represented by a two-dimensional matrix of size 20×2710 . The deep learning framework for amino acid sequence is similar to that for RNA sequence. The amino acid matrix is processed by a convolutional layer, a maxpooling layer and two fully connected layers with dropouts. Therefore, the size of intermedia outputs is also 1×202 and the size of final outputs is also 1×68 . Figure 4B shows the CNN model of extracting deep features for amino acid sequences.

For the dipeptide component representation, a two-dimensional matrix of size 30×440 is obtained by counting the numbers of 0-gap dipeptide combinations in the amino acid sequence and converting the one-dimensional vector of dipeptides into a dipeptide histogram facilitates. Since the size of dipeptide component matrix is less than others, the maxpooling layer is removed from the deep learning framework for dipeptide component features. After processing by a convolutional layer and two fully connected layers with dropouts, the intermedia outputs with size of 1×202 and the final outputs with size of 1×68 are gained. Figure 4C shows the CNN model of extracting deep features for dipeptide component.

The rectified linear unit (ReLU) incurs lower computational overhead and, to some extent, helps prevent

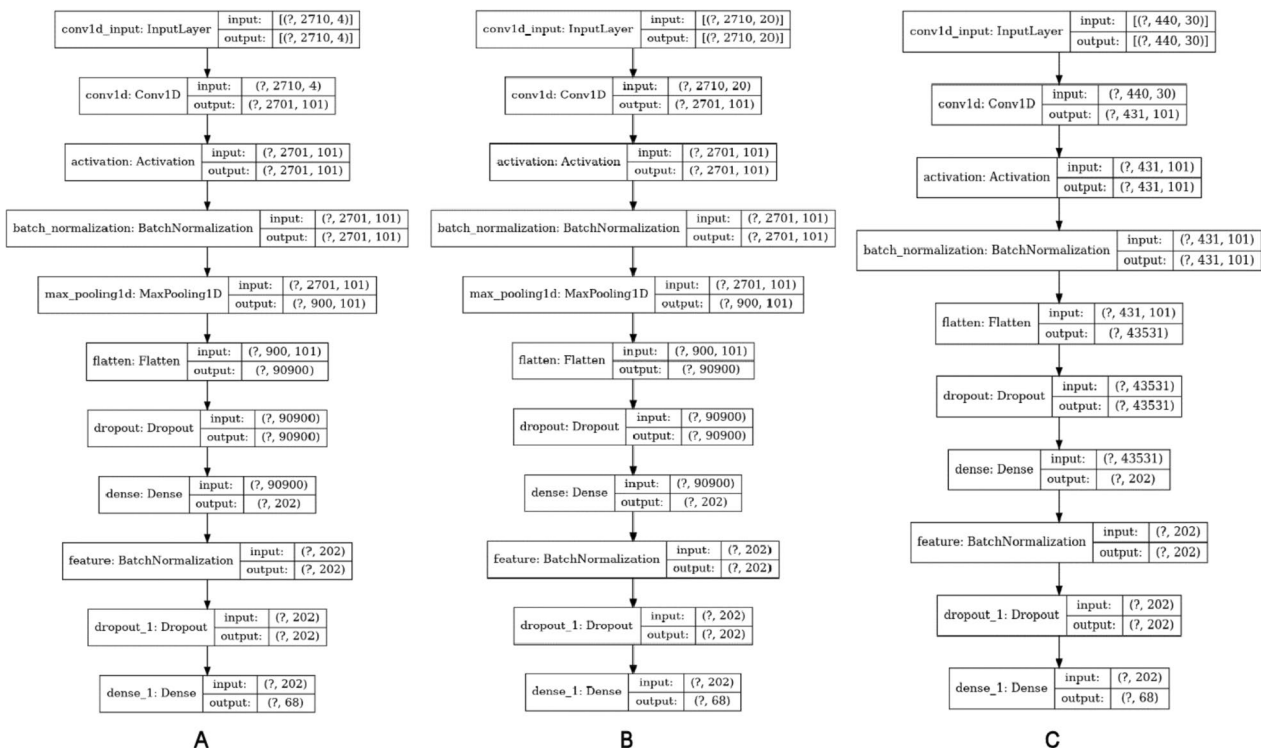


Fig. 4 Extracting deep features from the three views using deep CNN models

the vanishing gradient problem. Consequently, with the exception of the final layer, ReLU activation is employed in all layers of the three deep learning models. The formula for the ReLU activation function is given by:

$$f(z) = \max(0, z) \quad (1)$$

The advantages of the sigmoid activation function include an output range of $[0, 1]$. Due to this output range, the function normalizes the output for each neuron. It is particularly suitable for models that aim to predict probabilities as output. The expression for the sigmoid activation function is as follows:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

For the problem of binary classification, a binary cross entropy is generously used to compute the loss in each training iteration, the binary cross entropy is mathematically formalized as Eq. 3.

$$Loss = -\frac{1}{N} \sum_{i=1}^N y_i \bullet \log(P(y_i)) + (1 - y_i) \bullet \log(1 - P(y_i)) \quad (3)$$

In the above equation, $y_i = 1$ represents the instance i is positive sample, whereas $y_i = 0$ represents the instance i is negative sample. N is the number of instances. $Loss$ computes the average value of all instances. The probability of the instance i is belong to the current label is represented by $P(y_i)$. Binary cross entropy is used to evaluate the quality of predictions made by a binary classification model. In other words, for the case where the label y is 1, if the predicted value $P(y_i)$ approaches 1, then the value of the loss function should approach 0. Conversely, if the predicted value $P(y_i)$ approaches 0, then the value of the loss function should be very large, which is consistent with the properties of the logarithmic function. By summing up and averaging the individual output losses calculated for all outputs, we can obtain the loss of the model for a set of N outputs.

Binary Cross Entropy can handle predictions for multiple labels simultaneously and compute the loss value for each label. For each label, the Binary Cross Entropy loss function treats it as a binary classification problem and calculates the loss by comparing the predicted value with the true value. Therefore, Binary Cross Entropy is applied as the loss function for each deep learning model.

The objective of training three deep learning models is to obtain deep representations of multi-view features. The final outputs of each model, after passing through the final layer and the Sigmoid activation function, already exhibit a clear classification trend. However, this is not conducive to training the subsequent multi-label

feature learning model. Therefore, we utilize the intermediate outputs from the second-to-last layer, which consist of 202 dimensions, as the deep feature of the model.

The training of multi-view multi-label learning

Different from traditional single-view learning, multi-view learning enables more comprehensive exploration of information from multiple views. Due to the importance of different views, the existing multi-view multi-label learning algorithms typically calculate the weight of views by the number of correctly predicted labels. Such method ignores the contribution of data features of each view for multi label classification, therefore this paper used multi-view multi-label learning with view feature attention allocation (MMFA) to deal with multi-label classification problem.

Deep RNA sequence features $X^1 \in \mathcal{R}^{n \times 202}$, deep amino sequence features $X^2 \in \mathcal{R}^{n \times 202}$ and deep dipeptide composition features $X^3 \in \mathcal{R}^{n \times 202}$ are the original inputs of MMFA. The structure of MMFA framework is shown in Fig. 5, which is mainly constructed by two parts.

In the Part A, with the input of multi-view deep features, common feature extraction layer and private feature extraction layer will extract the common features and private features of each view. These feature extraction layers contain a fully connected layer with ReLU activation function. Common features are derived by minimizing the adversarial loss and incorporating the shared subspace multi-label loss. Subsequently, an orthogonal constraint is applied to eliminate these common features from the original set, resulting in the extraction of private features [36]. The feature dimensions of common features and private features are both represented by k . By feature extraction, the common features $C \in \mathcal{R}^{n \times k}$, the private features of deep RNA sequence features $Q^1 \in \mathcal{R}^{n \times k}$, the private features of deep amino sequence features $Q^2 \in \mathcal{R}^{n \times k}$ and the private features of deep dipeptide composition features $Q^3 \in \mathcal{R}^{n \times k}$ are generated. The combined features obtained by concatenating common and private features are referred to as synergistic features $P = [Q^1, Q^2, Q^3, C] \in \mathcal{R}^{n \times [(3+1) \times k]}$.

In the Part B, a multi-head attention mechanism is utilized to calculate the attention weights of view features. Considering the different contributions of multi-view features in label prediction, the multi-head attention mechanism is applied as a more comprehensive approach. The attention weights of multi-view features are computed by Eq. 4.

$$Attention(X^v, C, P) = softmax\left(\frac{X^v C^T}{\sqrt{t_c}}\right)P \quad (4)$$

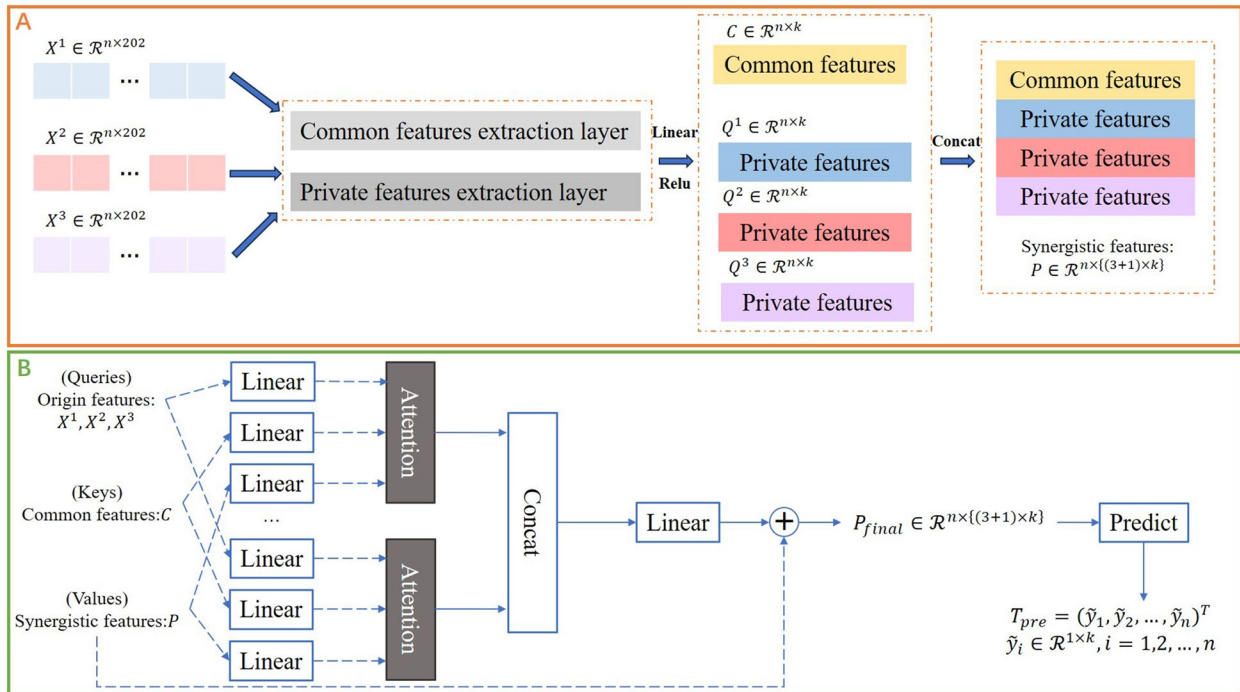


Fig. 5 Considering multi-view deep features by multi-view multi-label learning with view feature attention allocation

where t_c is a scaling factor, its value is the dimensionality of the common features. In this part, the original features are considered as queries, the common features as keys, and the synergistic features as values. X^v and C are used to calculate the similarity between original deep features and common features. Using scaling factor is a method of improving dot product attention to alleviate the impact of input vector dimensions on attention weights. After obtaining the similarity matrix, it will be normalized by softmax function [37] and then multiplied with synergistic features P to gain the attention weights. Inspired by the multi-head attention mechanism, three original deep features were taken as queries and divided into three parts. The formular can be represented as:

$$Multihead(X^1, X^2, X^3, C, P) = Concat(head_1, head_2, head_3)W^o \quad (5)$$

$$head_i = Attention(X^i, C, P) \quad (6)$$

in which, the attention of i -th view is represented by $head_i$. The fusion of the stitched results is indicated by W^o . These attention weights are combined with the synergistic features through a shortcut branch to produce the synergistic features with attention. The formular is expressed as follows:

$$P_{final} = P + Multihead(X^1, X^2, X^3, C, P) \quad (7)$$

The resulting synergistic features are then passed through a fully connected layer $H(\bullet)$ for multi-label prediction. Finally, the multi-label predictive results of MMFA are denoted by Eq. 9.

$$T_{out} = H(P_{final}) \quad (8)$$

$$T_{pre} = sign(T_{out}) \quad (9)$$

$$sign(x) = \begin{cases} 1, & \text{if } x - thres > 0 \\ 0, & \text{else} \end{cases} \quad (10)$$

The multi-head attention plays a key role in the multi-view multi-label classification neural network, and Adam was used as an adaptive optimizer that decouples the weight decay from the optimization step, allowing for separate optimization. In MMFA framework, the final loss function is mathematically formalized as follow:

$$L = l_{ml} + \lambda l_{common} + \gamma l_{private} \quad (11)$$

where λ and γ are the trade-off parameters for the loss terms, and l_{ml} represents the multi-label loss of the final multi-label prediction.

Experiments and results

Metrics

This paper employs two evaluation metrics, namely the AUC area and F1 score, to assess the classification and prediction performance of the model. AUC, defined as the area under the ROC curve, reflects a comprehensive measure of sensitivity and specificity [38]. A higher AUC value indicates better classification performance of the model. In the context of multi-label classification, imbalances in class samples can result in significant bias in performance metrics. Therefore, this paper introduces a Weight constraint. Weighted-AUC calculates the weight for each class based on the number of samples, and then performs a weighted sum of the AUC values for these classes. In addition to assessing the model's classification performance, F1-score is computed to evaluate the predictive performance of the model. The F1 score, a harmonic mean of precision and recall, is subject to the Weight constraint, similar to the AUC metric.

Exact Match Rate refers to the situation where, for each sample, the prediction is considered correct only if the predicted value is identical to the true value. In other words, if there is any difference in the predicted results for a particular category, it is considered incorrect. Therefore, the accuracy calculation formula is given by:

$$MR = \frac{1}{m} \sum_{i=1}^m I(y^{(i)} == \hat{y}^{(i)})$$

where m is the number of samples. $I(\bullet)$ is the indicator function, it is equal to 1 when $y^{(i)}$ is same as $\hat{y}^{(i)}$, otherwise it is equal to 0. As observed, a higher MR value indicates a higher accuracy in classification.

The One Error evaluation metric in multi-label classification measures the probability that at least one label is predicted incorrectly among all samples. Specifically, for each sample, if the model's prediction contains at least one incorrect label, the sample is considered an event of One Error. The formula for calculating the One Error metric is as follows:

$$OE = \frac{1}{m} \sum_{i=1}^m \delta_i$$

in which m is the number of samples. δ_i is the indicator function, which is 1 when at least one label is predicted incorrectly for the i -th sample and 0 otherwise. The One Error evaluation metric ranges from 0 to 1, with a lower value indicating better performance of the model in multi-label classification tasks.

Experiments

For the CNN models of learning multi-view deep features, the number of learning interactions is set to 50, and the learning rate and the decay are applied to 0.001 and 0.01 respectively in each epoch. The dropout rate is set to 0.5 for the dropout layer, and the batch size is set to 64. Adam [39] was used as an adaptive optimizer that decouples the weight decay from the optimization step, allowing for separate optimization.

For multi-view multi-label learning with view feature attention allocation, the number of learning interactions is set to 20 and the dimensions of common features and private features are both fixed as 68. Moreover, the best combination of main parameters was selected after trying many combinations. Finally, λ controlling the contribution of common information is set to 10^{-3} , and γ controlling the contribution of private information is set to 10^{-4} .

The running server is configured with an 8-core Intel CPU, NVIDIA RTX3090 GPU and 768 GB of memory and the software uses Linux operating system. PyTorch library was used to implement the RNA-RBP interaction recognition model, and the codes were written in Python3.7. For the filtered AURA dataset, 80% of the cases were randomly allocated to the training set, while the remaining 20% of cases were reserved for the testing set.

The experimental results are shown in Table 1, where the original RNA sequence features, the original amino sequence features and the original dipeptide composition features are denoted by X_o^1 , X_o^2 , X_o^3 . It can be observed that single-view models using deep learning achieve satisfactory results. Among them, the dipeptide composition view yields the best performance. This is attributed to the fact that the dipeptide composition view not only contains sequence order information but also encompasses sequence composition and structural details, making it the most informative perspective. To explore and leverage multiple deep features, the multi-view multi-label learning with view feature attention allocation (MMFA) is used to process the captured deep features to predict the RBPs. To validate the effectiveness of integrating multi-view features, experiments were conducted with double-view MMFA models and triple-view MMFA models. From the results, it is not difficult to observe that the performance of double-view MMFA models is superior to any single-view model using one of these two features. Especially, the triple-view MMFA model, which integrates all deep features of RNAs, outperforms any single-view model and double-view MMFA models in terms of AUC, F1 score, and Exact Match Rate. Furthermore, it exhibits a lowest One Error compared to any other models. This not only indicates that different

Table 1 Performance of RNA-RBP interactions prediction

	Model	Weighted-AUC↑	Weighted-F1↑	Exact Match Rate↑	One Error↓
Single-view Model	X_0^1 (CNN)	0.8843	0.6489	0.2923	0.3922
	X_0^2 (CNN)	0.8864	0.6667	0.3179	0.3633
	X_0^3 (CNN)	0.9129	0.6901	0.2983	0.3131
Double-view Model	$X_0^1 + X_0^2$ (CNN+MMFA)	0.8968	0.6672	0.3057	0.3389
	$X_0^1 + X_0^3$ (CNN+MMFA)	0.9147	0.6993	0.3605	0.3016
	$X_0^2 + X_0^3$ (CNN+MMFA)	0.9167	0.6967	0.3662	0.2923
Triple-view Model	$X_0^1 + X_0^2 + X_0^3$ (CNN+MMFA)	0.9195	0.7104	0.3738	0.2803

↑ indicates higher values represent better model performance, while ↓ indicates lower values represent better model performance

features of RNA can provide complementary information, but also demonstrates that the effectiveness of considering the importance of these features.

To provide further validation of the efficacy of assigning attention weights to view synergistic features, a component analysis was conducted. MMFA-MA, which does not assign weights to view features, was performed to validate the effectiveness of assigning attention weights to view features. The outcomes of component analysis are depicted in Fig. 6. It is evident that MMFA outperforms MMFA-MA across nearly all evaluation metrics, underscoring the effectiveness of MMFA in recognizing RNA-RBP interactions by considering the varying importance of the captured deep views features of RNA.

There are two main parameters of MMFA and the grid search was used to find the best parameters for the AURA dataset. One is λ controlling the contribution of

common information, another is γ controlling the contribution of private information. If the value of λ is too small, fewer common information is extracted and the attention weights of MMFA are diminished, it may lead to poor performance for each metric. On the other hand, when the value of γ is too large, there is an abundance of private information about the views, and the exchange between views decreases, which may diminish the algorithm’s performance. For discovering the best combination of parameters, this paper searched λ and γ in the range of $[10^{-4}, 10^{-2}]$. The experimental results on each metric are depicted in the Fig. 7, with specific numerical values for the top three rankings for each metric. In the case of $\lambda = 10^{-3}$ and $\gamma = 10^{-4}$, the weighted-AUC value ranks third, the weighted F1 value ranks first, EM value ranks second, and OE value ranks first. Additionally, only with this specific parameter value, all evaluation metrics

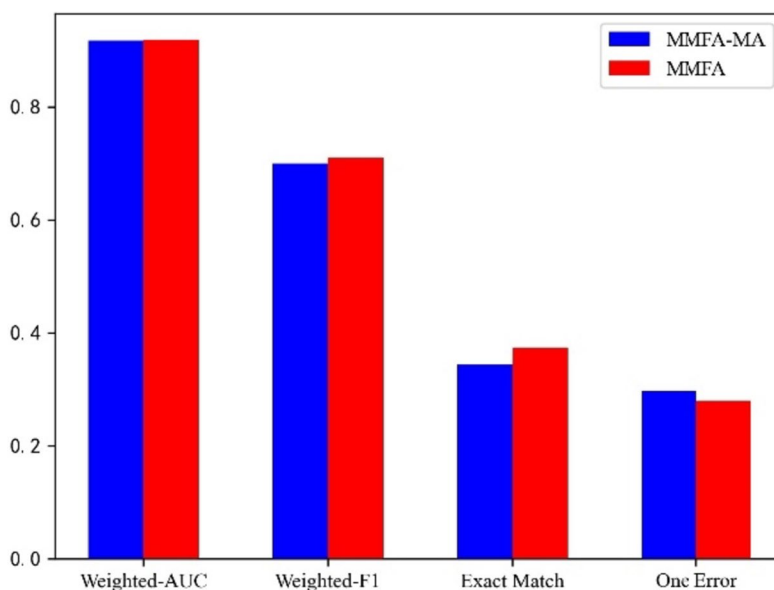


Fig. 6 Parameter sensitivity analysis on the RNA-RBP dataset

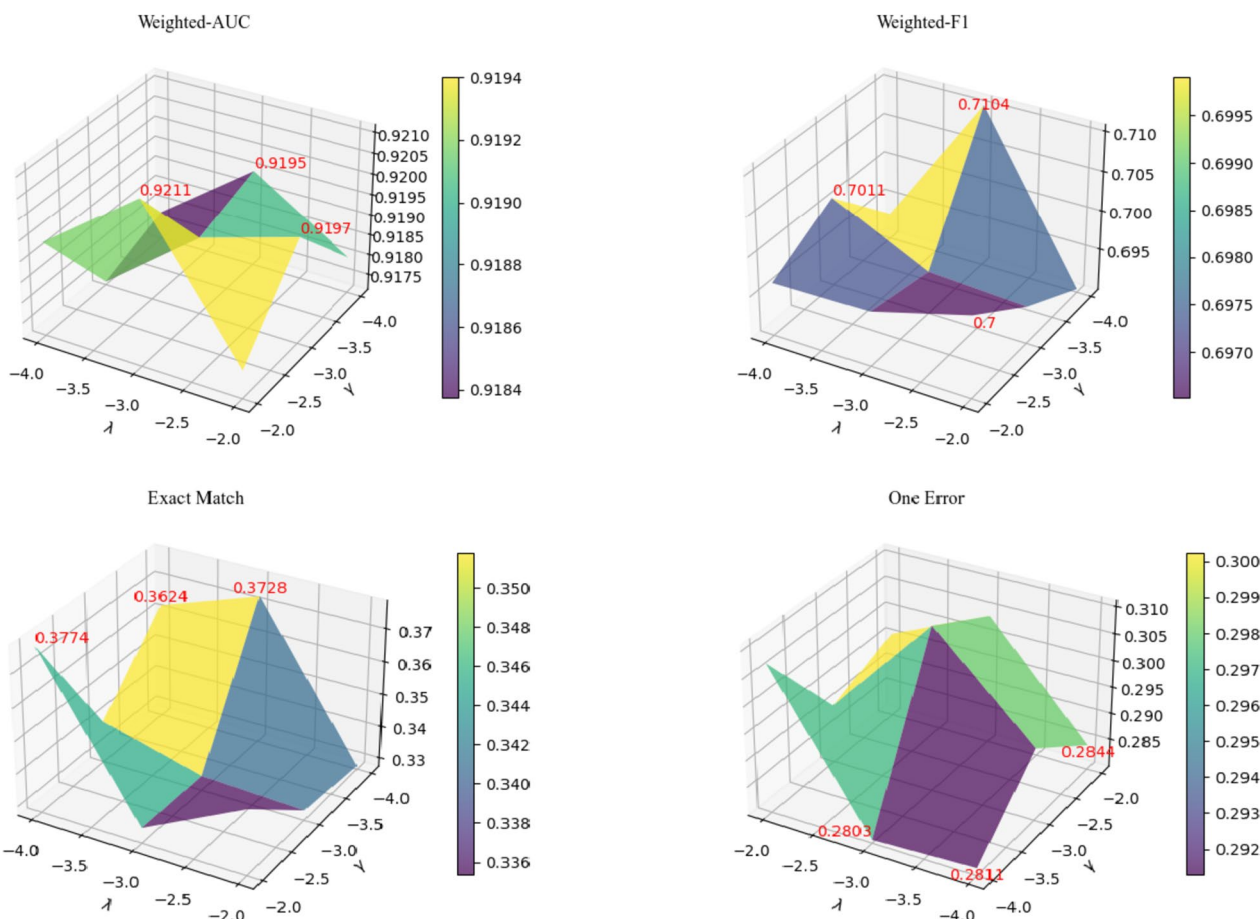


Fig. 7 Parameter sensitivity analysis on the RNA-RBP dataset

achieve a top-3 ranking within the designated value combination. Therefore, for this AURA dataset, the optimal λ is 10^{-3} and the optimal γ is 10^{-4} .

Conclusion

In this research, we presented a multi-view multi-label approach to RNA-RBP interactions recognition that integrates RNA sequence, the amino acids sequences of RNA, and the dipeptides of RNA by utilizing feature attention allocation. Although the performance of the proposed model has improved, there are still some shortcomings and areas for further improvement in this study. For instance, the model did not utilize information about RBPs, and did not consider the relationships among RBPs. Incorporating RBP information, including sequence and structural details, and utilizing the similarity between RBPs to guide the training of the multi-label learning model may potentially further improve the model’s performance. Additionally, the dataset used in this study exhibits a significant imbalance in the number of RNAs interacting with different

RBPs, posing a typical issue of label imbalance. This imbalance can have a considerable impact on the learning and classification effectiveness of the model. Future research should explore methods to mitigate the effects of label imbalance on RNA-RBP interactions recognition.

Acknowledgements

We would like to express our sincere thanks to the editors and reviewers for their insightful comments and suggestions on improving this paper.

Authors’ contributions

All authors were involved in the research for this paper. Huirui Han and Wei Liu led the entire work. Wei Liu and Limei Wang carried out the detailed research, including background study and review compilation. Huirui Han were responsible for the modeling and experimental analysis part of the whole paper. Bandeh Ali Talpur, Bilal Ahmed, Nadia Sarhan and Emad Mahrous Awwad were responsible for writing the thesis as well as the experimental evaluation. All authors read and approved the final draft.

Funding

This work is supported by Hainan Provincial Natural Science Foundation of China (no. 621QN246), the Research Cultivation Foundation of Hainan Medical University (no. HYPY2020039) and the Researchers Supporting Program number (RSPD2024R1052) of King Saud University, Riyadh, Saudi Arabia.

Declarations

Ethics approval and consent to participate

This article does not contain any studies with human participants or animals performed by any of the authors.

Competing interests

The authors declare no competing interests.

Received: 31 December 2023 Accepted: 10 February 2024

Published online: 07 March 2024

References

- Handshakes and Fights (2017) The regulatory interplay of RNA-binding proteins. *Front Mol Biosci* 4:67
- Asada K, Sakaue F, Nagata T, Zhang J, YoshidaTK Abe A, Yokota T (2021) Short DNA/RNA heteroduplex oligonucleotide interacting proteins are key regulators of target gene silencing. *Nucleic Acids Res* 49(9):4864–4876
- Ilaslan E, Sajek MP, Jaruzelska J, KuszZamelczyk K (2022) Emerging Roles of NANOS RNA-Binding Proteins in Cancer. *Int J Mol Sci* 23(16):9408–9408
- Kim W, Kim D, Lee K (2021) RNA-binding proteins and the complex pathophysiology of ALS. *Int J Mol Sci* 22(5):2598–2598
- Noyon C, Roumeguère T, Delporte C, Dufour D, Cortese M, Desmet J, Van AP (2017) The presence of modified nucleosides in extracellular fluids leads to the specific incorporation of 5-chlorocytidine into RNA and modulates the transcription and translation. *Mol Cell Biochem* 429(1–2):59–71
- Kristofich J, Nicchitta C (2023) Signal-noise metrics for RNA binding protein identification reveal broad spectrum protein-RNA interaction frequencies and dynamics. *Nat Commun* 14(1):5868–5868
- Mohammad A, Javad Z, Reza R, Morteza E (2016) rpiCOOL: a tool for In Silico RNA-protein interaction detection using random forest. *J Theor Biol* 402:1–8
- Pan X, Rijnbeek P, Yan J, Shen H (2018) Prediction of RNA-protein sequence and structure binding preferences using deep convolutional and recurrent neural networks. *BMC Genomics* 19(1):1–11
- Zheng F, Chen G, Deng H (2022) Identifying the focuses of hereditary gingival fibromatosis with bioinformatics strategies. *Am J Transl Res* 14(6):3741–3749
- Zhao G, Li P, Qiao X, Han X, Liu ZP (2022) Predicting lncRNA-Protein Interactions by Heterogenous Network Embedding. *Front Genet* 12:814073
- Suresh V, Liu L, Adjeroh D, Zhou X (2015) RPI-Pred: predicting ncRNA-protein interaction using sequence and structural information. *Nucleic Acids Res* 43(3):1370–9
- Gianluca C, Toma T, Fabrizio C, Paolo F, Andrea P (2016) RNAcommender: genome-wide recommendation of RNA-protein interactions. *Bioinformatics (Oxford, England)* 32(23):3627–3634
- Pan X, Fan Y, Jia J, Shen H (2019) Identifying RNA-binding proteins using multi-label deep learning. *Sci China Inform Sci* 62(1):213–215
- Li Z, Zhu J, Xu X, Yao Y (2020) RDense: A protein-RNA binding prediction model based on bidirectional recurrent neural network and densely connected convolutional networks. *IEEE Access* 8:14588–14605
- Zhang Y, Jiang Y, Zhang Q, Liu D (2023) Multi-label learning based on instance correlation and feature redundancy. *Pattern Recognition Letters* 176:123–130
- Zhu C, Zhao J, Hu S, Dong Y, Cao L, Zhou F, Zhou R (2023) A simple multiple-fold correlation-based multi-view multi-label learning. *Neural Comput Appl* 35(14):10407–10420
- Huang J, Qu X, Li G, Qin F, Zheng X, Huang Q (2019) Multi-view multi-label learning with view-label-specific features. *IEEE Access* 7:100979–100992
- Zhao D, Gao Q, Lu Y, Sun D, Cheng Y (2021) Consistency and diversity neural network multi-view multi-label learning. *Knowl-Based Syst* 218:106841
- Xiong B, Chen H, Li T, Yang X (2023) Robust multi-view clustering in latent low-rank space with discrepancy induction. *Appl Intell* 53(20):23655–23674
- Liu G, Ge H, Su S, Wang S (2022) Low-rank tensor multi-view subspace clustering via cooperative regularization. *Multimed Tools Appl* 82(24):38141–38164
- A. Gretton, O. Bousquet, A. Smola, B. Scholkopf, Measuring statistical dependence with Hilbert-Schmidt norms, in: International conference on algorithmic learning theory, Springer, Berlin, Heidelberg, 2005, pp. 63–77.
- Zhao D, Gao Q, Lu Y, Sun D (2021) Two-step multi-view and multi-label learning with missing label via subspace learning. *Applied Soft Computing*(prepublish), 107120
- Cheng Y, Li Q, Wang Y, Zheng W (2022) Multi-view multi-label learning with view feature attention allocation. *Neurocomputing* 501:857–874
- Dassi E, Re A, Leo S (2014) AURA 2: empowering discovery of post-transcriptional networks. *Translation* 2:e27738
- Lorenz R, Stadler P (2020) RNA secondary structures with limited base pair span: exact backtracking and an application. *Genes* 12(1):14
- Cao Y, Fu L (2021) Wu J UFold: fast and accurate RNA secondary structure prediction with deep learning. *Nucleic Acids Res.* <https://doi.org/10.1101/2020.08.17.254896>
- Ito K, Watanabe K, Kitagawa D (2020) The emerging role of ncRNAs and RNA-BINDING PROTEINS IN MITOTIC APPARATUS Formation. *Non-Coding RNA* 6(1):13–13
- Taha M, El-Mageed H, Lee M (2021) DFT study of cyclic glycine-alanine dipeptide binding to gold nanoclusters. *J Mol Graph Model* 103:107823
- Wakabayashi R, Hattori Y, Hosogi S, Toda Y, Ashihara E (2021) A novel dipeptide type inhibitor of the Wnt/ β -catenin pathway suppresses proliferation of acute myelogenous leukemia cells- Science Direct. *Biochem Biophys Res Commun* 535:73–79
- Feng P, Che W, Lin H (2016) Identifying Antioxidant Proteins by Using Optimal Dipeptide Compositions. *Interdiscip Sci Comput Life Sci* 8(2):186–191
- Hayate T, Kana M, Tohru H (2023) Trp-Tyr is a dipeptide structure that potently stimulates GLP-1 secretion in a murine enteroendocrine cell model, identified by comprehensive analysis. *Biochem Biophys Res Commun* 661:28–33
- Reza T, Rasouli F, Mahdi S, Ali M (2022) Convenient synthesis of dipeptide structures in solution phase assisted by a thioaza functionalized magnetic nanocatalyst. *Sci Rep* 12(1):4719–4719
- Liu K, vYe Y, Li S, Tang H (2023) Accurate de novo peptide sequencing using fully convolutional neural networks. *Nat Commun* 14(1):7974–7974
- Baldominos A, Saez Y, Isasi P (2020) On the automated, evolutionary design of neural networks: past, present, and future. *Neural Comput Appl* 32(3):519–545
- Yan Y, Yao X, Wang S, Zhang Y (2021) A survey of computer-aided tumor diagnosis based on convolutional neural network. *Biology* 10(11):1084–1084
- Wu X, Chen Q G, Hu Y, Wang D B, Chang X D, Wang X B, Zhang M L (2019) Multi-View Multi-Label Learning with View-Specific Information Extraction, in: *IJCAL*. 3884–3890
- Ba JL, Kiros JR, Hinton GE (2016) Layer normalization. *arXiv preprint arXiv:1607.06450*
- Ahmed H, Joseph B (2013) Estimation of weighted log partial area under the ROC curve and its application to MicroRNA expression data. *Stat Appl Genet Mol Biol* 12(6):743–55
- Kingma D (2014) Ba J Adam: A Method for Stochastic Optimization. *Comput Sci.* <https://doi.org/10.48550/arXiv.1412.6980>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.