

Article

MPMABP: A CNN and Bi-LSTM-Based Method for Predicting Multi-Activities of Bioactive Peptides

You Li ¹, Xueyong Li ¹, Yuewu Liu ², Yuhua Yao ³ and Guohua Huang ^{1,*} 

¹ School of Electrical Engineering, Shaoyang University, Shaoyang 422000, China; youli9609@163.com (Y.L.); xylihnsyu@163.com (X.L.)

² College of Information and Intelligence, Hunan Agricultural University, Changsha 410128, China; yuewuliu@whu.edu.cn

³ School of Mathematics and Statistics, Hainan Normal University, Haikou 571158, China; yaoyuhua2288@163.com

* Correspondence: 3280@hnsyu.edu.cn

Abstract: Bioactive peptides are typically small functional peptides with 2–20 amino acid residues and play versatile roles in metabolic and biological processes. Bioactive peptides are multi-functional, so it is vastly challenging to accurately detect all their functions simultaneously. We proposed a convolution neural network (CNN) and bi-directional long short-term memory (Bi-LSTM)-based deep learning method (called MPMABP) for recognizing multi-activities of bioactive peptides. The MPMABP stacked five CNNs at different scales, and used the residual network to preserve the information from loss. The empirical results showed that the MPMABP is superior to the state-of-the-art methods. Analysis on the distribution of amino acids indicated that the lysine preferred to appear in the anti-cancer peptide, the leucine in the anti-diabetic peptide, and the proline in the anti-hypertensive peptide. The method and analysis are beneficial to recognize multi-activities of bioactive peptides.

Keywords: bioactive peptide; convolution neural network; deep learning; long short-term memory; multi-label issues



Citation: Li, Y.; Li, X.; Liu, Y.; Yao, Y.; Huang, G. MPMABP: A CNN and Bi-LSTM-Based Method for Predicting Multi-Activities of Bioactive Peptides. *Pharmaceuticals* **2022**, *15*, 707. <https://doi.org/10.3390/ph15060707>

Academic Editors: Zhongjie Liang and Fanwang Meng

Received: 22 April 2022

Accepted: 30 May 2022

Published: 3 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bioactive peptides are small protein fragments, that generally contains 2–20 amino acid residues [1,2]. The bioactive peptides remain inactive when they are encrypted in the precursor protein, while the bioactive peptides are active once they are released from the precursor protein. Bioactive peptides are not only distributed widely in foods, plants, and animals [3], but also play versatile roles in the metabolic and biological processes. For example, some bioactive peptides were reported to resist the action of digestion peptidases [4], some were proved to be of anti-bacterial and anti-oxidant activity [1], while some had immunomodulatory and anti-cancer activities [5]. Therefore, it is of great importance to accurately identify activities of bioactive peptides in at least two respects: (1) it is helpful to promote understanding of the mechanism of bioactive peptides; and (2) it is fundamental to develop new natural foods and drugs to meet the demands for safety and health.

The bioactive peptides are organic substances comprising amino acids joined by covalent bonds. According to mode of action, the bioactive peptides are classified as the anti-microbial peptide (AMP), the anti-diabetic peptide (ADP), the anti-hypertensive peptide (AHP), the anti-inflammatory peptide (AIP), the anti-cancer peptide (ACP), the anti-oxidant peptide, the immunomodulatory peptide, and so on [3,6]. Since the AMP have anti-bacterial, anti-fungal, or anti-viral properties, they are also called host defense peptides (HDPs) which are distributed widely in the innate immune response system. When the host was invaded by foreigners such as a virus or bacterium, the AMP was induced to destroy or kill the invading foreigners with the membrane damage mechanism [1,7]. The AHP

has anti-hypertensive activity by mainly targeting the inhibiting angiotensin-converting enzyme (ACE) which exerts a crucial role through the renin angiotensin system (RAS) in the regulation of blood pressure and electrolyte balance [8]. The AIP is endogenous, and is able to inhibit antigen-specific T(H)1-driven responses [9]. The ACP is able to inhibit tumor cell proliferation or migration, or block the formation of tumor blood vessels, and is less likely to induce drug resistance [10]. The anti-oxidant peptide has anti-oxidant activity, which poses influences on the cells by removing free radicals, inhibiting lipid peroxidation, and interacting with metal ions [11].

Due to importance to food and health of humans, great efforts have been made to develop methods or techniques to identify activities of bioactive peptides over the past decades. Currently, there are two ways: the classical and the bioinformatics approaches [12]. Although the former can accurately identify bioactive peptides, it is too time-consuming, expensive, and labor-intensive. Especially for vast volumes of bioactive peptides, this approach is insufficient for the task. On the contrary, the bioinformatics approaches can remedy the limitation. Hundreds of bioinformatics approaches have been developed to address a wide range of issues in the field of molecular biology [13–18], including stress response protein identification [19], RNA modification identification [20–22], post-translational protein modification identification [23–28], and genomic island detection [18,29,30]. However, these bioinformatics approaches rely heavily on the accumulation of the known samples and the sophisticated design of algorithms. Recently, a large volume of annotated bioactive peptides has been deposited in many public databases, which facilitates greatly development of the bioinformatics approaches. These bioactive peptides databases include the ADP database BioDADPep [31], the BioPepDB [32], the therapeutic peptide database SATPdb [33], the ACPs database CancerPPD [34], the AHPs database AHTPDB [35], the anti-parasitic peptides database ParaPep [36], the database of peptide sequences PepBank [37], the Peptipedia [38], the anti-inflammatory database PreAIP [39], the target-unrelated peptides database TUPDB [40], and the anti-tubercular peptides AntiTbPdb [41]. For example, the BioPep-UWM, a database of bioactive peptides, deposited more than 3000 active peptides [42].

The annotation and the collection of bioactive peptides pave solid material foundations for computationally identifying its activities, while efficiencies depend heavily on design of the features and the learning algorithms. Since the bioactive peptides are made up of the amino acid residues, the information related to composition, the peptide chain, and the hydrophobic/hydrophilic nature of the amino acid would be influential for its activities. To the best of our knowledge, there are no less than 20 bioinformatics approaches for predicting activities of bioactive peptides [43–53]. Khatun et al. [39] developed a random forest-based method (PreAIP) to computationally recognize the AIP, which employed primary sequence as well as evolutionary and structural information. Manavalan et al. [54] presented an extremely randomized tree-based method for anti-tubercular peptides prediction, which utilized only sequence information. Usmani et al. [55] proposed a support-vector-machine-based method AntiTBpred, and Khatun et al. [56] presented a SVM and random-forest-based combination model for anti-tubercular peptides prediction. Zhang et al. [57] developed a classifier-chain-based ensemble learning method for anti-inflammatory peptides prediction. Hasan et al. [58] generated 66 optimal baseline models by combining 11 different encodings and six different classifiers, and then built a representation learning-based method for identifying neuropeptides. The methods for anti-angiogenic peptide prediction include the SVM-based AntiAngioPred [59], the generalized linear model [60], the AntAngioCOOL [61], the random-forest-based TargetAntiAngio [62], the convolution-network-based AAPred-CNN [63]. The PIP-EL [64], the ProInfam [65], and the ProIn-Fuse [66] are three methods for proinflammatory peptide predictions, while the HemoPI [67], the HemoPred [68], and HLPpred-Fuse [52] are three methods for hemolytic peptide prediction. The methods for discriminating between ACPs and non-ACPs include AntiCP [69,70], iACP [71], ACPP [72], iACP-GAEnsC [73], MLACP [74], TargetACP [75], ACPred [76], ACPred-FL [77], ACPred-Fuse [78], ACP-DL [79], and iACP-FSCM [80], while

the methods for distinguishing therapeutic peptides from non-therapeutic peptides include PEPred-Suite [81], PTPD [82], PPTPP [83], and PreTP-EL [84]. Most bioinformatics approaches suffered from the small number of bioactive peptides. He et al. [85] pioneered mutual information meta learning to address small samples of bioactive peptides prediction, while Zhang et al. [48] employed the pre-trained natural language model BERT [86] to predict the AMP.

All the previous methods are only suitable for differentiating specific activity of bioactive peptides. In practice, a bioactive peptide might simultaneously consist of multi-activities. Obviously, to computationally identify activities of bioactive peptides is a multi-label and multi-class issue. Recently, Tang et al. [87] presented a convolution neural network (CNN) and gated recurrent unit (GRU)-based deep learning method (called MLBP [87]) for multi-activities of bioactive peptide prediction. This is a promising avenue for identifying actual activities of bioactive peptides. For the deep learning method, the ability to learn a representation would depend on what components it adopted and ways of combining components. The MLBP [87] is a deep learning architecture with three different scale parallel CNNs followed by the GRU [88]. The CNN is the most widely used architecture of neural work especially in the field of image processing, which is capable of characterizing local properties [89,90], while the long short-term memory (LSTM) is a popular architecture to capture semantics in the context of text sequences [91]. The structure of the CNN followed directly by the LSTM would absorb merits of both components. However, the MLBP attached the GRU to three-scale CNNs, which causes multi-scale information loss. In addition, with the increase in the depth of the deep neural network, the original information about sequences would drop seriously. On the basis of the analysis above, we improved the MLBP [87] in two respects. One is that we used multi-branch CNNs, each followed directly by the semantic architecture to improve representation of peptides. The other is to use the residual network architecture to ensure no loss of information about peptides in the forward process. In addition, we replaced the GRU by the Bi-LSTM. The proposed method is abbreviated to MPMABP. The empirical experiments showed that the MPMABP outperformed the MLBP [87].

2. Results and Discussion

2.1. Optimization of Parameters

In the MPMABP, there are many user-defined hyper-parameters such as the embedding dimension, the learning rate, the dropout, and the pooling size which are influential in its predictive performance. We separated 20 percent from the training set as the validation set to investigate influence. We tested four embedding dimensions (50, 100, 150, and 200), four learning rates (0.1, 0.01, 0.005, 0.003, and 0.001), four dropouts (0.1, 0.2, 0.3, and 0.5), and four pooling sizes (2, 3, 4, and 5). As shown in Figure 1, the accuracies over various values of the same type of hyper-parameter are relatively stable, exhibiting only a slight difference between the various parameter values. According to the general experience, we set the embedding dimension to 100, the learning rate to 0.001, the dropout to 0.5, and the pooling size to 3, respectively. The details of other hyper-parameters in the MPMABP are listed in Table 1.

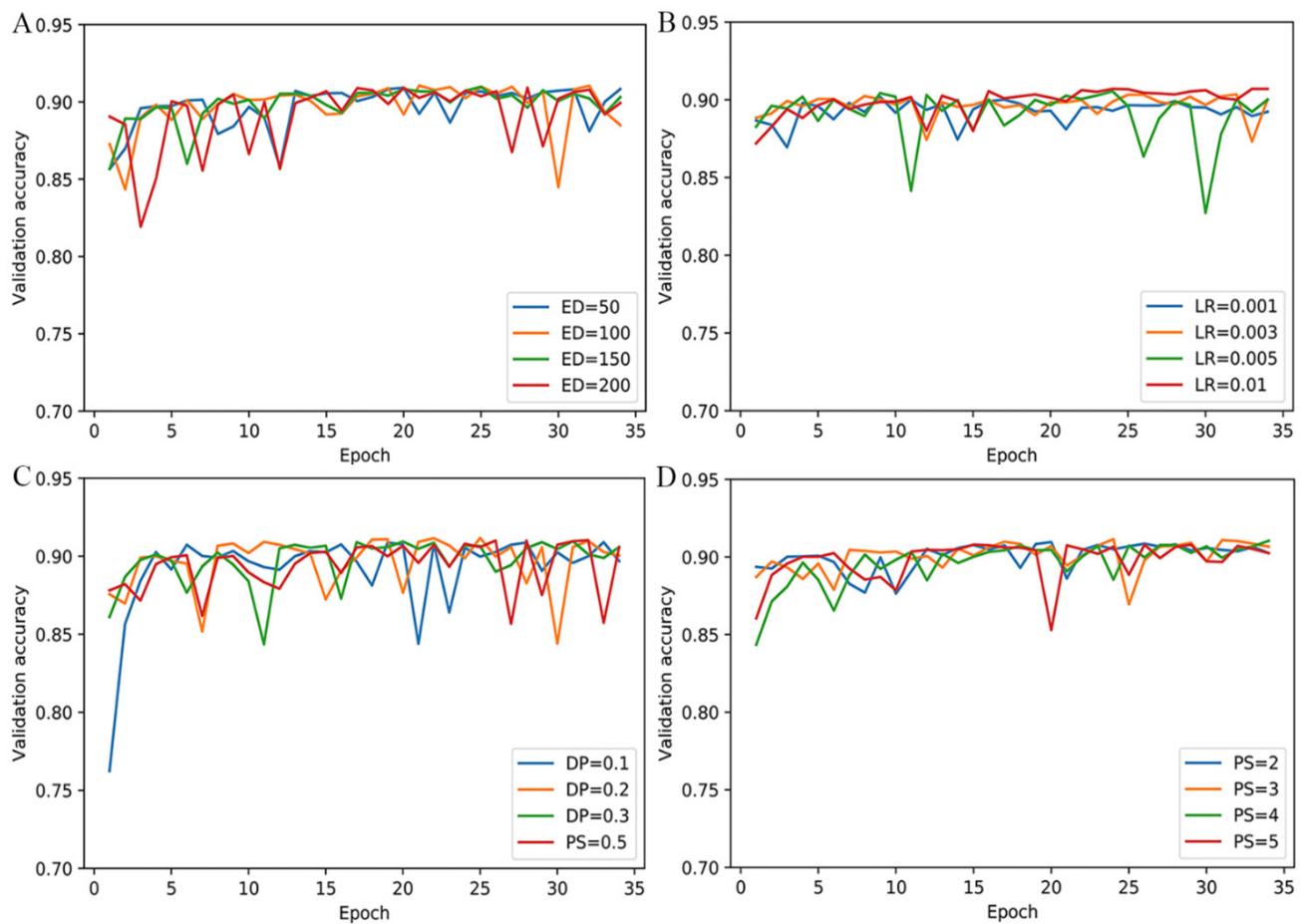


Figure 1. The predictive accuracy of various hyper-parameters. (A) ED, (B) LR, (C) DP, and (D) PS denote embedding dimension in the embedding layer, the learning rate, the dropout rate, and the pooling size in the pooling layer, respectively.

Table 1. The details of hyper-parameters in the MPMABP.

Layer	Super-Parameter	Value
Embedding	embedding dimensions	100
CNN layer 1	number of kernels	64
	size of kernels	3
CNN layer 2	number of kernels	64
	size of kernels	5
CNN layer 3	number of kernels	64
	size of kernels	8
CNN layer 4	number of kernels	64
	size of kernels	10
CNN layer 5	number of kernels	64
	size of kernels	12
Pooling layer	size of pooling	3
	stride	1
Bi-LSTM layer	number of neurons	32
Dense1	number of neurons	64
	activation function	relu
Dense2	number of neurons	128
	activation function	relu
Dense3	number of neurons	5
	activation function	relu

2.2. Comparison with State-of-the-Art Methods

To the best of our knowledge, the MLBP [87] is the latest method to classify multi-functional multi-label bioactive peptides so far. Of course, there are some multi-label algorithms which are applicable to predicting bioactive peptides, such as calibrated label ranking (CLR) [92], random k-label sets (RAKEL) [93], ranking support vector machine, and binary relevance with robust low-rank learning (RBRL) [94], and multi-label learning with deep forest (MLDF) [95]. We conducted the same experiments as the MLBP [87] for comparison. As shown in Table 2, the MPMABP outperformed the MLBP in terms of Precision, Coverage, Accuracy, and Absolute true. The MPMABP promoted the Precision by about 0.034, the Coverage by 0.037, the Accuracy by 0.027, and the Absolute true by 0.011. The lower the Absolute false, the better the predictive performance. The MPMABP decreased the Absolute false by 0.010. We compared five methods over the independent test. As shown in Table 3, we observed the best Precision, the best Coverage, the best Accuracy, the best Absolute true, and the worst Absolute false in the MPMABP, implying that the MPMABP is comprehensively superior to the state-of-the-art methods.

Table 2. The 5-fold cross-validation results of the training dataset.

Model	Precision	Coverage	Accuracy	Absolute True	Absolute False
MPMABP	0.731 ± 0.011	0.738 ± 0.012	0.722 ± 0.010	0.696 ± 0.013	0.099 ± 0.006
MLBP [87]	0.697 ± 0.012	0.701 ± 0.014	0.695 ± 0.012	0.685 ± 0.011	0.109 ± 0.004

Note: ± indicates standard deviation over the 5-fold cross-validations.

Table 3. The independent test results.

Model	Precision	Coverage	Accuracy	Absolute True	Absolute False
MPMABP	0.728	0.749	0.727	0.704	0.101
MLBP [87]	0.710	0.720	0.709	0.697	0.106
CLR [92]	0.667	0.677	0.666	0.655	0.133
RAKEL [93]	0.649	0.648	0.648	0.647	0.141
MLDF [95]	0.649	0.649	0.648	0.646	0.119
RBRL [94]	0.650	0.651	0.649	0.646	0.140

We compared predictive performances of five methods on single functional bioactive peptides. SN and SP were computed by Equations (7) and (8), where the investigated category of bioactive peptides is viewed positive and other as negative. For example, when we computed SN of the AIP, all the AIP bioactive peptides were viewed positive and the other as negative. As shown in Figure 2, predictive performance differs largely with categories. The MPMABP reached the best SN in terms of AIP and AHP, and the best SP in terms of ACP and ADP. However, the MPMABP is inferior to the MLBP in terms of AMP, ACP, and ADP. Since the predictive performances of the MPMABP over AHP are far greater than that of the MLBP, the MPMABP is, as a whole, superior to the MLBP.

Recently, many methods have been developed to identify single activity of bioactive peptides. In order to validate effectiveness and efficiency in classifying activities of bioactive peptides, we compared some state-of-the-art methods which provide web applications, i.e., IAMP-RAAC [96], mAHTPred [97], AHPPred [98], and AIPpred [99]. The IAMP-RAAC [96] is a reduced amino acid cluster-based method for distinguishing AMP from ACP, the mAHTPred [97] is a meta predictor for AHP, the AHPPred [98] is an CNN- and LSTM-based method for AHP prediction, and the AIPpred [99] is a random-forest-based predictor for AIP. Except the IAMP-RAAC [96], all the methods can only be applied to predict specific activity of bioactive peptides. For fair comparison, removing overlapping bioactive peptides with the training samples, we used the overlapping samples with these independent tests in these methods, respectively. Table 4 lists the performances (SN). Obviously, except the mAHTPred, the MPMABP outperformed the other three state-of-the-art methods.

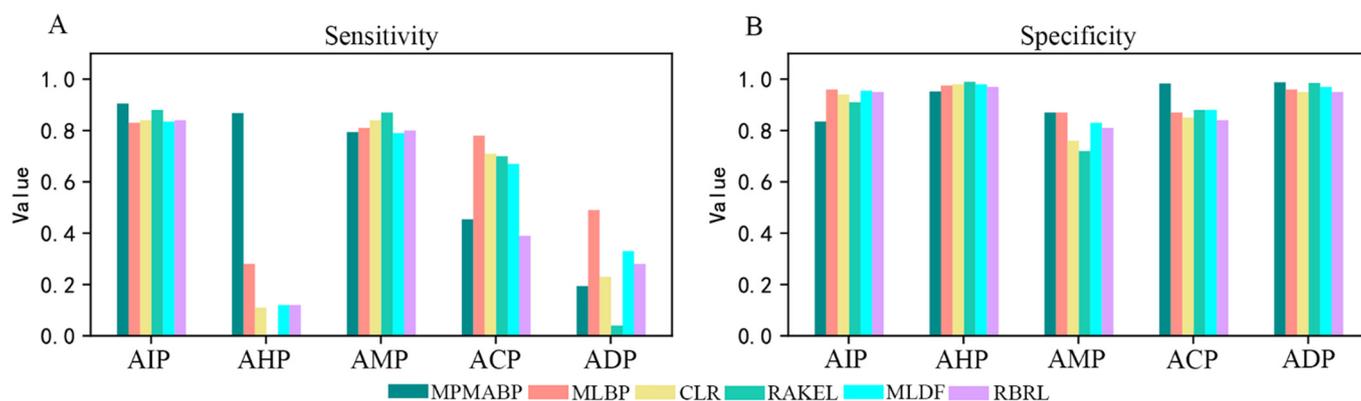


Figure 2. The predictive performance of single-functional bioactive peptides. (A) Comparison of MPMABP with other methods on SN, (B) Comparison of MPMABP with other methods on SP.

Table 4. Comparison with four existing state-of-the-art methods.

Type	Method	MPMABP	IAMP-RAAC [96]	mAHTPred [97]	AHPPred [98]	AIPpred [99]
AMP		0.872	0.788	-	-	-
ACP		0.505	0.333	-	-	-
AHP		0.889	-	0.986	0.361	-
AIP		0.914	-	-	-	0.827

2.3. Case Study

In order to prove further predicting ability of the MPMABP, we randomly chose 10 bioactive peptides to be predicted. Table 5 lists predictions by three methods over 10 bioactive peptides. The MPMABP correctly predicted multi-activities of all the bioactive peptides. The MLBP [87] predicted correctly 7, mistakenly 1, and partly correctly 2 of 10 bioactive peptides. MultiPep [100] is also a method which is able to predict up to 12 types of bioactive peptide. We utilized the webserver of the MultiPep: <https://agbg.shinyapps.io/MultiPep/> (accessed on 5 February 2022) to perform prediction. Obviously, the MultiPep predicted less classes than the true class for ADP-463 and AIP-1050 and predicted more classes for the other seven bioactive peptides. The 10 cases illustrated the superior predictive performance of the MPMABP over the MLBP [87] and the MultiPep [100].

Table 5. Comparison of MPMABP with two other algorithms by case study.

Sequence	True labels	Prediction		
		MPMABP	MLBP [87]	MultiPep [100]
ACP-499	ACP	ACP	ACP	AMP/anti-virus/ACP/anti-bacterial/anti-fungal
ADP-156	ADP	ADP	ADP	ACE inhibitor/AHP
AHP-665	AHP	AHP	AHP	Neuropeptide/peptidehormone
AIP-1046	AIP	AIP	AIP	AMP/anti-bacterial
AMP-1389	AMP	AMP	AMP	AMP/anti-bacterial
ACP-29	ACP/AMP	ACP/AMP	AMP	ACP/anti-bacterial/anti-fungal
ACP-220	ACP/AMP	ACP/AMP	None	AMP/anti-bacterial/anti-fungal
ADP-463	ADP/AHP	ADP/AHP	ADP	ADP
AIP-1050	AIP/ADP	ADP/AIP	ADP/AHP	ADP
AHP-483	AHP/ACP	AHP	AHP	Antioxidative/ACE inhibitor/AHP

2.4. Discussion

The MPMABP is a CNN and Bi-LSTM-based deep learning method for predicting multi-label bioactive peptides. The MPMABP stacked five CNN and Bi-LSTM modules in a parallel manner. The MPMABP utilized the ResNet to preserve necessary information in the forwarding process. We investigated, respectively, predictive performances of the

MPMABPs without the ResNet (called MPMABP_{wr}) and in a series-connection manner (called MPMABP_{sc}). Table 6 shows predictive performances over the 5-fold cross-validation and the independent test. Contrasting Table 6 with Tables 2 and 3, we found that the inclusion of the ResNet and the parallel manner remarkably improved the predictive performances, respectively.

Table 6. The predictive performances of MPMABP_{wr} and MPMABP_{sc}.

Model	Precision	Coverage	Accuracy	Absolute True	Absolute False
MPMABP _{wr} ^a	0.702	0.723	0.701	0.678	0.108
MPMABP _{wr} ^b	0.697 ± 0.013	0.704 ± 0.022	0.688 ± 0.013	0.663 ± 0.013	0.105 ± 0.003
MPMABP _{sc} ^a	0.697	0.719	0.696	0.672	0.109
MPMABP _{sc} ^b	0.704 ± 0.019	0.710 ± 0.023	0.694 ± 0.019	0.668 ± 0.018	0.103 ± 0.006

^a and ^b represent independent test and 5-fold cross-validation, respectively.

The CNN and the LSTM are two dominating components in deep learning, each with respective advantages. The CNN is good at characterizing local properties, while the LSTM does well in capturing semantic of words in context of the sequences. We combined two architectures to make full use of their merits. We experimented with many simpler architectures of deep neural network, that is, the MPMABP without the CNN, the MPMABP without the LSTM, and the MPMABP with only one branch. Tables 7 and 8 show the predictive performance by five-fold cross-validation and by the independent test. The exclusion of the CNN or the LSTM from the original MPMABP lead the predictive performance to decrease. The degeneration of the MPMABP also reduced the ability to accurately classify bioactive peptides

Table 7. The predictive performance of 5-fold cross-validation.

Model	Precision	Coverage	Accuracy	Absolute True	Absolute False
MPMABP	0.731 ± 0.011	0.738 ± 0.012	0.722 ± 0.010	0.696 ± 0.013	0.099 ± 0.006
No CNN	0.724 ± 0.011	0.729 ± 0.010	0.714 ± 0.011	0.689 ± 0.013	0.101 ± 0.004
No LSTM	0.708 ± 0.017	0.708 ± 0.014	0.698 ± 0.017	0.678 ± 0.020	0.102 ± 0.004
Degeneration	0.725 ± 0.015	0.733 ± 0.015	0.716 ± 0.014	0.688 ± 0.013	0.101 ± 0.009

Table 8. The predictive performance of the independent test.

Model	Precision	Coverage	Accuracy	Absolute True	Absolute False
MPMABP	0.728	0.749	0.727	0.704	0.101
No CNN	0.676	0.688	0.675	0.662	0.105
No LSTM	0.659	0.670	0.658	0.645	0.109
Degeneration	0.690	0.708	0.689	0.670	0.111

We investigated distribution of amino acid over five categories of bioactive peptides. As listed in Figure 3, some distributions are common in all classes, but some have remarkable differences across different types of bioactive peptides. The amino acid K appears more frequently in the ACP, P more frequently in the AHP, and L more frequently in the ADP.

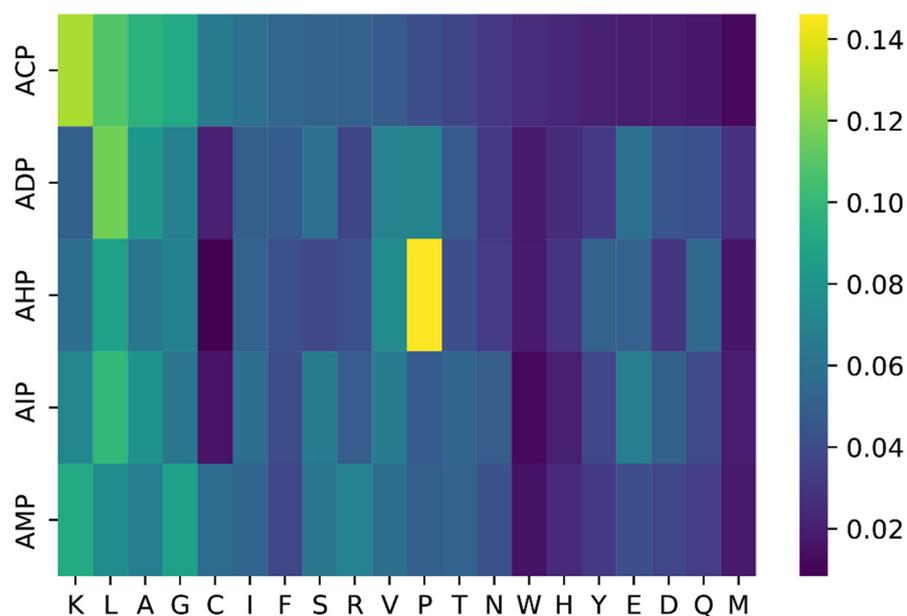


Figure 3. Heat map of amino acid distribution of five types of bioactive peptides.

3. Materials and Methods

3.1. Datasets

We used the same experimental dataset as in [87]. The dataset was retrieved by searching the Google Scholar engine with the keyword bioactive peptide, in 2020 [87]. The initial dataset included 18 types of bioactive peptides. Since the number of training samples is too small to train deep neural network favorably, the peptides of less than 500 residues were dropped out. Consequently, five types of functional peptides (AMP, ACP, ADP, AHP, and AIP) were preserved. The clustering tool CD-HIT [101] was used to remove or decrease redundancy and homology. The sequence identity was set to 0.9. The final numbers of the ACP, the ADP, the AHP, the AIP, and the AMP are, respectively, 646, 514, 868, 1678, and 2409, as shown in Figure 4. Obviously, most bioactive peptides are of only one type of activity, a small number of bioactive peptides are simultaneously of two types, and none belong to more than two types. This is a multi-class and multi-label issue. In total, 80 percent of all the peptides were randomly sampled as the training set and the remaining 20 percent were used the testing set.

3.2. Methodology

As shown in Figure 5, the proposed MPMABP is an end-to-end deep learning model which is made up of the 1D CNN, the LSTM, the embedding, the batch normalization, and the full-connected layer. The input to the MPMABP is amino acid sequences, which are subsequently transformed into continuous vectors by the embedding layer. Five parallel modules follow the batch-normalization layer to extract deep and abstract representations, of which each is constructed by linking 1D CNN, Bi-LSTM, and max pooling in order. To keep the information, the ResNet structure is used. All the representations are concatenated to be entered into the classification module, which consists of three fully connected layers and a dropout layer. The final fully connected layer has five neurons with the sigmoid function. The output of each neuron stands for a probability of belonging to a corresponding type of peptide.

convolution kernel in the element-wise manner and then to sum all the products. The convolution kernel serves as filters in the field of signals, and thus is also called filters. The size of the convolution kernel is influential for representation of the original features. The larger size could capture global structures, while the smaller size could characterize local structure. For extracting different scale representations from sequences, we used five convolution kernels with different sizes. As shown in Figure 4, the smallest size is 3, and the largest size is 12. Therefore, we obtained multi-scale representations of primary sequences.

3.2.3. Bi-LSTM

The LSTM proposed by Hochreiter et al. [107] is an improved recurrent neural network (RNN) [108–110]. The LSTM [107] introduces the gate mechanisms such as the forget gate, the output gate, and the input gate, and thus solves well the gradient vanishing or exploding issue occurring in the long sequence analysis. Compared with the traditional RNN, the LSTM is capable of capturing long-distance dependency. Therefore, the LSTM has been used in a wide range of fields including action recognition [111], succinylation prediction [28], and N4-Acetylcytidine prediction [21]. The single LSTM is unidirectional, which is generally able to uncover relationships with previous words. Therefore, the Bi-LSTM is used in practice. The Bi-LSTM [91,112] is composed of two LSTMs in opposite directions, one from the front to the back, and the other from the back to the front. Two LSTMs have identical inputs but have completely different learnable parameters. The outputs of two LSTMs are concatenated as the output of the Bi-LSTM.

3.2.4. Pooling

Pooling is a popular operation in the CNN, which serves as non-linear down-sampling. The pooling has dual roles. One is to decrease the dimensionality of representations, to save storage space, and to accelerate the calculation and another is to avoid over-fitting issues. The pooling operations include max pooling and average pooling. We used the max pooling herein.

3.3. ResNet

The ResNet [113] is actually the improved version of the CNN. The ResNet is very simple but effective. As shown in Figure 4, the ResNet consists mainly of two branches: one is to directly link the next layer and the other is the CNN. The sum of the input and the output of the CNN is the output of the ResNet. The ResNet enables construction of deeper neural networks without loss of information. These popular deep learning methods such as VGG, Transformer, and GoogleNet used the ResNet architecture. Here, we used the ResNet to fuse multi-scale representations and original information.

3.4. Fully Connected Layer

The fully connected layer is identical to the hidden layers in the multilayer perceptron, which is generally used as linear representations of inputs. Therefore, it is essential to classification or embedding representation in deep learning. We used three fully connected layers, of which the last has five neurons, each to represent a class of the functional peptide. Because this is a multi-label multi-class issue, we used the sigmoid activation function in the last fully connected layer. The neuron outputting more than 0.5 indicated that the input belonged to the corresponding functional peptides. We also used one dropout following the first and the second fully connected layers, respectively, so as to decrease overfitting.

3.5. Validation and Evaluation Metrics

We employed both hold-out and 5-fold cross-validation to examine the proposed method. In the hold-out, 80 percent of all the experimental peptides are sampled randomly as the training set, and the remaining 20 percent as the validation set. The model is trained by the training set and then validated by the validation set. In the 5-fold cross-validation,

the training set is separated into five parts on average. Four parts are used to train the model and the remaining is used to test the model. The process is repeated five times.

For convenient comparison with the state-of-the-art methods, we used the same evaluation metrics as the MLBP [87], the CLR [92], the Rakel [93], the MLDF [95], and the RBRL [94]. These metrics are defined below.

$$\text{Precision} = \frac{1}{N} \sum_{i=1}^N \frac{\|L_i \cap L_i^*\|}{\|L_i^*\|} \quad (1)$$

$$\text{Coverage} = \frac{1}{N} \sum_{i=1}^N \frac{\|L_i \cap L_i^*\|}{\|L_i\|} \quad (2)$$

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^N \frac{\|L_i \cap L_i^*\|}{\|L_i \cup L_i^*\|} \quad (3)$$

$$\text{Absolute true} = \frac{1}{N} \sum_{i=1}^N ID(L_i, L_i^*) \quad (4)$$

$$\text{Absolute false} = \frac{1}{N} \sum_{i=1}^N \frac{\|L_i \cup L_i^*\| - \|L_i \cap L_i^*\|}{M} \quad (5)$$

where L_i and L_i^* denote the set of actual labels and predicted labels for the sample I , respectively, N is the total number of the testing samples, \cup as well as \cap denote the union and intersection of the set, respectively, $\|A\|$ is the number of elements of the set A , and ID is defined as:

$$ID(L_i, L_i^*) = \begin{cases} 1 & L_i == L_i^* \\ 0 & \text{other} \end{cases} \quad (6)$$

For Precision, Coverage, Accuracy, and Absolute true, the greater the value meant the better predictive performance. On the contrary, the less Absolute false indicated better predictive performance.

We employed the sensitivity (SN) and specificity (SP) which are the frequently used evaluation metrics in the binary classification. Below are SN and SP definitions:

$$\text{SN} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{SP} = \frac{TN}{TN + FP} \quad (8)$$

where TP as well as TN are the numbers of the true positive and true negative samples, respectively, and FP as well as FN are the number of false positive and false negative samples, respectively. This is a multi-label and multi-class issue, not a binary classification. Therefore, we viewed it as five binary classifications. Namely, for a given class, all the samples with such class are positive and others are negative. For example, when we computed SN and SP for the AMP, all the peptides of AMP are positive, and peptides with other classes are negative.

4. Conclusions

Most bioactive peptides play therapeutic roles such as resisting microbes and cancer, being potential, safe, and natural organic substances. We presented a CNN and Bi-LSTM deep learning method for classifying multi-label bioactive peptides from the primary protein sequences. Compared with the latest state-of-the-art method (MLBP), the presented method made two remarkable improvements: stacking CNN and Bi-LSTM module in a parallel manner and utilizing the ResNet. The former allows for extracting multi-scale information from sequences, while the latter keeps the information loss lower in the forward process. The inclusion of both improves the predictive performance. We also found that distribution of amino acids varies with category of bioactive peptide. The amino acid P

was enriched in the AHP, the L was enriched in the ADP, while the K was enriched in the ACP. The finding is helpful for determining activities of bioactive peptides.

Author Contributions: Conceptualization, G.H. and X.L.; methodology, G.H. and Y.L. (You Li); software, Y.L. (You Li); validation, Y.L. (You Li); formal analysis, Y.L. (Yuewu Liu) and Y.Y.; investigation, Y.L. (You Li); resources, Y.L. (You Li); data curation, Y.L. (You Li); writing—original draft preparation, Y.L. (You Li); writing—review and editing, G.H., X.L., Y.L. (Yuewu Liu) and Y.Y.; supervision, G.H.; project administration, G.H.; funding acquisition, G.H., Y.L. (Yuewu Liu) and Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 62162025), by the Scientific Research Fund of Hunan Provincial Education Department (grant number 21A0466, 19A215), by the Natural Science Foundation of Hunan Province (grant number 2020JJ4034), by the open project of Hunan Key Laboratory for Computation and Simulation in Science and Engineering (grant number 2019LCESE03), and Shaoyang University Innovation Foundation for Postgraduates (grant number CX2021SY031).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data and source code can be found here: <https://github.com/Good-Ly/MPMABP> (accessed on 21 April 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, S.; Luo, L.; Sun, X.; Ma, A. Bioactive Peptides: A Promising Alternative to Chemical Preservatives for Food Preservation. *J. Agric. Food Chem.* **2021**, *69*, 12369–12384. [[CrossRef](#)] [[PubMed](#)]
2. Manikkam, V.; Vasiljevic, T.; Donkor, O.N.; Mathai, M.L. A Review of Potential Marine-derived Hypotensive and Anti-obesity Peptides. *Crit. Rev. Food Sci. Nutr.* **2016**, *56*, 92–112. [[CrossRef](#)] [[PubMed](#)]
3. Sánchez, A.; Vázquez, A. Bioactive peptides: A review. *Food Qual. Saf.* **2017**, *1*, 29–46. [[CrossRef](#)]
4. Kadam, S.U.; Tiwari, B.K.; Álvarez, C.; O'Donnell, C.P. Ultrasound applications for the extraction, identification and delivery of food proteins and bioactive peptides. *Trends Food Sci. Technol.* **2015**, *46*, 60–67. [[CrossRef](#)]
5. Chalamaiah, M.; Yu, W.; Wu, J. Immunomodulatory and anticancer protein hydrolysates (peptides) from food proteins: A review. *Food Chem.* **2018**, *245*, 205–222. [[CrossRef](#)]
6. Pavlicevic, M.; Marmiroli, N.; Maestri, E. Immunomodulatory peptides—A promising source for novel functional food production and drug discovery. *Peptides* **2022**, *148*, 170696. [[CrossRef](#)]
7. Hussain, M.A.; Sumon, T.A.; Mazumder, S.K.; Ali, M.M.; Jang, W.J.; Abualreesh, M.H.; Sharifuzzaman, S.M.; Brown, C.L.; Lee, H.-T.; Lee, E.-W.; et al. Essential oils and chitosan as alternatives to chemical preservatives for fish and fisheries products: A review. *Food Control* **2021**, *129*, 108244. [[CrossRef](#)]
8. Majumder, K.; Wu, J. Molecular Targets of Antihypertensive Peptides: Understanding the Mechanisms of Action Based on the Pathophysiology of Hypertension. *Int. J. Mol. Sci.* **2015**, *16*, 256–283. [[CrossRef](#)]
9. Gupta, S.; Sharma, A.K.; Shastri, V.; Madhu, M.K.; Sharma, V.K. Prediction of anti-inflammatory proteins/peptides: An insilico approach. *J. Transl. Med.* **2017**, *15*, 7. [[CrossRef](#)]
10. Xie, M.; Liu, D.; Yang, Y. Anti-cancer peptides: Classification, mechanism of action, reconstruction and modification. *Open Biol.* **2020**, *10*, 200004.
11. Zhao, J.; Bai, L.; Ren, X.-k.; Guo, J.; Xia, S.; Zhang, W.; Feng, Y. Co-immobilization of ACH11 antithrombotic peptide and CAG cell-adhesive peptide onto vascular grafts for improved hemocompatibility and endothelialization. *Acta Biomater.* **2019**, *97*, 344–359. [[CrossRef](#)] [[PubMed](#)]
12. Udenigwe, C.C. Bioinformatics approaches, prospects and challenges of food bioactive peptide research. *Trends Food Sci. Technol.* **2014**, *36*, 137–143. [[CrossRef](#)]
13. Li, Y.; Lyu, J.; Wu, Y.; Liu, Y.; Huang, G. PRIP: A Protein-RNA Interface Predictor Based on Semantics of Sequences. *Life* **2022**, *12*, 307. [[CrossRef](#)] [[PubMed](#)]
14. Hussain, W.; Rasool, N.; Khan, Y.D. A sequence-based predictor of Zika virus proteins developed by integration of PseAAC and statistical moments. *Comb. Chem. High Throughput Screen.* **2020**, *23*, 797–804. [[CrossRef](#)]
15. Aranha, M.P.; Spooner, C.; Demerdash, O.; Czejdo, B.; Smith, J.C.; Mitchell, J.C. Prediction of peptide binding to MHC using machine learning with sequence and structure-based feature sets. *Biochim. Et Biophys. Acta (BBA)-Gen. Subj.* **2020**, *1864*, 129535. [[CrossRef](#)]
16. Nielsen, M.; Andreatta, M.; Peters, B.; Buus, S. Immunoinformatics: Predicting peptide–MHC binding. *Annu. Rev. Biomed. Data Sci.* **2020**, *3*, 191–215. [[CrossRef](#)]

17. Yang, Z.; Yi, W.; Tao, J.; Liu, X.; Zhang, M.Q.; Chen, G.; Dai, Q. HPVMD-C: A disease-based mutation database of human papillomavirus in China. *Database* **2022**, *2022*, baac018. [[CrossRef](#)]
18. Kong, R.; Xu, X.; Liu, X.; He, P.; Zhang, M.Q.; Dai, Q. 2SigFinder: The combined use of small-scale and large-scale statistical testing for genomic island detection from a single genome. *BMC Bioinform.* **2020**, *21*, 159. [[CrossRef](#)]
19. Alzahrani, E.; Alghamdi, W.; Ullah, M.Z.; Khan, Y.D. Identification of stress response proteins through fusion of machine learning models and statistical paradigms. *Sci. Rep.* **2021**, *11*, 21767. [[CrossRef](#)]
20. Yang, S.; Wang, Y.; Chen, Y.; Dai, Q. MASQC: Next Generation Sequencing Assists Third Generation Sequencing for Quality Control in N6-Methyladenine DNA Identification. *Front. Genet.* **2020**, *11*, 269. [[CrossRef](#)]
21. Zhang, G.; Luo, W.; Lyu, J.; Yu, Z.-G.; Huang, G. CNNLSTMac4CPred: A Hybrid Model for N4-Acetylcytidine Prediction. *Interdiscip. Sci. Comput. Life Sci.* **2022**, *14*, 439–451. [[CrossRef](#)] [[PubMed](#)]
22. Tang, X.; Zheng, P.; Li, X.; Wu, H.; Wei, D.-Q.; Liu, Y.; Huang, G. Deep6mAPred: A CNN and Bi-LSTM-based deep learning method for predicting DNA N6-methyladenosine sites across plant species. *Methods* **2022**, *204*, 142–150. [[CrossRef](#)] [[PubMed](#)]
23. Naseer, S.; Hussain, W.; Khan, Y.D.; Rasool, N. iPhosS(Deep)-PseAAC: Identify Phosphoserine Sites in Proteins using Deep Learning on General Pseudo Amino Acid Compositions via Modified 5-Steps Rule. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *1*. [[CrossRef](#)] [[PubMed](#)]
24. Naseer, S.; Hussain, W.; Khan, Y.D.; Rasool, N. NPalmitylDeep-PseAAC: A predictor of N-palmitoylation sites in proteins using deep representations of proteins and PseAAC via modified 5-steps rule. *Curr. Bioinform.* **2021**, *16*, 294–305. [[CrossRef](#)]
25. Naseer, S.; Hussain, W.; Khan, Y.D.; Rasool, N. Sequence-based identification of arginine amidation sites in proteins using deep representations of proteins and PseAAC. *Curr. Bioinform.* **2020**, *15*, 937–948. [[CrossRef](#)]
26. Shah, A.A.; Khan, Y.D. Identification of 4-carboxyglutamate residue sites based on position based statistical feature and multiple classification. *Sci. Rep.* **2020**, *10*, 16913. [[CrossRef](#)]
27. Naseer, S.; Hussain, W.; Khan, Y.D.; Rasool, N. Optimization of serine phosphorylation prediction in proteins by comparing human engineered features and deep representations. *Anal. Biochem.* **2021**, *615*, 114069. [[CrossRef](#)]
28. Huang, G.; Shen, Q.; Zhang, G.; Wang, P.; Yu, Z.G. LSTMCNNsucc: A Bidirectional LSTM and CNN-Based Deep Learning Method for Predicting Lysine Succinylation Sites. *Biomed Res. Int.* **2021**, *2021*, 9923112. [[CrossRef](#)]
29. Onesime, M.; Yang, Z.; Dai, Q. Genomic Island Prediction via Chi-Square Test and Random Forest Algorithm. *Comput. Math. Methods Med.* **2021**, *2021*, 9969751. [[CrossRef](#)]
30. Dai, Q.; Bao, C.; Hai, Y.; Ma, S.; Zhou, T.; Wang, C.; Wang, Y.; Huo, W.; Liu, X.; Yao, Y. MTGIpick allows robust identification of genomic islands from a single genome. *Brief. Bioinform.* **2018**, *19*, 361–373. [[CrossRef](#)]
31. Roy, S.; Teron, R. BioDADPep: A Bioinformatics database for anti diabetic peptides. *Bioinformatics* **2019**, *15*, 780–783. [[CrossRef](#)] [[PubMed](#)]
32. Li, Q.; Zhang, C.; Chen, H.; Xue, J.; Guo, X.; Liang, M.; Chen, M. BioPepDB: An integrated data platform for food-derived bioactive peptides. *Int. J. Food Sci. Nutr.* **2018**, *69*, 963–968. [[CrossRef](#)] [[PubMed](#)]
33. Singh, S.; Chaudhary, K.; Dhanda, S.K.; Bhalla, S.; Usmani, S.S.; Gautam, A.; Tuknait, A.; Agrawal, P.; Mathur, D.; Raghava, G.P.S. SATPdb: A database of structurally annotated therapeutic peptides. *Nucleic Acids Res.* **2016**, *44*, D1119–D1126. [[CrossRef](#)] [[PubMed](#)]
34. Tyagi, A.; Tuknait, A.; Anand, P.; Gupta, S.; Sharma, M.; Mathur, D.; Joshi, A.; Singh, S.; Gautam, A.; Raghava, G.P.S. CancerPPD: A database of anticancer peptides and proteins. *Nucleic Acids Res.* **2015**, *43*, D837–D843. [[CrossRef](#)] [[PubMed](#)]
35. Kumar, R.; Chaudhary, K.; Sharma, M.; Nagpal, G.; Chauhan, J.S.; Singh, S.; Gautam, A.; Raghava, G.P.S. AHTPDB: A comprehensive platform for analysis and presentation of antihypertensive peptides. *Nucleic Acids Res.* **2015**, *43*, D956–D962. [[CrossRef](#)]
36. Mehta, D.; Anand, P.; Kumar, V.; Joshi, A.; Mathur, D.; Singh, S.; Tuknait, A.; Chaudhary, K.; Gautam, S.K.; Gautam, A.; et al. ParaPep: A web resource for experimentally validated antiparasitic peptide sequences and their structures. *Database* **2014**, *2014*, bau051. [[CrossRef](#)]
37. Shtatland, T.; Guettler, D.; Kossodo, M.; Pivovarov, M.; Weissleder, R. PepBank—A database of peptides based on sequence text mining and public peptide data sources. *BMC Bioinform.* **2007**, *8*, 280. [[CrossRef](#)]
38. Quiroz, C.; Saavedra, Y.B.; Armijo-Galdames, B.; Amado-Hinojosa, J.; Olivera-Nappa, Á.; Sanchez-Daza, A.; Medina-Ortiz, D. Peptipedia: A user-friendly web application and a comprehensive database for peptide research supported by Machine Learning approach. *Database* **2021**, *2021*, baab055. [[CrossRef](#)]
39. Khatun, M.S.; Hasan, M.M.; Kurata, H. PreAIP: Computational Prediction of Anti-inflammatory Peptides by Integrating Multiple Complementary Features. *Front. Genet.* **2019**, *10*, 129. [[CrossRef](#)]
40. He, B.; Yang, S.; Long, J.; Chen, X.; Zhang, Q.; Gao, H.; Chen, H.; Huang, J. TUPDB: Target-Unrelated Peptide Data Bank. *Interdiscip. Sci. Comput. Life Sci.* **2021**, *13*, 426–432. [[CrossRef](#)]
41. Usmani, S.S.; Kumar, R.; Kumar, V.; Singh, S.; Raghava, G.P.S. AntiTbPdb: A knowledgebase of anti-tubercular peptides. *Database* **2018**, *2018*, bay025. [[CrossRef](#)] [[PubMed](#)]
42. Minkiewicz, P.; Iwaniak, A.; Darewicz, M. BIOPEP-UWM Database of Bioactive Peptides: Current Opportunities. *Int. J. Mol. Sci.* **2019**, *20*, 5978. [[CrossRef](#)] [[PubMed](#)]
43. Basith, S.; Manavalan, B.; Shin, T.H.; Lee, D.Y.; Lee, G. Evolution of machine learning algorithms in the prediction and design of anticancer peptides. *Curr. Protein Pept. Sci.* **2020**, *21*, 1242–1250. [[CrossRef](#)] [[PubMed](#)]
44. Alotaibi, F.; Attique, M.; Khan, Y.D. AntiFlamPred: An Anti-Inflammatory Peptide Predictor for Drug Selection Strategies. *CMC-Comput. Mater. Contin.* **2021**, *69*, 1039–1055. [[CrossRef](#)]

45. Charoenkwan, P.; Chiangjong, W.; Hasan, M.M.; Nantasenamat, C.; Shoombuatong, W. Review and Comparative Analysis of Machine Learning-based Predictors for Predicting and Analyzing Anti-angiogenic Peptides. *Curr. Med. Chem.* **2022**, *29*, 849–864. [[CrossRef](#)] [[PubMed](#)]
46. Attique, M.; Farooq, M.S.; Khelifi, A.; Abid, A. Prediction of Therapeutic Peptides Using Machine Learning: Computational Models, Datasets, and Feature Encodings. *IEEE Access* **2020**, *8*, 148570–148594. [[CrossRef](#)]
47. Lertampaiporn, S.; Vorapreeda, T.; Hongsthong, A.; Thammarongtham, C. Ensemble-AMPPred: Robust AMP Prediction and Recognition Using the Ensemble Learning Method with a New Hybrid Feature for Differentiating AMPs. *Genes* **2021**, *12*, 137. [[CrossRef](#)]
48. Zhang, Y.; Lin, J.; Zhao, L.; Zeng, X.; Liu, X. A novel antibacterial peptide recognition algorithm based on BERT. *Brief. Bioinform.* **2021**, *22*, bbab200. [[CrossRef](#)]
49. Yan, J.; Bhadra, P.; Li, A.; Sethiya, P.; Qin, L.; Tai, H.K.; Wong, K.H.; Siu, S.W.I. Deep-AmPEP30: Improve Short Antimicrobial Peptides Prediction with Deep Learning. *Mol. Ther.-Nucleic Acids* **2020**, *20*, 882–894. [[CrossRef](#)]
50. Hussain, W. sAMP-PFPDeep: Improving accuracy of short antimicrobial peptides prediction using three different sequence encodings and deep neural networks. *Brief. Bioinform.* **2022**, *23*, bbab487. [[CrossRef](#)]
51. Arif, M.; Ahmed, S.; Ge, F.; Kabir, M.; Khan, Y.D.; Yu, D.-J.; Thafar, M. StackACPred: Prediction of anticancer peptides by integrating optimized multiple feature descriptors with stacked ensemble approach. *Chemom. Intell. Lab. Syst.* **2022**, *220*, 104458. [[CrossRef](#)]
52. Hasan, M.M.; Schaduagrath, N.; Basith, S.; Lee, G.; Shoombuatong, W.; Manavalan, B. HLPpred-Fuse: Improved and robust prediction of hemolytic peptide and its activity by fusing multiple feature representation. *Bioinformatics* **2020**, *36*, 3350–3356. [[CrossRef](#)] [[PubMed](#)]
53. Lawrence, T.J.; Carper, D.L.; Spangler, M.K.; Carrell, A.A.; Rush, T.A.; Minter, S.J.; Weston, D.J.; Labbé, J.L. amPEPpy 1.0: A portable and accurate antimicrobial peptide prediction tool. *Bioinformatics* **2021**, *37*, 2058–2060. [[CrossRef](#)] [[PubMed](#)]
54. Manavalan, B.; Basith, S.; Shin, T.H.; Wei, L.; Lee, G. AtbPpred: A Robust Sequence-Based Prediction of Anti-Tubercular Peptides Using Extremely Randomized Trees. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 972–981. [[CrossRef](#)]
55. Usmani, S.S.; Bhalla, S.; Raghava, G.P.S. Prediction of Antitubercular Peptides from Sequence Information Using Ensemble Classifier and Hybrid Features. *Front. Pharmacol.* **2018**, *9*, 954. [[CrossRef](#)]
56. Khatun, S.; Hasan, M.; Kurata, H. Efficient computational model for identification of antitubercular peptides by integrating amino acid patterns and properties. *FEBS Lett.* **2019**, *593*, 3029–3039. [[CrossRef](#)]
57. Zhang, J.; Zhang, Z.; Pu, L.; Tang, J.; Guo, F. AIEpred: An Ensemble Predictive Model of Classifier Chain to Identify Anti-Inflammatory Peptides. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 1831–1840. [[CrossRef](#)]
58. Hasan, M.M.; Alam, M.A.; Shoombuatong, W.; Deng, H.-W.; Manavalan, B.; Kurata, H. NeuroPred-FRL: An interpretable prediction model for identifying neuropeptide using feature representation learning. *Brief. Bioinform.* **2021**, *22*, bbab167. [[CrossRef](#)]
59. Ettayapuram Ramaprasad, A.S.; Singh, S.; Gajendra, P.S.R.; Venkatesan, S. AntiAngioPred: A server for prediction of anti-angiogenic peptides. *PLoS ONE* **2015**, *10*, e0136990. [[CrossRef](#)]
60. Blanco, J.L.; Porto-Pazos, A.B.; Pazos, A.; Fernandez-Lozano, C. Prediction of high anti-angiogenic activity peptides in silico using a generalized linear model and feature selection. *Sci. Rep.* **2018**, *8*, 15688. [[CrossRef](#)]
61. Khorsand, J.Z.; Ali Akbar, Y.; Kargar, M.; Ramin Shirali Hossein, Z.; Mahdevar, G. AntAngioCOOL: Computational detection of anti-angiogenic peptides. *J. Transl. Med.* **2019**, *17*, 71.
62. Laengsri, V.; Nantasenamat, C.; Schaduagrath, N.; Nuchnoi, P.; Prachayasittikul, V.; Shoombuatong, W. TargetAntiAngio: A Sequence-Based Tool for the Prediction and Analysis of Anti-Angiogenic Peptides. *Int. J. Mol. Sci.* **2019**, *20*, 2950. [[CrossRef](#)] [[PubMed](#)]
63. Lin, C.; Wang, L.; Shi, L. AAPred-CNN: Accurate predictor based on deep convolution neural network for identification of anti-angiogenic peptides. *Methods*, 2022; in press. [[CrossRef](#)] [[PubMed](#)]
64. Manavalan, B.; Shin, T.H.; Kim, M.O.; Lee, G. PIP-EL: A New Ensemble Learning Method for Improved Proinflammatory Peptide Predictions. *Front. Immunol.* **2018**, *9*, 1783. [[CrossRef](#)] [[PubMed](#)]
65. Gupta, S.; Madhu, M.K.; Sharma, A.K.; Sharma, V.K. ProInflam: A webserver for the prediction of proinflammatory antigenicity of peptides and proteins. *J. Transl. Med.* **2016**, *14*, 178. [[CrossRef](#)]
66. Khatun, M.S.; Hasan, M.M.; Shoombuatong, W.; Kurata, H. ProIn-Fuse: Improved and robust prediction of proinflammatory peptides by fusing of multiple feature representations. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 1229–1236. [[CrossRef](#)]
67. Chaudhary, K.; Kumar, R.; Singh, S.; Tuknait, A.; Gautam, A.; Mathur, D.; Anand, P.; Varshney, G.C.; Raghava, G.P.S. A Web Server and Mobile App for Computing Hemolytic Potency of Peptides. *Sci. Rep.* **2016**, *6*, 22843. [[CrossRef](#)]
68. Win, T.S.; Malik, A.A.; Prachayasittikul, V.S.; Wikberg, J.E.; Nantasenamat, C.; Shoombuatong, W. HemoPred: A web server for predicting the hemolytic activity of peptides. *Future Med. Chem.* **2017**, *9*, 275–291. [[CrossRef](#)]
69. Chiangjong, W.; Chutipongtanate, S.; Hongeng, S. Anticancer peptide: Physicochemical property, functional aspect and trend in clinical application. *Int. J. Oncol.* **2020**, *57*, 678–696. [[CrossRef](#)]
70. Agrawal, P.; Bhagat, D.; Mahalwal, M.; Sharma, N.; Raghava, G.P.S. AntiCP 2.0: An updated model for predicting anticancer peptides. *Brief. Bioinform.* **2021**, *22*, bbaa153. [[CrossRef](#)]

71. Chen, W.; Ding, H.; Feng, P.; Lin, H.; Chou, K.-C. iACP: A sequence-based tool for identifying anticancer peptides. *Oncotarget* **2016**, *7*, 16895. [[CrossRef](#)] [[PubMed](#)]
72. Vijayakumar, S.; Ptv, L. ACP: A Web Server for Prediction and Design of Anti-cancer Peptides. *Int. J. Pept. Res. Ther.* **2015**, *21*, 99–106. [[CrossRef](#)]
73. Akbar, S.; Hayat, M.; Iqbal, M.; Jan, M.A. iACP-GAEnsC: Evolutionary genetic algorithm based ensemble classification of anticancer peptides by utilizing hybrid feature space. *Artif. Intell. Med.* **2017**, *79*, 62–70. [[CrossRef](#)] [[PubMed](#)]
74. Manavalan, B.; Basith, S.; Shin, T.H.; Choi, S.; Kim, M.O.; Lee, G. MLACP: Machine-learning-based prediction of anticancer peptides. *Oncotarget* **2017**, *8*, 77121. [[CrossRef](#)] [[PubMed](#)]
75. Kabir, M.; Arif, M.; Ahmad, S.; Ali, Z.; Swati, Z.N.K.; Yu, D.-J. Intelligent computational method for discrimination of anticancer peptides by incorporating sequential and evolutionary profiles information. *Chemom. Intell. Lab. Syst.* **2018**, *182*, 158–165. [[CrossRef](#)]
76. Schaduagrath, N.; Nantasenamat, C.; Prachayasittikul, V.; Shoombuatong, W. ACPred: A Computational Tool for the Prediction and Analysis of Anticancer Peptides. *Molecules* **2019**, *24*, 1973. [[CrossRef](#)]
77. Wei, L.; Zhou, C.; Chen, H.; Song, J.; Su, R. ACPred-FL: A sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides. *Bioinformatics* **2018**, *34*, 4007–4016. [[CrossRef](#)]
78. Rao, B.; Zhou, C.; Zhang, G.; Su, R.; Wei, L. ACPred-Fuse: Fusing multi-view information improves the prediction of anticancer peptides. *Brief. Bioinform.* **2020**, *21*, 1846–1855. [[CrossRef](#)]
79. 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**, *17*, 1–9. [[CrossRef](#)]
80. Charoenkwan, P.; Chiangjong, W.; Lee, V.S.; Nantasenamat, C.; Hasan, M.M.; Shoombuatong, W. Improved prediction and characterization of anticancer activities of peptides using a novel flexible scoring card method. *Sci. Rep.* **2021**, *11*, 1–13. [[CrossRef](#)]
81. Wei, L.; Zhou, C.; Su, R.; Zou, Q. PEPred-Suite: Improved and robust prediction of therapeutic peptides using adaptive feature representation learning. *Bioinformatics* **2019**, *35*, 4272–4280. [[CrossRef](#)] [[PubMed](#)]
82. Wu, C.; Gao, R.; Zhang, Y.; De Marinis, Y. PTPD: Predicting therapeutic peptides by deep learning and word2vec. *BMC Bioinform.* **2019**, *20*, 456. [[CrossRef](#)] [[PubMed](#)]
83. Zhang, Y.P.; Zou, Q. PPTPP: A novel therapeutic peptide prediction method using physicochemical property encoding and adaptive feature representation learning. *Bioinformatics* **2020**, *36*, 3982–3987. [[CrossRef](#)] [[PubMed](#)]
84. Guo, Y.; Yan, K.; Lv, H.; Liu, B. PreTP-EL: Prediction of therapeutic peptides based on ensemble learning. *Brief. Bioinform.* **2021**, *22*, bbab358. [[CrossRef](#)]
85. He, W.; Jiang, Y.; Jin, J.; Li, Z.; Zhao, J.; Manavalan, B.; Su, R.; Gao, X.; Wei, L. Accelerating bioactive peptide discovery via mutual information-based meta-learning. *Brief. Bioinform.* **2022**, *23*, bbab499. [[CrossRef](#)]
86. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
87. Tang, W.; Dai, R.; Yan, W.; Zhang, W.; Bin, Y.; Xia, E.; Xia, J. Identifying multi-functional bioactive peptide functions using multi-label deep learning. *Brief. Bioinform.* **2021**, *23*, bbab414. [[CrossRef](#)]
88. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.
89. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
90. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]
91. Kiperwasser, E.; Goldberg, Y. Simple and Accurate Dependency Parsing Using Bidirectional LSTM Feature Representations. *Trans. Assoc. Comput. Linguist.* **2016**, *4*, 313–327. [[CrossRef](#)]
92. Fürnkranz, J.; Hüllermeier, E.; Mencía, E.L.; Brinker, K.J.M.I. Multilabel classification via calibrated label ranking. *Mach. Learn.* **2008**, *73*, 133–153. [[CrossRef](#)]
93. Tsoumakas, G.; Vlahavas, I. Random k-labelsets: An ensemble method for multilabel classification. In Proceedings of the European Conference on Machine Learning, Warsaw, Poland, 17–21 September 2007; pp. 406–417.
94. Wu, G.; Zheng, R.; Tian, Y.; Liu, D.J.N.N. Joint ranking SVM and binary relevance with robust low-rank learning for multi-label classification. *Neural Netw.* **2020**, *122*, 24–39. [[CrossRef](#)]
95. Yang, L.; Wu, X.-Z.; Jiang, Y.; Zhou, Z.-H. Multi-label learning with deep forest. *arXiv* **2019**, arXiv:1911.06557.
96. Dong, G.; Zheng, L.; Huang, S.-H.; Gao, J.; Zuo, Y. Amino acid reduction can help to improve the identification of antimicrobial peptides and their functional activities. *Front. Genet.* **2021**, *12*, 549. [[CrossRef](#)] [[PubMed](#)]
97. Manavalan, B.; Basith, S.; Shin, T.H.; Wei, L.; Lee, G. mAHTPred: A sequence-based meta-predictor for improving the prediction of anti-hypertensive peptides using effective feature representation. *Bioinformatics* **2019**, *35*, 2757–2765. [[CrossRef](#)] [[PubMed](#)]
98. Shi, H.; Zhang, S. Accurate Prediction of Anti-hypertensive Peptides Based on Convolutional Neural Network and Gated Recurrent unit. *Interdiscip. Sci. Comput. Life Sci.* **2022**, 1–6. [[CrossRef](#)]
99. Manavalan, B.; Shin, T.H.; Kim, M.O.; Lee, G. AIPpred: Sequence-based prediction of anti-inflammatory peptides using random forest. *Front. Pharmacol.* **2018**, *9*, 276. [[CrossRef](#)]

100. Grønning, A.G.; Kacprowski, T.; Schéele, C.J.B.M. MultiPep: A hierarchical deep learning approach for multi-label classification of peptide bioactivities. *Biol. Methods Protoc.* **2021**, *6*, bpab021.
101. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [[CrossRef](#)]
102. Fukushima, K.; Miyake, S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognit.* **1982**, *15*, 455–469. [[CrossRef](#)]
103. Hubel, D.H.; Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **1962**, *160*, 106–154. [[CrossRef](#)] [[PubMed](#)]
104. LeCun, Y.; Boser, B.; Denker, J.; Henderson, D.; Howard, R.; Hubbard, W.; Jackel, L. Handwritten digit recognition with a back-propagation network. *Adv. Neural Inf. Process. Syst.* **1990**, *2*, 396–404.
105. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
106. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
107. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
108. Pearlmutter, B.A. Learning State Space Trajectories in Recurrent Neural Networks. *Neural Comput.* **1989**, *1*, 263–269. [[CrossRef](#)]
109. Pearlmutter, B.A. Dynamic Recurrent Neural Networks. 1990. Unpublished work. Available online: <https://mural.maynoothuniversity.ie/5505/> (accessed on 21 April 2022).
110. Snyders, S.; Omlin, C.W. Inductive bias in recurrent neural networks. In Proceedings of the International Work-Conference on Artificial Neural Networks, Granada, Spain, 13–15 June 2001; pp. 339–346.
111. Ullah, A.; Ahmad, J.; Muhammad, K.; Sajjad, M.; Baik, S.W. Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features. *IEEE Access* **2018**, *6*, 1155–1166. [[CrossRef](#)]
112. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The Performance of LSTM and BiLSTM in Forecasting Time Series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.
113. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.