

RESEARCH

Open Access



# CL-ACP: a parallel combination of CNN and LSTM anticancer peptide recognition model

Huiqing Wang, Jian Zhao\*, Hong Zhao, Haolin Li and Juan Wang

\*Correspondence:  
zhaojian0384@link.tyut.edu.cn  
College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China

## Abstract

**Background:** Anticancer peptides are defence substances with innate immune functions that can selectively act on cancer cells without harming normal cells and many studies have been conducted to identify anticancer peptides. In this paper, we introduce the anticancer peptide secondary structures as additional features and propose an effective computational model, CL-ACP, that uses a combined network and attention mechanism to predict anticancer peptides.

**Results:** The CL-ACP model uses secondary structures and original sequences of anticancer peptides to construct the feature space. The long short-term memory and convolutional neural network are used to extract the contextual dependence and local correlations of the feature space. Furthermore, a multi-head self-attention mechanism is used to strengthen the anticancer peptide sequences. Finally, three categories of feature information are classified by cascading. CL-ACP was validated using two types of datasets, anticancer peptide datasets and antimicrobial peptide datasets, on which it achieved good results compared to previous methods. CL-ACP achieved the highest AUC values of 0.935 and 0.972 on the anticancer peptide and antimicrobial peptide datasets, respectively.

**Conclusions:** CL-ACP can effectively recognize antimicrobial peptides, especially anticancer peptides, and the parallel combined neural network structure of CL-ACP does not require complex feature design and high time cost. It is suitable for application as a useful tool in antimicrobial peptide design.

**Keywords:** Anticancer peptide, Secondary structure, Neural network model, Attention mechanism

## Background

Antimicrobial peptides (AMPs) are natural and vital defence substances involved in innate immunity to certain diseases. Anticancer peptides (ACPs) are a class of antimicrobial peptides composed of 10–50 amino acids that have a killing effect on cancer cells [1]. ACPs interact with the cancer cell membrane and effectively destroy its structure, thereby inhibiting the proliferation and growth of cancer cells and inducing apoptosis [2, 3]. Studies have shown that ACPs have clear inhibition and elimination effects on cervical cancer cells, rectal cancer cells, and hepatocellular carcinoma cells [4]. ACPs can



© The Author(s), 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

effectively remove cancer cells from the body and improve the body's immune function to resist invasion by tumour cells. Presently, many ACPs targeting different types of cancer have undergone clinical application [5–8]. Therefore, for the treatment and research of cancer, it is important to determine whether AMPs have antitumour activity, which would provide a new development direction for novel ACPs.

As small molecular peptides, ACPs have specific rules governing their peptide chains, and many researchers have conducted related studies [9, 10]. Hajisharifi et al. introduced peptide sequences as characteristic information to develop a model that successfully predicted ACPs [11]. To compensate for the uniqueness of peptide sequences, Tyagi et al. added information such as the dipeptide compositions and amino acid compositional differences between the N-terminal and C-terminal as features, achieving a higher accuracy than that of Hajisharifi [12]. Chen et al. combined the pseudo amino acid composition, mean chemical shift, and simplified amino acid composition to significantly improve ACP prediction accuracy [13]. However, these methods considered only original sequences and physicochemical properties of amino acids and neglected peptide structural characteristics at the spatial level.

Studies have shown that although the types and sequence length of ACPs vary, most ACPs contain characteristic structural information, such as  $\alpha$ -helix or  $\beta$ -chain structures [14–16], which allow ACPs to act selectively on cancer cells [17]. With the development of protein structural property prediction techniques [18], researchers have begun to examine the role of protein structural information in AMPs. Chen et al. changed the hydrophobicity and amphiphilicity of peptide molecules by amino acid substitution in different regions of the peptide sequences [19, 20], which proved that the secondary structures play a critical role in the antitumour activity of ACPs. Based on the mechanisms of ACPs, Hammami et al. analysed the direct involvement of structural information in the formation of amphiphile side chains of ACPs [21]. They concluded that protein structural characteristics are the basis of the selective action of ACPs on cancer cells. Therefore, protein structural properties contain highly potent local and global features that provide strong evidence for the prediction of ACPs.

Most ACP prediction models are designed based on traditional machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), Naive Bayesian (NB) or statistical models. Chen et al. used SVM to predict ACPs and achieved a high prediction accuracy [13]. Wei et al. used amino acid compositions and other information, combined with SVM to construct 40 submodels to predict ACPs and achieved good results [22]. In addition, some ACP prediction methods are based on the combination of multiple classifiers and the fusion of multiple sequence features [23–25]. Although these methods have made some achievements, the feature construction and extraction process is still tedious and depends on feature design and prior knowledge to some extent. In addition, the algorithm designs of these models are relatively complex, and their performances depend mainly on the number of feature types and the scale of the models.

A neural network can automatically learn advanced representation from raw data, providing a suitable means to solve the problems mentioned above. These networks have been successfully applied in many fields, such as image recognition, machine reading and bioinformatics [26–30]. Yi et al. predicted ACPs by integrating binary profile features and a k-mer sparse matrix with simplified amino acid identification and realised

automatic feature extraction by long short-term memory (LSTM) to address the time-dependence problem in sequences [31]. Yi's work was the first attempt to apply a deep recurrent neural network (RNN) to predict ACPs. Wu et al. mapped peptide sequences to word vectors using the word2vec [32] tool and obtained multiangle features from different sizes of receptor fields using a text-convolutional neural network (text-CNN) [33, 34].

Feature extraction methods vary among different neural networks. LSTM automatically learns dependencies in sequence data through its memory units and gate mechanism. However, the limitation of its learning mechanism causes difficulty in learning local features in sequence data. As compensation, CNN compensate for this limitation by capturing local relevant features in input through convolution kernels. Therefore, a combined network can effectively improve model prediction abilities [35]. Wang et al. proposed a hybrid deep learning model for miRNA prediction based on integrating CNN and bidirectional long short-term memory (BiLSTM) [27], which improved the prediction quality by capturing complex local features of nucleotides via CNN and long-term interdependence between nucleotides by BiLSTM. Guo et al. developed DeepACLSTM by combining an asymmetric CNN and a BiLSTM network to predict protein secondary structures effectively [28]. Therefore, the combination of CNN and LSTM can simultaneously focus on the local spatial and long-term dependence information in the original data, effectively reducing information loss and improving the ACP prediction performance.

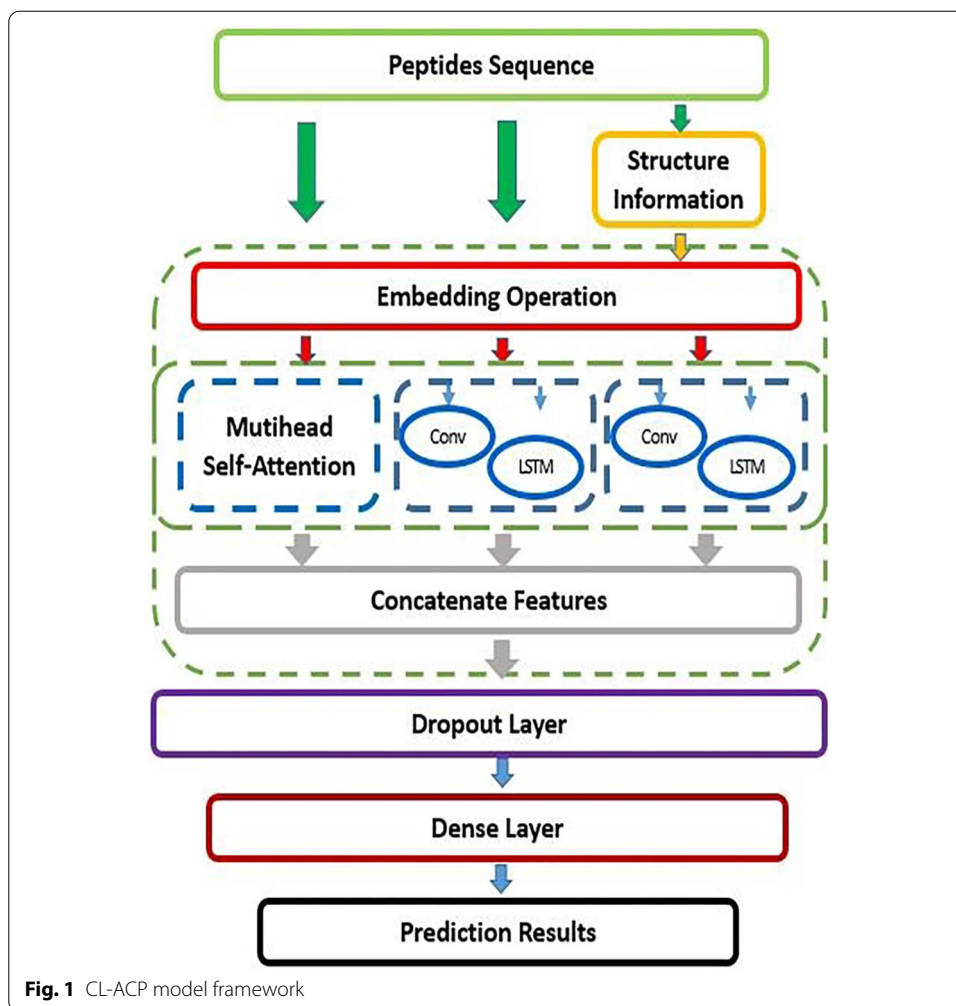
Based on the above problems, we effectively combined CNN and LSTM to propose a new neural network model, CL-ACP, for ACP recognition (Fig. 1). The CL-ACP model constructed a feature space from two aspects—ACP sequences and secondary structures. In addition, it used multi-head self-attention [36] to enhance peptide sequence representations. Finally, the CNN and LSTM parallel combined network model was applied to effectively capture the temporal and spatial feature information of peptide sequences and structural characteristics. To evaluate the predictive performance of CL-ACP, we conducted an experiment on the datasets of Yi et al. and compared CL-ACP with existing methods. The fivefold cross-validation experimental results show that CL-ACP can automatically learn the effective characteristics of complex correlation patterns in the data and further identify ACPs effectively.

## Methods

In this paper, we introduced the secondary structure information of ACPs and the multi-head self-attention mechanism and proposed a parallel combination of CNN and LSTM to predict ACPs. We evaluated the model predictive performance on the benchmark datasets ACP736 and ACP240 and compared it with machine learning and neural network models.

## Datasets

In this paper, we selected the ACP736 and ACP240 datasets collected by Yi et al. as the benchmark datasets and constructed a new dataset, ACP539. These ACP datasets contain the most experimentally verified positive samples in the literature thus far, as shown in Table 1.



**Table 1** Summary of the scale of AMP datasets supplied in this work

Peptides	Number	Positive	Negative
AAP	214	107	107
ABP	1600	800	800
ACP240	240	129	111
ACP736	736	375	361
ACP539	539	189	350
CPP	740	370	370
QSP	400	200	200
PBP	160	80	80

Positive samples from the ACP736 and ACP240 datasets were experimentally verified, and negative samples were AMPs without anticancer functions. The benchmark datasets were all passed through the CD-HIT [37] tool to remove redundant sequences, with the threshold set to 0.9 to ensure that there were no duplicates. We also constructed a supplementary dataset, ACP539, to further verify the performance

of the model. We generated positive samples by utilizing the previous works [12, 13, 22, 38] and the Antimicrobial peptide date (APD), which contain a total of 793 experimentally verified ACPs. For the collection of negative samples, we utilized AMPs and random peptides from the previous works [12, 38], wherein AMPs had been extracted from several databases including, the antimicrobial peptide date (APD), collection of anti-microbial peptides (CAMP) and database of anuran defense peptides (DADP), for which no anticancer activity has been reported in the literature. For random peptides, we assume that they are non-ACPs. Subsequently, we eliminated the duplicate samples of ACP739 and ACP240 from all the samples. To avoid performance over-estimation introduced by the homology bias, the remaining peptide sequences in both the positive and negative datasets with more than 90% sequence identity were removed using the CD-HIT program with the threshold set at 0.9. Finally, we obtained ACP539 dataset that contained 189 positive samples and 350 negative samples. The data volume ratio of positive and negative samples in ACP539 is approximately 1:2.

To verify the model generalizability, we selected other AMP datasets [39–45] to verify the prediction ability of the model for short peptide data. The AMP datasets are also shown in Table 1.

#### Extraction of peptide secondary structures

ACPs can target cancer cells because of their special mechanism of membrane cleavage. The main mechanism representations are the bucket plate model, blanket model, and annular pore model [46–48]. In the disintegration of the cancer cell membrane, the bucket plate and annular pore models form ionic channels penetrating the plasma membrane, which require at least three peptide molecules with a particular secondary structure. Relevant studies have proven that many ACPs exert their anticancer effects through the bucket plate and annular hole models. For example, melittin, which was discovered in 1994, has anti-inflammatory, analgesic, antibacterial, anti-HIV, antitumour, and other pharmacological activities. It also has a broad spectrum of antitumour activities, including against human hepatocellular carcinoma, leukaemia, and breast cancer [49, 50]. ACPs from *Litoria raniformis* have strong effects on leukaemia, lung cancer, colon cancer, and other diseases [51]. *Magainins* from *Xenopus laevis* can destroy the cell membrane of human cervical cancer cells through annular pore action [52]. Therefore, the action mechanism of most ACPs is based on their secondary structures, and the accurate extraction of peptide secondary structures thus plays a vital role.

To accurately extract peptide secondary structures, we analysed the performances of various secondary structure predictors and selected the SPIDER3 [18]. We also eliminated four peptide sequences that lacked secondary structure from the ACP736 dataset.

#### Representation of the features

The input feature space of CL-ACP was composed of three kinds of feature information, including peptide original sequences and secondary structures encoded by one-hot coding [53] and peptide sequences enhanced by the multi-head self-attention mechanism.

### ***One-hot coding and multi-head self-attention mechanism***

ACPs are usually composed of 5–40 amino acids of 20 different types. Each amino acid was encoded using one-hot coding of length 20, representing 20 dimensions corresponding to the 20 amino acids. Thus, an original sequence of length  $L$  was encoded as a vector representation of dimension  $L \times 20$ . The structure information included three exponents:  $\alpha$ -helix  $P$  (H),  $\beta$ -chain  $P$  (C), and  $\gamma$ -strand  $P$  (E). Therefore, we obtained a vector representation of  $L \times 3$ -dimensional protein structure property information for motifs of length  $L$ .

Although one-hot coding is a simple and effective coding method, its premise is to assume that coding elements are independent of each other; however, residues in peptide sequences are not independent of each other. Moreover, the proportions of certain residues in peptide sequences are relatively high, which indicates that these residues play relatively important roles in antitumour activity. Obviously, using one-hot coding alone cannot reflect the degree of distance between elements and fully characterize sequences. To strengthen the expression of peptide sequences and extract high-quality feature information, we introduced a multi-head self-attention mechanism to focus on the relatively important residues in the sequences. The multi-head self-attention mechanism is a variant of the attention mechanism, which has been widely used in tasks such as machine reading, text summarization, and image description. Compared with the self-attention mechanism, multiple heads can form multiple subspaces, allowing the attention mechanism to evaluate the importance of residues from different subspaces [54]. To the best of our knowledge, this paper introduces the multi-head self-attention mechanism into peptide sequences coding for the first time. The input sequence vectors are calculated by the multi-head self-attention mechanism to obtain new characterization vectors, allowing the model input to represent more context information. Moreover, the multi-head self-attention mechanism associates any two amino acid residues in the sequences by calculating the similarity between the elements without limiting the distance between them. Therefore, it does not need to fix the length of peptide sequences and can dynamically adjust the weights of different amino acids in sequences to preserve complete feature information.

In addition, to select a suitable number of heads in the multi-head self-attention mechanism, we set the numbers of heads as 1, 2, 4, 8 and 16 and evaluated the proposed model performance, as shown in Additional file 2: Table S1. Compared with the model with a multi-head self-attention mechanism, the comprehensive performance of the model using the common self-attention mechanism (only 1 head) was relatively poor, which indicates that the multi-head self-attention mechanism can comprehensively evaluate the importance of residues in sequences from multiple perspectives. Moreover, the number of heads is an important hyperparameter in the multi-head self-attention mechanism, and the number of heads is not necessarily proportional to the effect of the model [55]. The results in Table S1 show that when the number of heads increased from 2 to 16, the performance of the model decreased. After we added regularization to each head [56], this phenomenon improved to some extent. As shown in Additional file 3: Table S2, when the number of heads was large, more redundant subspaces were generated, leading to high similarity between heads. Although heads regularization can increase the diversity among multiple attention heads, but it also increased time cost

when the number of heads was large. Therefore, we selected 2 as the optimal number of heads to avoid serious similarity problems among heads, and the resulting model had the best comprehensive performance and a low time cost.

The multi-head self-attention mechanism contains multiple identical self-attention structures, and each attention head uses different initialization parameters to learn different attention spaces. The self-attention mechanism uses scaled dot-product attention to calculate similarity scores. The calculation of similarity scores is shown in Eq. 1.

$$Score = \frac{Query * Key^T}{\sqrt{d_w}} \quad (1)$$

where *Query* represents an amino acid and *Key* represents each amino acid in a peptide sequence,  $d_w$  represents the word vector dimension, and *Score* represents the similarity between the evaluated amino acid and each amino acid in a peptide sequence.

The similarity scores are then normalized by softmax and converted into a probability distribution with the sum of weights equals to 1, thus highlighting the correlation between the two elements, as shown in Eq. 2.

$$\alpha = Softmax(Score) \quad (2)$$

Finally, the attention score of the current amino acid is obtained by multiplying the normalized similarity score by the current amino acid. The calculation process is shown in Eq. 3.

$$Attention(Query, Key, Value) = \alpha \cdot Value \quad (3)$$

In the multi-head self-attention mechanism, *Value* represents the same value as *Query*. *Query*, *Key*, and *Value* are mapped to multiple parallel heads for repeated attention calculations through different parameter matrices. Each head can process different information, and the calculation process is shown in Eq. 4.

$$head_i = Attention\left(Query \cdot W_i^{Query}, Key \cdot W_i^{Key}, Value \cdot W_i^{Value}\right) \quad (4)$$

The weight parameters  $W_i^{Query} \in R^{d_w/h \times d_w}$ ,  $W_i^{Key} \in R^{d_w/h \times d_w}$  and  $W_i^{Value} \in R^{d_w/h \times d_w}$  are learnable parameters for linear calculation.

The multi-head self-attention mechanism can process different parts of a sequence to extract richer sequence features and combine the results of multiple attention operations into vector stitching, as shown in Eq. 5.

$$MultiHead(Query, Key, Value) = W^M [head_1, head_2, \dots, head_h] \quad (5)$$

where  $h$  is the number of parallel heads, and  $W^M \in R^{d_w \times d_w}$  is used to connect several attention results, which can maintain the original output dimension. The final calculation process of the multi-head self-attention mechanism is shown in Eq. 6.

$$S' = \{w'_1, w'_2, \dots, w'_n\} = MultiHead(S, S, S) \quad (6)$$

In the multi-head self-attention mechanism, *Query*, *Key*, and *Value* all represent the original sequence  $S$ , and  $S'$  is the output.  $w_i$  is a new output of amino acids in the

sequence calculated by the multi-head self-attention mechanism, which contains richer sequence information. The final  $S' \in \mathbb{R}^{n^{\text{dw}}}$  is a new representation of a peptide sequence.

### The framework of CL-ACP

In this paper, we propose CL-ACP to effectively predict potential ACPs. To prevent cross-talk between the peptide original sequences and secondary structures, we used two sets of parallel CNN and LSTM composite structures to extract the features respectively, and then combined the extracted features with the enhanced sequence features of multi-head self-attention to obtain advanced features. Finally, the advanced features were input to a fully connected layer to predict ACPs.

### Convolutional neural network

Due to the different characteristics of the sequence information carried by the original sequences and secondary structures, we introduced two sets of single-layer two-dimensional convolutional neural networks to extract features from the two types of information, with each branch consisting of a convolutional layer and rectifying linear unit (ReLU). The convolutional layer can obtain local features by convolving the sequences encoding space and rectifying linear elements to sparse the convolution layer output. Due to the short length of the peptide sequences, we did not pool the features after convolution, thus preserving the feature integrity.

The convolutional layer of peptide sequence and structure information consists of 300 and 150 convolution kernels, and the sizes of the convolution kernels are  $5 \times 5$  and  $3 \times 3$ , respectively. The convolution kernels are convolved with the input peptide sequences to output a series of weight numbers indicating the convolution kernel matching degree with each window. The inner product of the output matrix of the convolutional layer is shown in Eq. 7.

$$C_l = \sum_{b=1}^8 \sum_{j=1}^{20} K_{b,j} X_{b,j+1} \quad (7)$$

where  $X \in \{0, 1\}^{T \times L}$  is the input matrix after encoding,  $T$  is the number of different elements in the sequences,  $K$  is the convolution kernel of  $5 \times 5$  or  $3 \times 3$ , and  $C_l$  eliminates the negative matches in the convolutional layer and maintains positive matches by ReLU processing. Finally, the original sequences and the convolution branch of the secondary structures are stitched and input to the fully connected layer.

### Long short-term memory network

To identify category information hidden in the original and secondary structure sequences, we added LSTM, which incorporates long-term dependence information to aid in prediction. As the LSTM scans each element of the input sequences, first, the forget gate determines what information to discard based on the previous input. Then, the input gate determines how much new information should be added to the cell state to update the current state value. Finally, the output gate arranges the values to determine which values to output. These gating operations enable the LSTM to automatically



extract and learn all relevant information from the sequences that is useful for the overall classification task.

The numbers of storage units in the LSTM hidden layer of CL-ACP are 45 and 20 for feature extraction in the original sequences and secondary structures, respectively. The gating mechanism of LSTM and the update state of each step are shown in Eqs. 8–12.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (8)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (9)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (10)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (11)$$

$$h_t = o_t \tanh(c_t) \quad (12)$$

where  $\sigma$  is the sigmoid function, and  $i_t, f_t, o_t$  and  $c_t$  represent the input gate, forget gate, output gate and cell activation vector, respectively.  $X, h$ , and  $c$  represent input vectors, hidden states, and memory locations, respectively.  $W$  and  $b$  are weights and offsets that need to be learned. We selected the sigmoid function as the activation function, and Eq. 13 shows the calculation process.

$$\sigma = \text{sigmoid}(x) = \frac{1}{(1 + e^{-x})} \quad (13)$$

Simultaneously, we chose the corresponding binary cross-entropy loss function in the binary classification tasks to adjust the neural network. Equation 14 is the definition of the loss function.

$$\text{logloss}(t, p) = -((1 - p) \times \log(1 - p) + t \times \log(p)) \quad (14)$$

where  $P$  and  $T$  represent the predicted and target values of the model, respectively. Finally, the *Adam* optimizer commonly adopted in the neural network was used to update the network weight.

Considering the limited ACP data and the complex network model structure that may lead to overfitting problems, we used dropout [57] and early stopping regularization methods to optimize the model and reduce the model parameters. The loss rate  $p$  was set to 0.45, and dropout was only used during training. In addition, the CNN, LSTM and the multi-head self-attention mechanism were combined in parallel to reduce the number of network layers, model complexity and time consumption and maintain the richness of feature dimensions.

#### **Performance evaluation criteria**

We considered several statistical measures to evaluate the performance of the proposed model and other comparative models, including accuracy (Acc), sensitivity (Sens), specificity (Spec), precision (Prec) and Matthew's correlation coefficient (Mcc). The definitions are shown in Eqs. 15–19.

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \quad (15)$$

$$Sens = \frac{TP}{TP + FN} \quad (16)$$

$$Spec = \frac{TN}{TN + FP} \quad (17)$$

$$Prec = \frac{TP}{TP + FP} \quad (18)$$

$$Mcc = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FN) \times (TN + FP)}} \quad (19)$$

where  $TP$  denotes true positives,  $TN$  denotes true negatives,  $FP$  denotes false positives, and  $FN$  denotes false negatives.  $Acc$  measures the total number of correctly identified ACPs and non-ACPs.  $Sens$  evaluates the accuracy of the model in identifying ACPs.  $Spec$  assesses the ability of predictors to recognize non-ACPs.  $Prec$  evaluates the number of correctly predicted ACPs in the identified data. When positive and negative samples are unbalanced,  $Mcc$  can measure the classification quality of a classifier. In addition, the area under the receiver operating characteristic (ROC) curve (AUC) was used to measure the overall performance of the model. The higher the values of these indicators are, the better the overall performance of the model.

## Results

### The performances of CL-ACP on the benchmark datasets

To evaluate the CL-ACP model ACP predictive ability, we conducted fivefold cross-validation on the benchmark datasets ACP736 and ACP240. Detailed information about the fivefold cross-validation experiment on the benchmark datasets is shown in Table 2.

The average  $Acc$  of fivefold cross-validation on ACP736 was 83.83%, the average  $Sens$  was 82.93%, the average  $Spec$  was 84.76%, the mean  $Prec$  was 85.15% and the average  $Mcc$  was 67.86%. CL-ACP had an AUC of 0.909, as shown in Fig. 2a. On dataset ACP240, the average  $Acc$  of fivefold cross-validation was 87.92%, the average  $Sens$  was 90.74%, the mean  $Spec$  was 84.72%, the average  $Prec$  was 88.41%, the average  $Mcc$  was 76.56%, the AUC was 0.935, and the ROC curve is shown in Fig. 2b. The CL-ACP model achieved a good prediction performance on both datasets, especially on the smaller ACP240 dataset, indicating that CL-ACP has good ACP prediction ability and robustness.

To further verify the CL-ACP model performance, we conducted experiments on the newly constructed ACP539 dataset. The ratio of positive samples to negative samples in the ACP539 dataset is 1:2. Table 2 shows the fivefold cross-validation results of CL-ACP on ACP539. As Table 2 shows, on dataset ACP539, the average  $Acc$  of fivefold cross-validation was 84.41%, the average  $Prec$  was 78.46%, the average  $Sens$  was 77.48%, the average  $Spec$  was 88.23%, and the average  $Mcc$  was 65.98%. The AUC was 0.921, and the ROC curve is shown in Fig. 2c. The results of fivefold cross-validation show that the

**Table 2** The 5-fold cross-validation details in the ACP datasets

Fold	Acc(%)	Sens(%)	Spec(%)	Prec(%)	Mcc(%)
ACP736					
1	83.78	84.00	83.56	84.00	67.56
2	83.00	85.33	80.55	82.05	66.00
3	85.03	82.67	87.50	87.32	70.19
4	84.35	86.67	81.94	83.33	68.73
5	82.99	76.00	90.28	89.06	66.82
Average	83.83	82.93	84.76	85.15	67.86
ACP240					
1	89.58	96.15	81.82	86.21	79.45
2	81.25	92.31	68.18	77.42	63.02
3	89.58	84.62	95.45	95.65	79.86
4	91.67	84.62	99.89	99.89	84.62
5	87.50	96.00	78.26	82.76	75.86
Average	87.92	90.74	84.76	88.41	76.56
ACP539					
1	87.04	81.58	90.00	81.58	71.58
2	80.56	65.79	88.57	75.76	56.36
3	83.33	74.37	88.41	78.38	63.52
4	89.81	84.62	92.75	86.84	77.82
5	81.31	81.08	81.43	69.77	60.64
Average	84.41	77.48	88.23	78.46	65.98

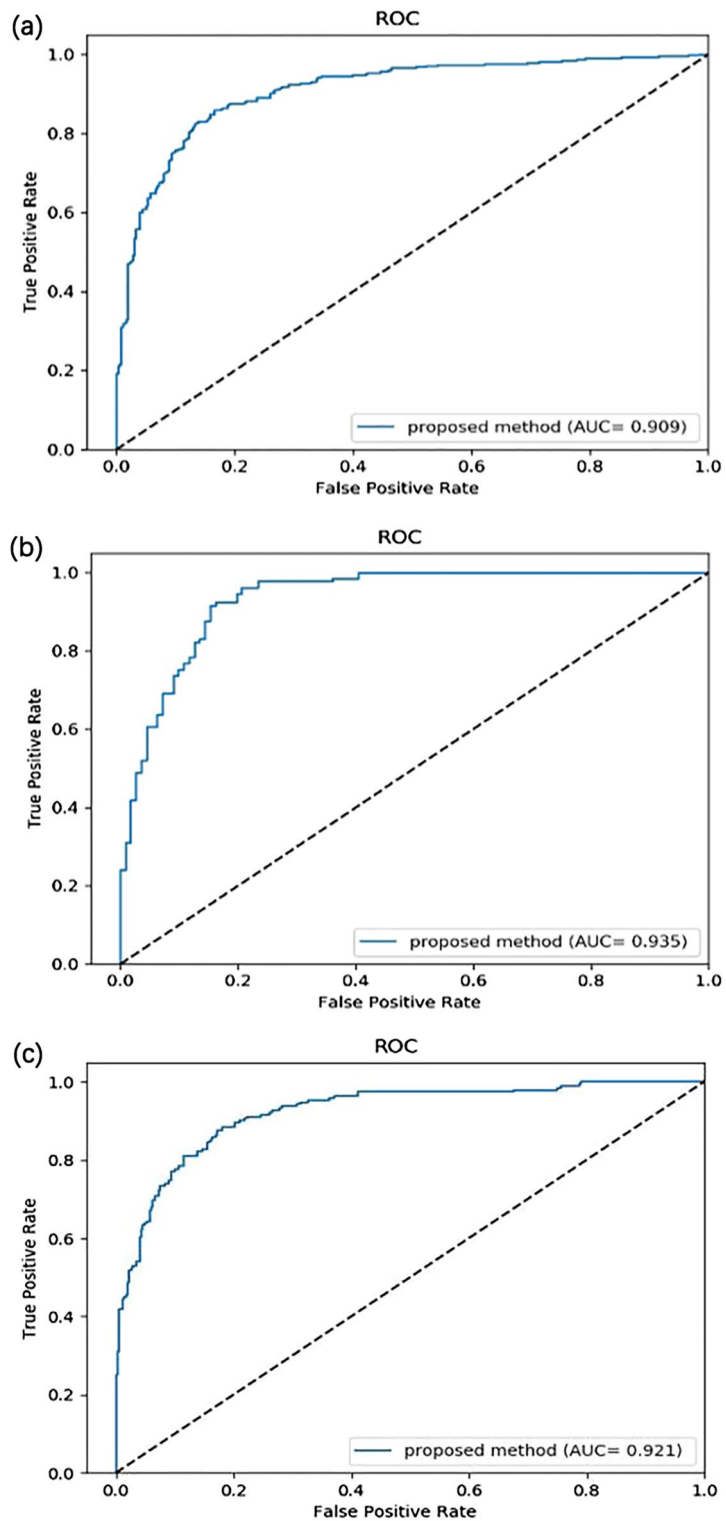
accuracy, specificity, and ROC value of CL-ACP on the ACP539 dataset were promising. However, the accuracy and sensitivity were lower than those on the two benchmark datasets, mainly because the ACP539 dataset was slightly unbalanced and contained noise in the negative samples.

In addition, to analyse the robustness of the proposed model, we further performed k-fold cross-validation, setting  $k=6, 8$  and  $10$ . The results are shown in Additional file 4: Table S3. From Additional file 4: Table S3, we can see that there was no significant fluctuation among the index values with different values of  $k$ . The congruence of k-fold cross-validation results indicates the promising performance and robustness of CL-ACP.

### Ablation experiments

To verify the vital role of protein structural characteristic information in predicting ACPs and each CL-ACP component's necessity, we used fivefold cross-validation to conduct ablation experiments on the benchmark datasets. The procedure mainly included the introduction of two-dimensional CNN, the use of multi-head self-attention mechanism, and the introduction of peptide secondary structures. The experimental results are shown in Table 3. The baseline model used the original sequence information of the motifs as input and applied LSTM to extract features and predict ACPs.

The introduction of two-dimensional convolution dramatically improved all indicators of the baseline model for ACP736 and ACP240, as shown in the fourth column of Table 3. The improvements show that the two-dimensional convolutional network can capture spatial feature information in peptide sequences. Compared with the model using LSTM alone, the CNN and LSTM parallel combined structure can fully



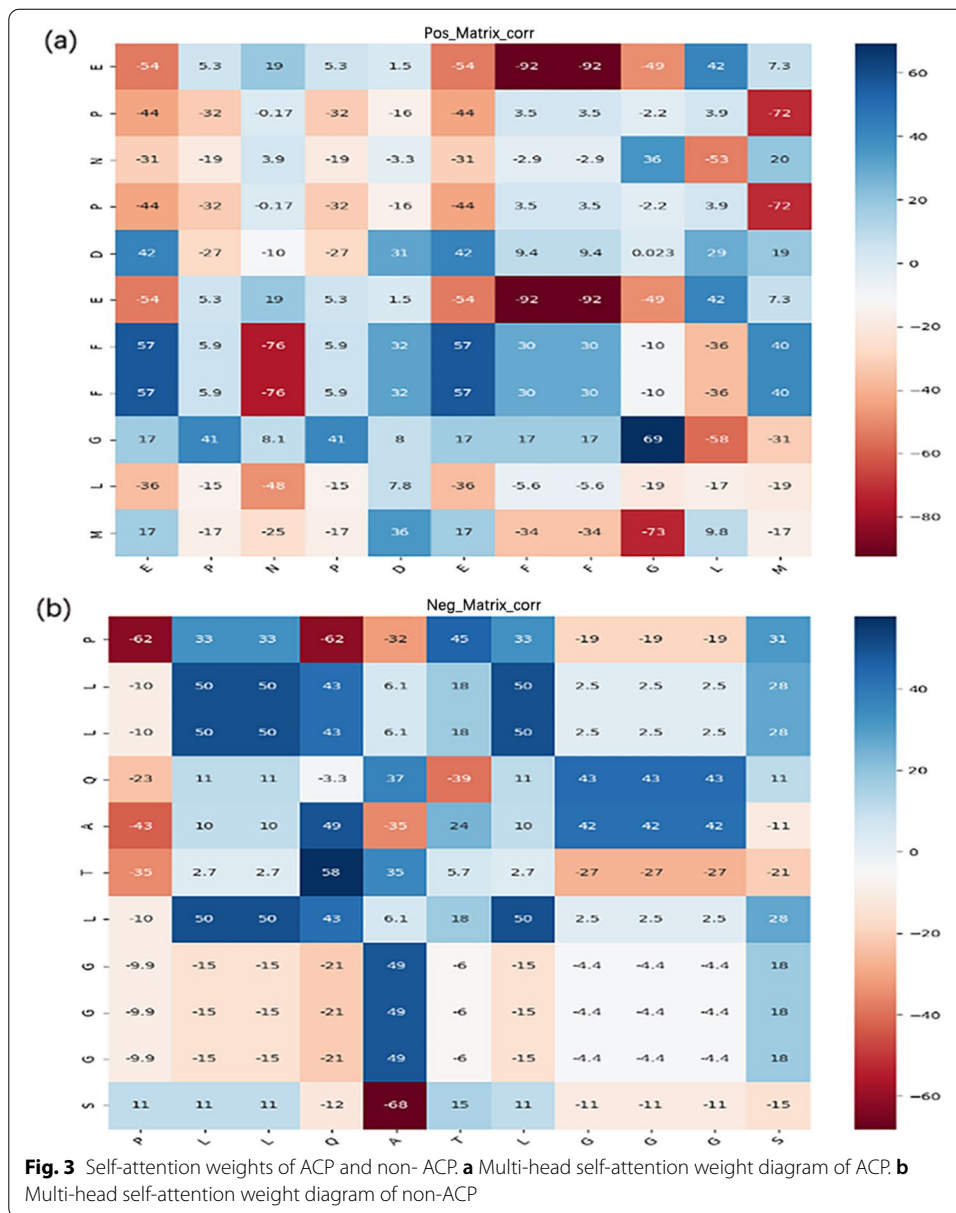
**Fig. 2** ROC curves of CL-ACP on the ACP datasets. **a** ROC curve of the CL-ACP model on ACP736. **b** ROC curve of the CL-ACP model on ACP240. **c** ROC curve of the CL-ACP model on ACP539

**Table 3** Ablation experiment results on the benchmark datasets

Component	LSTM	✓	✓	✓	✓	✓
	CNN		✓	✓	✓	✓
	Multi-SA (skip-connection)			✓		✓
	Structure information				✓	✓
	Multi-SA(cascade)				✓	
ACP736	Acc (%)	78.94	82.88	83.55	83.55	<b>83.83</b>
	Sens (%)	78.66	80.06	82.26	83.20	<b>82.93</b>
	Spec (%)	79.21	80.62	82.82	83.94	<b>84.76</b>
	Prec (%)	80.04	82.12	84.00	84.54	<b>85.15</b>
	Mcc (%)	58.12	65.84	67.37	67.27	<b>67.86</b>
	AUC	0.862	0.897	0.904	0.900	<b>0.909</b>
ACP240	Acc (%)	83.33	85.66	86.25	85.83	<b>87.92</b>
	Sens (%)	88.43	89.93	90.00	86.89	<b>90.74</b>
	Spec (%)	77.55	81.00	82.05	84.74	<b>84.76</b>
	Prec (%)	82.77	84.01	86.34	87.86	<b>88.41</b>
	Mcc (%)	67.83	73.03	73.18	72.23	<b>76.56</b>
	AUC	0.867	0.915	0.920	0.914	<b>0.935</b>

Bold indicates the results of the final model (CL-ACP)

extract sequence feature information from multiple angles and obtain a higher-quality abstract representation. In the fifth column of Table 3, we found that the introduction of a multi-head self-attention mechanism improved multiple indicators of the model. These improvements show that the multi-head self-attention mechanism can focus the model on more critical residue information in peptide sequences and strengthen the peptide character representations, thereby enhancing the representation ability of the network. Yi's work indicated that the contents of amino acids Cys (C), Phe (F), Gly (G), His (H), Ile (I), Asn (N), Ser (S), and Tyr (Y) accounted for a significantly higher proportion of ACPs than non-ACPs. However, the amino acids Glu (E), Leu (L), Met (M), Gln (Q), Arg (R), and Trp (W) accounted for a greater proportion in non-ACPs than ACPs. Visualization shows that the multi-head self-attention mechanism could effectively capture essential features of peptide sequences, as shown in Fig. 3. In the multi-head self-attention visualization diagram of randomly selected an ACP, the weights of amino acids Phe and Gly with respect to the whole sequence (the blue row in the matrix) are larger than those of other amino acids, as shown in Fig. 3a. Similarly, in randomly selected a non-ACP, the weight of the amino acid Leu in the whole sequence was higher than those of other amino acids, as shown in Fig. 3b. These visualizations are consistent with the findings of the previous study [31]. Considering that ACPs are relatively short, the model combination will not only yield more abundant characteristic information but also increase the input noise and useless information. Therefore, we added the multi-head self-attention mechanism to the fully connected layer by skip connection. The skip connection reduced the parameters and time cost of the model and improved the model efficiency. As shown in the sixth and seventh columns of Table 3, the performance of cascading multiple self-attention mechanism into a parallel network was poorer than that of using a skip connection. In addition,



as shown in Additional file 5: Table S4, the number of parameters in the cascading mode was 2.16 times that of the skip connection mode, and the running time of the former was approximately 4 times that of the skip connection mode. These results show that the introduction of skip connection is effective. As shown in the seventh column of Table 3, the addition of peptide secondary structures caused a significant increase in the model indicators. Compared with using original sequence information alone, the values of Acc, Sens, Spec, Prec, Mcc and ROC on ACP736 increased by 0.15%, 0.77%, 0.94%, 1.39%, 0.21% and 0.5%, respectively. Similarly, the values of Acc, Sens, Spec, Prec, Mcc and ROC on ACP240 increased by 1.67%, 0.74%, 2.71%, 2.07%, 3.38%, and 1.5%, respectively. Compared with the model without secondary structure information, all model indicators significantly improved after secondary

structure information was added, with most improved by approximately 2%. These improvements indicate that ACP secondary structures contain critical information about antitumour activity. The introduction of ACP secondary structures enriches the feature space and facilitates ACP identification.

### **Comparison with other classification models on benchmark datasets**

To further assess the prediction performance of the proposed model, we compared it on the benchmark datasets with other models, including SVM, RF, NB, AntiCP2.0, ACP-DL, PTPD and iACP-DRLF models [24, 25]. The first four models are machine learning methods, and the last three models are neural network methods. Among them, AntiCP 2.0 and iACP-DRLF only provide web servers or trained models; thus, we used the data from each fold in the fivefold cross-validation to evaluate the trained model and took the average value as the result. This verification method was also used in related works [58, 59]. In addition, iACP\_DRLF provided two trained models, with one trained by a dataset composed of ACPs and non-ACPs and the other trained by a dataset composed of ACPs and AMPs, which were denoted as iACP\_DRLF(a) and iACP\_DRLF(b), respectively. Similar to iACP\_DRLE, AntiCP2.0 also provided two trained models, AntiCP2.0(a) and AntiCP2.0(b). We validated all four models on our ACP datasets, and the detailed results are shown in Additional file 6: Table S5. Considering the similar compositions of the datasets, we chose iACP\_DRLF(a) and AntiCP2.0(a) for comparison with the proposed method using the baseline datasets and iACP\_DRLF(b) and AntiCP2.0(b) for comparison with the proposed method using ACP539. All methods were evaluated with the same evaluation indicators. The comparison results are shown in Table 4. On the dataset ACP736 dataset, the specificity of CL-ACP was lower than that of the NB model, but the other indicators of CL-ACP were the highest. The NB sensitivity and other indicators were lower than those of the other models. On dataset ACP240, all the indicators of CL-ACP were the highest. Overall, CL-ACP achieved better performance. Especially on ACP240, which contains a small amount of data, the CL-ACP model showed a better performance than the machine learning models. Although the results of AntiCP2.0 were better than those of other machine learning methods, its performance is based on feature construction, including location preference, which is a complex construction process and has certain limitations. The comparison results show that CL-ACP can extract high-quality features better than the machine learning models using the same features and coding methods. In addition, CL-ACP does not require manual feature design, and it has better robustness even when the amount of data is small.

Among the neural network models, ACP-DL, PTPD and iACP-DRLF were selected for comparison. We used fivefold cross-validation and the same evaluation indicators to evaluate the models. Considering that we used the datasets collected by Yi et al. as the benchmark datasets, we used ACP-DL as the main comparison method. On the ACP240 dataset, all indicators of CL-ACP were higher than those of the comparison models. Except for the sensitivity indicator, all indicators showed more than 4% improvement compared with ACP-DL, and the ROC value was significantly improved. On the ACP736 dataset, the sensitivity and ROC values were improved by approximately 2% compared with ACP-DL, and the other indicators were improved by more than 3%. The experimental results of PTPD and ACP-DL were worse than

**Table 4** Performance of comparison models and CL-ACP on the ACP datasets

Dataset	Methods	Acc (%)	Sens (%)	Spec (%)	Prec (%)	Mcc (%)	AUC
ACP736	SVM	80.97	81.86	80.06	81.06	61.97	0.810
	RF	81.52	81.06	82.00	82.44	63.08	0.815
	NB	75.41	<b>90.13</b>	60.14	70.34	52.87	0.751
	PTPD	80.97	81.86	80.06	81.06	61.97	0.884
	ACP-DL	80.81	81.39	80.22	81.00	61.67	0.890
	AntiCP2.0	81.21	87.59	74.85	79.13	62.87	0.843
	iACP-DRLF	80.72	86.68	74.24	78.74	61.38	0.859
	CL-ACP	<b>83.83</b>	82.93	<b>84.76</b>	<b>85.15</b>	<b>67.86</b>	<b>0.909</b>
ACP240	SVM	79.58	83.01	75.61	80.22	59.59	0.793
	RF	81.66	84.58	78.30	82.05	63.48	0.814
	NB	70.83	88.40	50.43	67.35	43.01	0.694
	PTPD	79.58	83.01	75.61	80.22	59.59	0.784
	ACP-DL	83.75	88.40	78.45	83.16	68.29	0.903
	AntiCP2.0	84.00	88.64	76.16	84.18	71.19	0.894
	iACP-DRLF	84.11	88.01	74.35	84.03	70.35	0.903
	CL-ACP	<b>87.92</b>	<b>90.74</b>	<b>84.76</b>	<b>88.41</b>	<b>76.56</b>	<b>0.935</b>
ACP539	SVM	76.80	38.71	<b>97.70</b>	<b>89.88</b>	48.34	0.682
	RF	76.80	45.46	93.96	79.93	46.88	0.698
	NB	75.41	55.13	90.14	78.34	50.87	0.606
	PTPD	74.94	37.09	95.70	82.65	42.82	0.740
	ACP-DL	72.72	60.08	80.34	65.43	41.37	0.831
	AntiCP2.0	82.38	69.25	95.00	85.27	60.09	0.881
	iACP-DRLF	82.56	65.21	92.00	82.16	60.99	0.882
	CL-ACP	<b>84.41</b>	<b>77.48</b>	88.23	78.46	<b>65.98</b>	<b>0.921</b>

Bold indicates the highest value

those of CL-ACP, mainly because CL-ACP's LSTM component can capture important sequence information. In contrast to the ACP-DL model, the CNN component of CL-ACP can capture ACP spatial information. Therefore, CL-ACP can be combined with neural networks with different structures to obtain sufficient characteristic information, which can be well applied to the identification and prediction of ACPs. The iACP-DRLF method performed well on both benchmark datasets. This mainly because it used two sequence embedding techniques and deep learning to characterize embedded sequences. However, sequence embedding required a high time cost, and the verification time of iACP-DRLF was the highest among all comparison methods (please see Additional file 5: Table S4).

To further verify the model prediction performance, we conducted a comparative experiment on ACP539 dataset, and the verification results are shown in Table 4. The average Acc of CL-ACP on the ACP539 dataset was 84.41%, the average Sens was 77.48%, the average Mcc was 65.98%, and the ROC value was 0.921, all of which were the highest among all comparison models. For the evaluation of unbalanced data, the Mcc value can be used to measure the classifier's quality, and the ROC value can measure overall model performance. It can be seen from Table 4 and Additional file 1: Figure S1 that the Mcc and ROC values of the machine learning models were lower among the comparison methods, and the ROC values of the neural network models



were higher than those of the machine learning models. Moreover, the Mcc and ROC values of CL-ACP were the highest, indicating that CL-ACP still had a better performance when the data were slightly unbalanced.

It is worth noting that the specificity of each model was higher than its respective sensitivity for the ACP539 dataset because sensitivity and specificity are antagonistic to a certain extent [58]. This phenomenon also shows that these models missed some true positive samples. However, CL-ACP maintained high specificity with the highest sensitivity and better recognized true positive and true negative data. There are two main reasons. The first reason is that the ACP539 was unbalanced. The number of negative samples was about twice that of positive samples, leading to more false negatives predicted by the model. Another reason was that since the negative samples in the ACP539 dataset were composed of AMPs, which shared high similarity with ACPs [60]. This data-trained model may result in large false negatives for prediction. Especially when the dataset is small, the model will overfit the data and generate more false negative data, resulting in low sensitivity. As shown in Table 4, machine learning models have low sensitivity and high specificity because sensitivity and specificity are antagonistic. CL-ACP adopted a variety of regularization methods to avoid overfitting ACP539, especially negative samples, so fewer false negatives were generated compared with the machine learning models and obtained higher sensitivity. Moreover, sensitivity and specificity are antagonistic, so the true positive and false positive data of CL-ACP were both higher, resulting in lower specificity and accuracy.

#### The performances of CL-ACP on antimicrobial peptide datasets

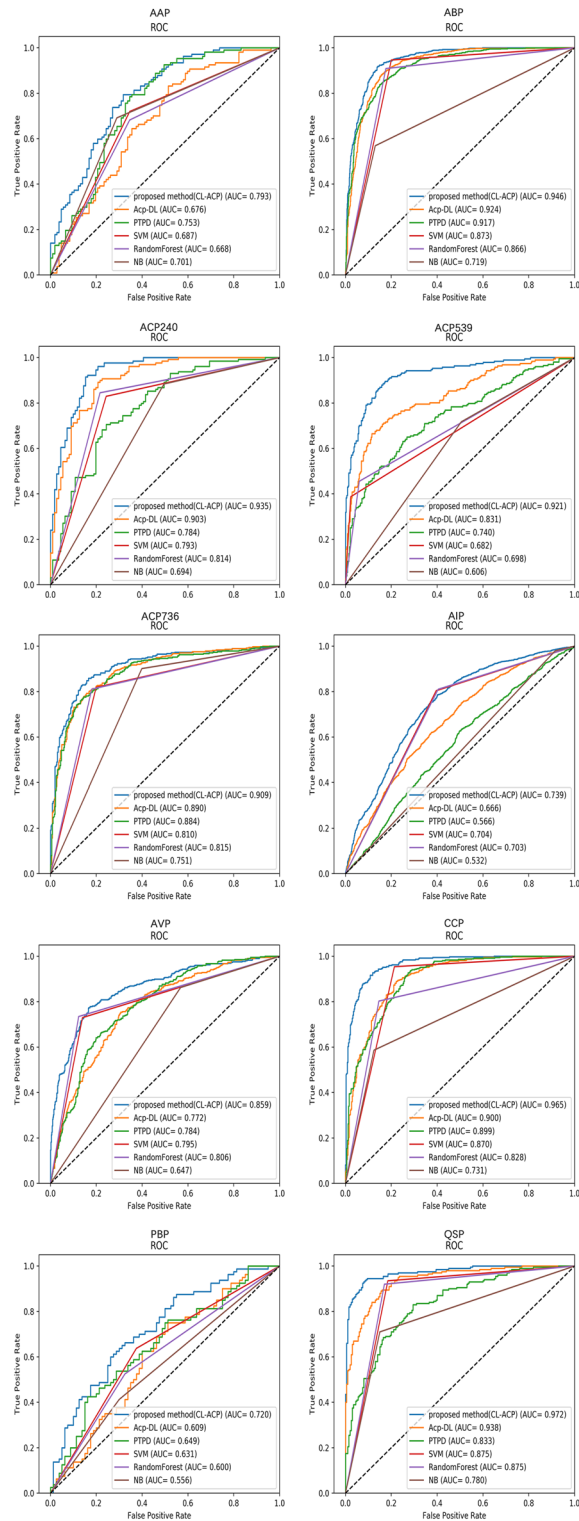
To verify the generalization ability of CL-ACP on short peptide data, we used other AMP datasets collected from APD and previous related works, including AAP, ABP, ACP, AIP, AVP, CPP, QSP, and PBP. Since most ACP data were contained in the ACP736 and ACP240 datasets, this procedure did not include testing on ACP datasets.

Figure 4 shows the ROC curves for various models. The AUC values of CL-ACP on ABP, CPP, and QSP were 0.945, 0.965, and 0.973, respectively, which were comparable to those on the ACP benchmark datasets, and these results were achieved by using only peptide sequence-level and structural information, as well as the advanced features automatically extracted by the effective architecture of CL-ACP.

The areas under the ROC curve of the AMP datasets AAP, ABP, ACP736, ACP240, ACP539, AIP, AVP, CPP, PBP and QSP for CL-ACP were 0.793, 0.946, 0.909, 0.935, 0.921, 0.739, 0.859, 0.965, 0.720 and 0.972, respectively. Compared with other models, the AUC values of CL-ACP were the highest, which further confirms the model generalizability and potential for the use of CL-ACP in AMP predictions. We will explore the incorporation of additional feature information to improve the model general AMP prediction performance in future work.

#### Discussion

Experiments showed that CL-ACP had a good predictive performance and robustness relative to the comparison methods. In experiments on other AMPs, CL-ACP also showed better generalizability.



**Fig. 4** The ROC curves of the antimicrobial peptide datasets for CL-ACP and the comparison models

The performance of CL-ACP benefits from several major factors. (1) The peptide secondary structures contain key information about antitumour activity of ACPs, and the introduction of peptide secondary structures improves the feature richness. (2) The introduction of the parallel combined network model can fully extract local features and long-term dependence information from the feature space, effectively reduce the model complexity and prevent the problem of overfitting. (3) The introduction of the multi-head self-attention mechanism strengthens the representation of sequence information, as also indicated by its visualization. This is also the first attempt to introduce attention mechanism into ACP prediction.

## Conclusions

In this paper, we proposed an ACP prediction model constructed with a hybrid CNN and LSTM, termed CL-ACP. It used a multi-head self-attention mechanism to enhance the peptide sequence expression and incorporated peptide secondary structure characteristics to better characterize the feature space. CNN networks were used to obtain the local hidden characteristics of ACPs. The sequence dependence information of amino acid residues was captured by LSTM, which reduced information loss. Finally, the extracted advanced features were input to the fully connected layer for prediction. Comparative experiments on benchmark datasets showed that CL-ACP had a better predictive performance than existing prediction models, improving ACP identification. Comparative experiments on the ACP539 dataset showed that even when the negative data contained noise, the performance of CL-ACP was better than those of comparison models, indicating good robustness of CL-ACP. Comparative experiments on AMPs data showed that CL-ACP is not limited to the prediction of ACPs but can also automatically extract practical features. CL-ACP can learn efficient abstract representations of short peptide data to discover novel ACPs and AMPs, providing helpful information for drug development to treat various cancers and other diseases.

Although CL-ACP had a good performance in predicting ACPs, it still lacked meaningful biological explanation. For example, the multi-head self-attention mechanism can enhance weights of the essential residues in peptide sequences. However, the biological rationale is unclear and will therefore be a focus of future work. Simultaneously, we will consider effective feature fusion methods and model structures, such as capsule networks [58], to improve the performance of the model.

## Abbreviations

AAC: Amino acid composition; Acc: Accuracy; AAP: Antiangiogenic peptide; ABP: Antibacterial peptide; ACP: Anticancer peptide; AIP: Anti-inflammatory peptide; Arg: Arginine; Asn: Asparagine; AVP: Antiviral peptide; AMP: Antimicrobial peptide; AUC: The area under ROC curve; ACP-DL: A deep learning long short-term memory model to predict anticancer peptides using high-efficiency feature representation; AntiCP 2.0: An updated model for predicting anticancer peptides; BiLSTM: Bi-directional long short-term memory network; CPP: Cell penetrating peptide; CNN: Convolution neural network; CL-ACP: A parallel combination of CNN and LSTM Anticancer Peptide Recognition Model; Cys: Cysteine; DeepACLSTM: Deep asymmetric convolutional long short-term memory neural models for protein secondary structure prediction; FN: False negative; FP: False positive; Gly: Glycine; Glu: Glutamic acid; Gln: Glutamine; His: Histidine; Ile: Isoleucine; Leu: Leucine; iACP-DRLF: Anticancer peptides prediction with deep representation learning features; LSTM: Long short-term memory network; Met: Methionine; MCC: Matthew's correlation coefficient; NB: Naive bayesian model; Phe: Phenylalanine; PBP: Polystyrene surface binding peptide; Prec: Precision; PSI-BLAST: Position specific iterated BLAST; PSSM: Position specific scoring matrices; PTPD: Prediction of therapeutic peptide by deep learning and word2Vec; QSP: Quorum sensing peptide; RAAAC: Reduced amino acid alphabet; RAAC: Reduced amino acid composition; ReLU: Rectified linear unit; RF: Random forest; ROC: Receiver operating characteristic; Ser: Serine; SPIDER3: Capturing non-local interactions by long short term memory bidirectional recurrent neural networks for improving prediction of protein

secondary structure; Sens: Sensitivity; Spec: Specificity; SVM: Support vector machine; Text-CNN: Text-convolutional neural network; TN: True negative; TP: True positive; Tyr: Tyrosine; Trp: Tryptophan.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-021-04433-9>.

**Additional file 1. Figure S1.** **a** ROC curves of ACP240 dataset on CL-ACP and comparison methods. **b** ROC curves of ACP736 dataset on CL-ACP and comparison methods. **c** ROC curves of ACP539 dataset on CL-ACP and comparison methods.

**Additional file 2. Table S1.** The results of CL-ACP using multiple numbers of head on the benchmark datasets.

**Additional file 3. Table S2.** Comparison results of using regularized and non-regularized multi-head self-attention mechanism.

**Additional file 4. Table S3.** k-fold cross-validation results of the proposed model CL-ACP on the benchmark datasets.

**Additional file 5. Table S4.** The amount of parameters and time spent on the 5-fold cross-validation of the model on the ACP datasets (The 5-fold cross-validation time of SVM, NB and RF and the validation time of AntiCP2.0 are very fast, so they are ignored).

**Additional file 6. Table S5.** The performance of two models of iACP\_DRLF and AntiCP2.0 on ACP datasets.

## Acknowledgements

Not applicable.

## Authors' contributions

HW and JZ conceived the algorithm, carried out analyses, prepared the datasets, carried out experiments, and wrote the manuscript. Other authors designed, performed, and analysed experiments and wrote the manuscript. All authors read and approved the final manuscript.

## Funding

This research was funded by the National Natural Science Foundation of China, grant number 61976150, and the Key Research and Development Plan of Shanxi Province, Grant Number 201903D121151.

## Availability of data and materials

The datasets supporting the conclusions of this article are ACP datasets available from the previous studies and the Anti-microbial Peptide Database (APD) [12, 13, 22, 38]. The Amp datasets are from the studies of Balachandran et al. [39–45]. All data generated or analysed during this study can be obtained from <https://github.com/zjlyn1314/CL-ACP/tree/main/Datasets>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 10 May 2021 Accepted: 5 October 2021

Published online: 20 October 2021

## References

1. Domingues MM, Felício M, Gonalves S: Antimicrobial peptides: effect on bacterial cells: methods and protocols. *Atomic Force Microsc.*; 2019.
2. Shibue T, Weinberg RA: EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat Rev Clin Oncol.* 2017.
3. Barras D, Widmann C: Promises of apoptosis-inducing peptides in cancer therapeutics. *Curr Pharm Biotechnol.* 2011; 12(8).
4. Pérez-Peinado C, Dias SA, Domingues MM, Benfield AH, Freire JM, Rádis-Baptista G, Gaspar D, Castanho M, Craik DJ, Henriques ST. Mechanisms of bacterial membrane permeabilization by crotalidicin (Ctn) and its fragment Ctn(15–34), antimicrobial peptides from rattlesnake venom. *J Biol Chem.* 2018;293(5):1536.
5. Zafar S, Beg S, Panda SK, Rahman M, Ahmad FJ: Novel therapeutic interventions in cancer treatment using protein and peptide-based targeted smart systems. *Semin Cancer Biol.* 2019.
6. Sah BNP, Vasiljevic T, McKechnie S, Donkor ON: Identification of anticancer peptides from bovine milk proteins and their potential roles in management of cancer: a critical review. *Compr Rev Food Sci Food Saf* 2015; 14(2).

7. Araste F, Abnous K, Hashemi M, Taghdisi SM, Ramezani M, Alibolandi M. Peptide-based targeted therapeutics: focus on cancer treatment. *J Controlled Release* 2018; 292:141–62.
8. Teerasak E-K, Pennapa T, Sittiruk R, Ladda M, Pramote C. Prediction of anticancer peptides against MCF-7 breast cancer cells from the peptidomes of *Achatina fulica* mucus fractions. *Comput Struct Biotechnol J*. 2016, 14.
9. Findlay F, Proudfoot L, Stevens C, Barlow PG. Cationic host defense peptides; novel antimicrobial therapeutics against Category A pathogens and emerging infections. *Pathog Global Health*. 2016;110(4):1–11.
10. Melicherčík P, Nešuta O, Čefovský V. Antimicrobial peptides for topical treatment of osteomyelitis and implant-related infections: study in the spongy bone. *Pharmaceuticals* 2018.
11. Hajisharifi Z, Piryai Ee M, Beigi MM, Behbahani M, Mohabatkar H. Predicting anticancer peptides with Chou's pseudo amino acid composition and investigating their mutagenicity via Ames test. *J Theor Biol*. 2014;341:34–40.
12. Tyagi A, Kapoor P, Kumar R, Chaudhary K, Gautam A, Raghava GPS. In silico models for designing and discovering novel anticancer peptides. *Sci Rep*. 2013, 3(1).
13. Wei C, Hui D, Pengmian F, Hao L, Kuo-Chen C. iACP: a sequence-based tool for identifying anticancer peptides. *Oncotarget* 2016, 7(13).
14. Haney EF, Mansour SC, Hancock R. Antimicrobial peptides: an introduction. *Antimicrobial Peptides*; 2017.
15. Huang Y, Qi F, Yan Q, Hao X, Chen Y: Alpha-helical cationic anticancer peptides: a promising candidate for novel anticancer drugs. *Mini Rev Med Chem*. 2015, 15(1).
16. Zandsalimi F, Talaei S, Ahari MN, Aghamiri S, Zadeh ZZ. Antimicrobial peptides: a promising strategy for lung cancer drug discovery? *Expert Opin Drug Discov*. 2020;1:1–12.
17. Nhung D, Günther W, Lisa G, Mario S, Beate K, Christian KH, Monika S-K. Cationic membrane-active peptides - anticancer and antifungal activity as well as penetration into human skin. *Exp Dermatol*. 2014, 23(5).
18. Rhys H, Yang Y, Kuldip P, Zhou Y. Capturing non-local interactions by long short term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers, and solvent accessibility. *Bioinformatics*. 2017;18:18.
19. Chen Y, Guarnieri MT, Vasil AI, Vasil ML, Mant CT, Hodges RS. Role of peptide hydrophobicity in the mechanism of action of alpha-helical antimicrobial peptides. *Antimicrobial Agents Chemother*. 2007; 51(4).
20. Chen Y, Vasil AI, Rehaume L, Mant CT, Burns JL, Vasil ML, Hancock R, Hodges RS. Comparison of biophysical and biologic properties of  $\alpha$ -helical enantiomeric antimicrobial peptides. *Chem Biol Drug Des*. 2006.
21. Hammami R, Fliss I. Current trends in antimicrobial agent research: chemo- and bioinformatics approaches. *Drug Discovery Today* 2010; 15(13).
22. Leyi W, Chen Z, Huangrong C, Jiangning S, Ran S. ACPred-FL: a sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides. *Bioinformatics (Oxford, England)* 2018; 34(23).
23. Akbar S, Hayat M, Iqbal M, Jan MA. iACP-GAEnsC: evolutionary genetic algorithm based ensemble classification of anticancer peptides by utilizing hybrid feature space. *Artif Intell Med*. 2017, 79.
24. Agrawal P, Bhagat D, Mahalwal M, Sharma N, Raghava GP. AntiCP 2.0: an updated model for predicting anticancer peptides. *Brief Bioinform*. 2021, 22(3).
25. Lv Z, Cui F, Zou Q, Zhang L, Xu L. Anticancer peptides prediction with deep representation learning features. *Brief Bioinform*. 2021.
26. Fenglin L, Minghui W, Yu L, Xing-Ming Z, Ao L: DeepPhos: prediction of protein phosphorylation sites with deep learning. *Bioinformatics (Oxford, England)* 2019, 35(16).
27. Huiqing W, Yue M, Chunlin D, Chun L, Jingjing W, Dan L: CL-PMI: A precursor MicroRNA identification method based on convolutional and long short-term memory networks. *Front Genet*. 2019, 10.
28. Guo Y, Li W, Wang B, Liu H, Zhou D. DeepACLSTM: deep asymmetric convolutional long short-term memory neural models for protein secondary structure prediction. *BioMed Central* 2019, 20(1).
29. Fei H, Rui W, Jiagen L, Lingling B, Dong X, Xiaowei Z. Large-scale prediction of protein ubiquitination sites using a multimodal deep architecture. *BMC Syst Biol*. 2018, 12(Suppl 6).
30. Zheng X, Fu X, Wang K, Wang M. Deep neural networks for human microRNA precursor detection. *BMC Bioinform*. 2020, 21(1).
31. Yi H-C, You Z-H, Zhou X, Cheng L, Li X, Jiang T-H, Chen Z-H. ACP-DL: a deep learning long short-term memory model to predict anticancer peptides using high efficiency feature representation. *Mol Ther - Nucleic Acids* 2019.
32. Zhang D, Xu H, Su Z, Xu Y. Chinese comments sentiment classification based on word2vec and SVM perf. *Expert Syst Appl* 2015; 42(4).
33. Chuanyan W, Rui G, Yusen Z, Yang DM. PTPD: predicting therapeutic peptides by deep learning and word2vec. *BMC Bioinform* 2019; 20(1).
34. Kim Y. Convolutional neural networks for sentence classification. *Eprint Arxiv* 2014.
35. Jeff D, Anne HL, Marcus R, Subhashini V, Sergio G, Kate S, Trevor D. Long-term recurrent convolutional networks for visual recognition and description. *IEEE Trans Pattern Anal Mach Intell* 2017; 39(4).
36. Lin Z, Feng M, Santos C, Yu M, Xiang B, Zhou B, Bengio Y. A structured self-attentive sentence embedding. 2017.
37. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006, 22(13).
38. Boopathi V, Subramaniyam S, Malik A, Lee G, Yang DC. mACPpred: a support vector machine-based meta-predictor for identification of anticancer peptides. *Int J Mol Sci* 2019.
39. Tramontano A, Ramaprasad AE, Singh S, Gajendra P, Venkatesan S. AntiAngioPred: a server for prediction of anti-angiogenic peptides. *PLoS ONE* 2015; 10(9):e0136990
40. Sharma BK, Sneha L, Gps R. Analysis and prediction of antibacterial peptides. *BMC Bioinform*. 2007;8(1):1–10.
41. Balachandran M, Shin TH, Kim MO, Gwang L. AIPpred: sequence-based prediction of anti-inflammatory peptides using random forest. *Front Pharmacol*. 2018, 9:276.
42. Nishant T, Abid Q, Manoj K. AVPPred: collection and prediction of highly effective antiviral peptides. *Nucleic Acids Res*. 2012;W1:W199–204.
43. Wei L, Xing P, Ran S, Shi G, Quan Z: CPPred-RF: a sequence-based predictor for identifying cell-penetrating peptides and their uptake efficiency. *J Proteome Res*. 2017, 16(5).

44. Akanksha R, Kumar GA, Manoj K, Lukasz K. Prediction and analysis of quorum sensing peptides based on sequence features. *Plos One* 2015, 10(3):e0120066.
45. Li N, Kang J, Jiang L, He B, Hao L, Huang J. PSBinder: a web service for predicting polystyrene surface-binding peptides. *BioMed Res Int*. 2017; (2017-12-27), 2017:1–5.
46. Prashant K, Jayachandran K, Suzana S. Antimicrobial peptides: diversity, mechanism of action and strategies to improve the activity and biocompatibility in vivo. *Biomolecules* 2018; 8(1):4.
47. Lombardi L, Stellato MI, Oliva R, Falanga A, Galdiero M, Petraccone L, D'Errico G, Santis AD, Galdiero S, Vecchio PD. Antimicrobial peptides at work: interaction of myxiniidin and its mutant WMR with lipid bilayers mimicking the *P. aeruginosa* and *E. coli* membranes. *Sci Rep*. 2017; 7:44425.
48. Amos S, Vermeer LS, Ferguson PM, Kozłowska J, Davy M, Bui TT, Drake AF, Lorenz CD, Mason AJ. Antimicrobial peptide potency is facilitated by greater conformational flexibility when binding to gram-negative bacterial inner membranes. *Sci Rep*. 2016;6:37639.
49. Berthony D, Di YP. Antimicrobial peptides with selective antitumor mechanisms: prospect for anticancer applications. *Oncotarget*. 2017;8(28):46635–51.
50. Wimley WC. How does Melittin Permeabilize membranes? *Biophys J*. 2018, 114(2).
51. Sani MA, Separovic F: How membrane-active peptides get into lipid membranes. *Acc Chem Res*. 2016:1130–1138.
52. Lehmann J, Retz M, Sidhu SS, Suttman H, Sell M, Paulsen F, Harder J, Unteregger G, Stöckle M. Antitumor activity of the antimicrobial peptide Magainin II against Bladder Cancer Cell Lines. *Eur Urol*. 2006, 50(1)
53. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng*. 2009;21(9):1263–84.
54. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. *arXiv* 2017.
55. Voita E, Talbot D, Moiseev F, Sennrich R, Titov I. Analyzing multi-head self-attention: specialized heads do the heavy lifting, the rest can be pruned. In *Meeting of the Association for Computational Linguistics*: 2019.
56. Jian L, Tu Z, Tong Z. Multi-head attention with disagreement regularization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*: 2018.
57. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(1):1929–58.
58. Duolin W, Yanchun L, Dong X. Capsule network for protein post-translational modification site prediction. *Bioinformatics* 2019; 35(14).
59. Ning Q, Zhao X, Bao L, Ma Z, Zhao X. Detecting Succinylation sites from protein sequences using ensemble support vector machine. *BMC Bioinform*. 2018;19(1):237.
60. Vijayakumar S, Ptv L. ACP: a web server for prediction and design of anti-cancer peptides. *Int J Pept Res Ther*. 2015;21(1):99–106.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

