

Research Article

ACNNT3: Attention-CNN Framework for Prediction of Sequence-Based Bacterial Type III Secreted Effectors

Jie Li ¹, Zhong Li ^{1,2}, Jiesi Luo,³ and Yuhua Yao ⁴

¹School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China

²School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China

³Key Laboratory for Aging and Regenerative Medicine, Department of Pharmacology, School of Pharmacy, Southwest Medical University, Luzhou 646000, China

⁴School of Mathematics and Statistics, Hainan Normal University, Haikou 571158, China

Correspondence should be addressed to Zhong Li; lizhong@zstu.edu.cn and Yuhua Yao; yaoyuhua2288@163.com

Received 6 February 2020; Revised 9 March 2020; Accepted 17 March 2020; Published 2 April 2020

Guest Editor: Lei Chen

Copyright © 2020 Jie Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The type III secretion system (T3SS) is a special protein delivery system in Gram-negative bacteria which delivers T3SS-secreted effectors (T3SEs) to host cells causing pathological changes. Numerous experiments have verified that T3SEs play important roles in many biological activities and in host-pathogen interactions. Accurate identification of T3SEs is therefore essential to help understand the pathogenic mechanism of bacteria; however, many existing biological experimental methods are time-consuming and expensive. New deep-learning methods have recently been successfully applied to T3SE recognition, but improving the recognition accuracy of T3SEs is still a challenge. In this study, we developed a new deep-learning framework, ACNNT3, based on the attention mechanism. We converted 100 residues of the N-terminal of the protein sequence into a fusion feature vector of protein primary structure information (one-hot encoding) and position-specific scoring matrix (PSSM) which are used as the feature input of the network model. We then embedded the attention layer into CNN to learn the characteristic preferences of type III effector proteins, which can accurately classify any protein directly as either T3SEs or non-T3SEs. We found that the introduction of new protein features can improve the recognition accuracy of the model. Our method combines the advantages of CNN and the attention mechanism and is superior in many indicators when compared to other popular methods. Using the common independent dataset, our method is more accurate than the previous method, showing an improvement of 4.1-20.0%.

1. Introduction

Gram-negative bacteria can secrete proteins into host cells through a variety of secretion systems which affect the cell and its external environment. This process can be mediated by a variety of secretory systems, which can be divided into eight categories: type I to VIII secretory systems (T1SS-T8SS) [1]. Type I and III secretory systems are independent of signal sequences (sec), while types II and IV depend on signal sequences. The proteins secreted by the sec-dependent secretion system have a signal peptide sequence mainly composed of N-terminal hydrophobic amino acids which can guide the protein through the cell membrane.

When the protein reaches the periplasm, the signal peptide is cut off. Type II and IV secretion systems remove the N-terminal part of the secreted protein in the periplasm. The difference between systems is that proteins pass through the outer membrane in different ways. When protein secreted by the type II secretion system passes through the outer membrane, an additional set of inner membrane and outer membrane proteins is needed to assist, while the type IV secretion system includes a series of autotransporters which form a hole in the outer membrane to make the protein pass through, autolytically cut, and then release the protein. Neither the type I nor III secretion system processes the terminal amino acid of the secreted protein, nor does it appear that

the secreted protein stays in the periplasm. The secretion signal of the protein secreted by the type I secretion system is located at about 60 amino acids from the C-terminal end of the protein. This secretory signal appears to be subfamily specific, and the secreted proteins are not easily cut by proteolytic enzymes. The type V secretion system is related to the secretion of macromolecular proteins and may also be sec-dependent.

The type III secretion system is a transmembrane channel formed by the multicomponent protein complex that has been widely encoded in many Gram-negative bacteria including *Escherichia*, *Shigella*, *Yersinia*, *Salmonella*, and *Pseudomonas* [2, 3]. It can change the signal transduction [4] and innate immune response [5] of host cells by secreting proteins or injecting these virulent proteins directly into host cells. Type III secretion systems (T3SSs) have been widely studied because they are critical for virulence in various human pathogens. There are many *in vivo* and *in vitro* methods for predicting T3SEs, and while some of them obtain good predictions, the experiments are complicated and time-consuming.

Some machine-learning methods have successfully been used to predict T3SEs, such as the Naïve Bayes (NB) [6], artificial neural network (ANN) [7], support vector machine (SVM) [8, 9], and random forest (RF) [10]. The disadvantage of these machine-learning methods is that features must be defined in advance, the appropriate selection of features will affect the prediction accuracy, and the flexibility of model change or update is limited [11]. Many deep-learning methods have been recently proposed, such as LSTM [12], ResNet [13], DenseNet [14], and VGG16 [15], and these methods can also be used in bioinformatics and other related fields [16, 17]. The deep-learning method DeepT3 [11] trained deep CNN using only one-hot encoding as the model feature input and achieved good prediction results in terms of accuracy. Since only one feature is input and CNN cannot connect the sequence context well when extracting sequence features, this method can be improved upon for predicting T3SEs.

The attention mechanism has recently gained popularity in neural networks because it can weigh the input features to measure the importance of each feature to the object recognition. It has widely been applied for text and image classification [18, 19], machine translation [20], and bioinformatics [21]. In this study, we propose a method for predicting T3SEs using N-terminal sequences based on the Attention-CNN. Our model extracts features of one-hot and PSSM from 100 residues of the N-terminal sequence and fuses them as the feature input. The attention layer in the model can well connect the front and back of the sequence, and the CNN module can well extract the features of the sequence. We combine these two modules to make the entire framework learn the features of the sequence to their maximum extent. The results show that our method is effective in predicting T3SEs; not only can it accurately capture protein transport target information, but it also performs better than the existing methods.

2. Materials and Methods

2.1. Dataset. We collected a comprehensive dataset from multiple bacterial species known as T3SEs and non-T3SEs

from previous studies by Yang et al. [10], Wang et al. [22], and Tay et al. [23]. CD-HIT [24] with the sequence identity cutoff of 30% was used for sequence alignment to remove proteins with high similarity, and by skipping proteins with less than 100 amino acids, we obtained a balance dataset containing 283 T3 proteins and 311 non-T3 proteins.

We established our negative sample set by selecting type I to VIII secreted proteins of Gram-negative bacteria and removing type III secreted proteins and their homologues. In order to establish a 1:3 ratio of positive to negative [11], we randomly selected negative samples from the previous work of Dong et al. [8] and eliminated protein sequences with high similarity, resulting in a total of 835 negative samples.

There are two test sets used to evaluate our method. The independent dataset collected from Li et al. [11] contains 35 type III effectors and 86 non-type III effectors. The other test dataset is from the plant pathogen *P. syringae*. 85 type III effectors and 14 non-type III effectors that were not included in all models were collected from Baltrus et al. [25].

At present, most tools are based on the full sequence information of proteins or only 100 C-terminal residues [26]. In previous studies, N-terminal residues have been shown to also provide targeted information for protein transport [27–29], and the target information of T3SEs is usually located in the 50-100 N-terminal residues in different bacteria [30, 31]. Therefore, we have only used the N-terminal sequences in all the following calculations.

2.2. Feature Extraction. The feature input of our model is the combination of one-hot encoding and the PSSM of a protein. Each sequence is transformed into a one-hot matrix with 100 rows and 20 columns and a PSSM matrix with 100 rows and 20 columns, which are integrated into a combination matrix with 200 rows and 20 columns as the feature input. The 20 columns in the one-hot matrix correspond to 20 amino acids. One-hot encoding solves the problem of the classifier not effectively processing attribute data and expands the features to a certain extent. However, compared to PSSM, one-hot encoding is weaker with regard to protein feature extraction. Here, the introduction of PSSM enables the network model to better learn the characteristic preference of proteins, because PSSM features consider the position weight, number, and other parameters of each amino acid in the protein. PSSM also considers evolutionary information, so even the same residue may generate different characteristics, and it can effectively extract information from amino acid sequences. We used the PSI-BLAST [32] search database from UniprotKB/Swiss-Prot [33] to obtain the PSSM of the target protein. The matrix is an $L \times 20$ matrix, where L represents the total number of residues in the target protein's amino acid sequence. At the same time, we use 1, 2, 3, ..., 20 to represent the individual characters of the ordered 20 basic amino acids and get the corresponding number of columns. In summary, $U_{i \rightarrow j}^{\oplus}$ indicates the possibility that the i position of the amino acid sequence of the target protein is encoded as the basic amino acid j during the evolution process.

$$\text{PSSM} = \begin{bmatrix} U_{1 \rightarrow 1}^{\oplus} & U_{1 \rightarrow 2}^{\oplus} & \cdots & U_{1 \rightarrow 20}^{\oplus} \\ U_{2 \rightarrow 1}^{\oplus} & U_{2 \rightarrow 2}^{\oplus} & \cdots & U_{2 \rightarrow 20}^{\oplus} \\ \vdots & \vdots & \vdots & \vdots \\ U_{L \rightarrow 1}^{\oplus} & U_{L \rightarrow 2}^{\oplus} & \cdots & U_{L \rightarrow 20}^{\oplus} \end{bmatrix}. \quad (1)$$

2.3. Overview of Attention-CNN Model. The traditional CNN model includes convolution, pooling, and full connection layers, and it can be used to extract the sequence features of proteins. However, the sequence of a protein is more like a piece of text composed of amino acids, and since, when one amino acid may be related to the amino acids around it or those even farther away, it is not enough to extract these features using only the CNN mechanism. We also need to consider the information before and after the protein sequence and the correlation between discontinuous amino acids. Intuitively, an amino acid or a segment of amino acids may have a great impact on the protein sequence, so we can set a higher weight to this or this part of amino acids and have thus introduced the attention layer into the network.

Attention is a network structure model layer based on encode-decode, which has achieved satisfying prediction results compared to other traditional models in many fields including machine translation, picture description, and speech recognition. This implementation of the attention mechanism retains the intermediate output results of the input sequence of an LSTM encoder, then trains a model to selectively learn these inputs and associate the output sequence with them when the model outputs.

We added the attention and full connection layers in parallel after the convolution and pool layers, so that the model can not only take advantage of the mechanism of attention to learn the front and back features of the sequence, but also use the advantages of CNN feature extraction.

Our framework, ACNNT3, first uses multiple convolution and pooling layers to automatically learn protein sequence features, then takes the output feature vector as the input of the attention layer to calculate the score showing whether the neural network pays attention to the sequence features of the location. We define the output after convolution and pooling as a matrix $M^c (d \times q)$, where d is the number of convolution kernels, q is the whole position after sequence pooling, and the column j of the feature map matrix M^c can be viewed as a feature vector (denoted by V_j). W_j is the normalized importance score that is used to further average the columns of the feature map matrix M^c . The dense matrix output through the attention network is M^a , i.e.,

$$M^a = \sum_{j=1}^q w_j v_j, \quad (2)$$

$$w_j = \frac{\exp(e_j)}{\sum_{j=1}^q \exp(e_j)}, \quad (3)$$

where e_j is the importance score of the shared network and W_j is the relevant standardization score.

In order to integrate the features after convolution-pooling and the feature output by the attention layer, we first connect all the values in M^c and project them linearly to a value that represents the contribution of the whole sequence, represented by S^c , then we concatenate it with the dense representation M^a and input it into the logistic regression classifier to obtain the prediction score, namely,

$$\text{Pred}(s) = \text{sigm}(\text{concat}(M^a, S^c)), \quad (4)$$

where s represents a position in the integrated sequence.

$$S^c = \text{dense}(\text{pool}(\text{conv}(\text{encode}(s))))), \quad (5)$$

where $\text{encode}(\cdot)$, $\text{conv}(\cdot)$, $\text{pool}(\cdot)$, $\text{concat}(\cdot)$, $\text{dense}(\cdot)$, and $\text{sigmoid}(\cdot)$ represent the unification of one-hot and PSSM encoding, convolution, maximum pooling, concatenation, dense connection, and sigmoid operation, respectively. At the same time, for a specific sequence, we can also output a weight vector, i.e.,

$$\text{AttMap}(s) = (w_1, \dots, w_q). \quad (6)$$

This formula is used to express the attention of the model to each position of the input sequence.

2.4. Model Training. ACNNT3 is composed of a series of modules which use the fusion features of 100 amino acids at the N-terminal of the protein as input to predict T3SEs (Figure 1). The ACNNT3 model consists of convolution, pooling, attention, and fully connected layers. We use cross-validation to train our model and improve the generalization ability. The loss function uses a binary cross entropy loss function, and the optimizer uses the Adam algorithm. In Figure 2, we give the accuracy (ACC) comparison on the independent datasets under different epochs and batches. Since the dataset is not very large, the number of training epochs is set as 50 and the best batch value on the crossvalidation set is 10 as the optimal setting.

2.5. Performance Evaluation. We used 5-fold crossvalidation to estimate the classification performance of our model. Namely, we repeated the process five times and recorded the training parameters and average performance parameters for each time. The commonly used evaluation indexes for two-class classification are precision (PRE), sensitivity (SN), specificity (SP), F1 score, accuracy (ACC), and Matthew's correlation coefficient (MCC):

$$\text{PRE} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (7)$$

$$\text{SN} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (8)$$

$$\text{SP} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (9)$$

$$\text{F1 score} = 2 \times \frac{\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \quad (10)$$

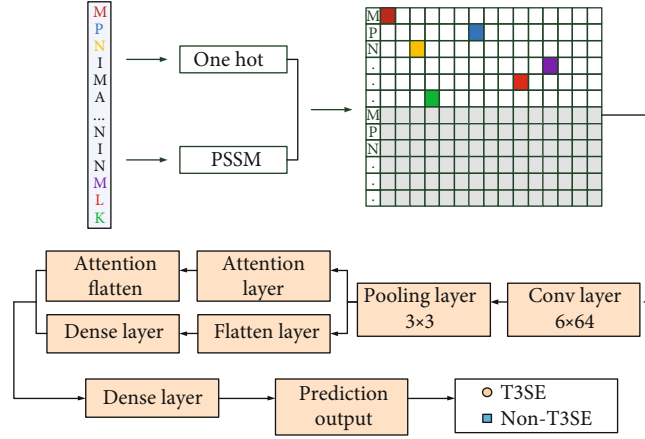


FIGURE 1: ACNNT3 architecture for T3SE prediction. Firstly, 64 1D convolution kernels with a length of 6 are convoluted to generate a 195×64 feature map, and then a 65×64 feature map is obtained through a 3×1 maximum pooling layer. The feature map is then input to the attention and full connection layers, and the two output results are combined to get 66 nodes. Finally, the 66 nodes are fully connected to the two output nodes, and the sigmoid function is used to activate to get the prediction probability of T3SE and non-T3SE.

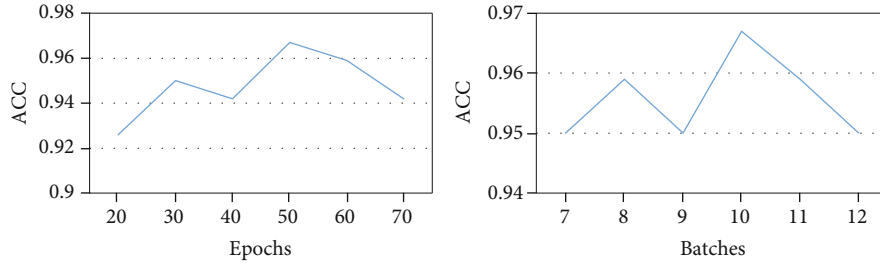


FIGURE 2: ACC comparison on the independent dataset under different epochs and batches.

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \quad (11)$$

$$\text{MCC} = \frac{(\text{TP} \times \text{TN}) - (\text{FN} \times \text{FP})}{\sqrt{(\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TP} + \text{FP}) \times (\text{TN} + \text{FN})}}, \quad (12)$$

where TP, TN, FP, and FN represent the number of true positive, true negative, false positive, and false negative protein datasets, respectively.

The ROC curve is the relationship between the true positive and false positive rates, which is used to measure the comprehensive performance of different methods. The area under the ROC curve (AUC) is commonly used as a summary measure of diagnostic accuracy. In the ROC curve, the horizontal axis is the FPR (false positive rate, i.e., the ratio of wrongly predicted pairs over the total number of negative pairs), and the vertical axis is the TPR (true positive rate, i.e., the ratio of correctly predicted pairs over the total number of positive pairs). The maximum AUC is 1, which means a perfect prediction, and the AUC obtained by a random guess is 0.5.

3. Results

We have constructed a new prediction model to identify T3SEs by using a neural network that combines attention

with CNN. In order to study the influence of the negative sample set on performance, we divided the training set into two parts. The positive to negative ratio of training set 1 is 1:1, and the positive to negative ratio of training set 2 is 1:3. The ACNNT3 model was trained using training sets 1 and 2, respectively. To evaluate the classification performance of our ACNNT3 model, we use ROC and AUC as the evaluation criterion. The ROC charts of 5-fold crossvalidation curves under training sets 1 and 2 are shown in Figures 3(a) and 3(b). We can see that the ACNNT3 model achieved a good performance on the ROC chart. The mean AUC of the model is 0.95 on training set 1 and 0.98 on training set 2. These results show that our ACNNT3 model can accurately classify T3SEs and non-T3SEs on both training sets.

3.1. Comparison of Different Features on the Same Network. We take the one-hot single feature and the fused feature containing the one-hot matrix and PSSM as inputs, respectively, using ACNNT3 as the training model, and use the independent dataset to evaluate the two models. The results show that in all evaluation indexes, the model with the fusion feature is superior to the one with single feature training, thus verifying the proposed fusion feature's effectiveness (Figure 4). Compared to the one-hot single feature, the fusion feature is more comprehensive for the extraction of protein sequence information, and it can be seen from the

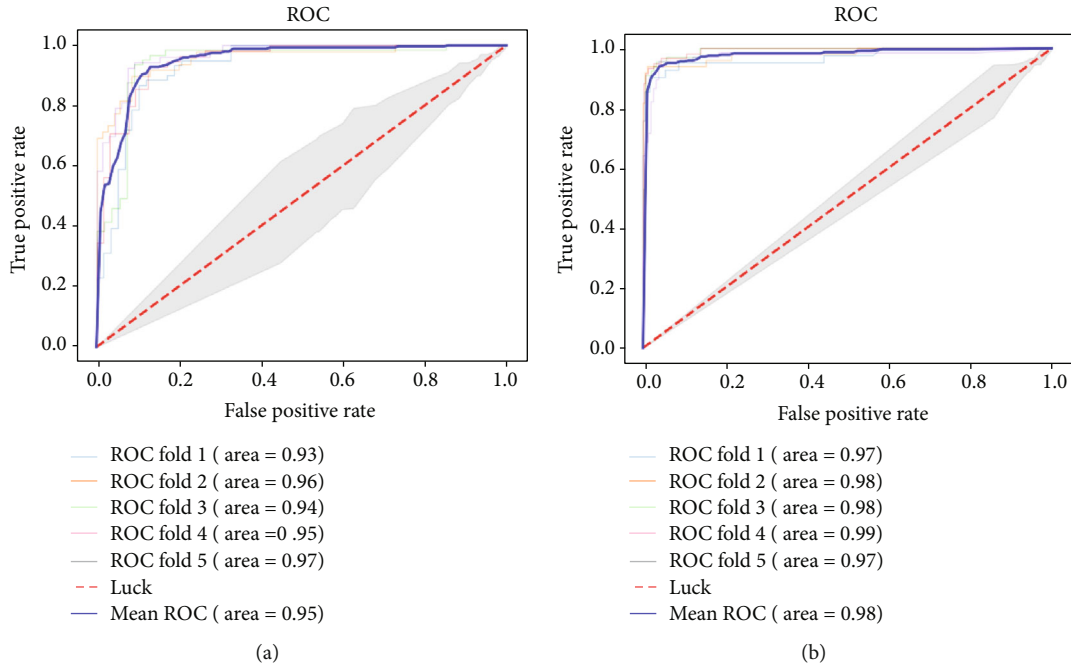


FIGURE 3: ROC curves on different training sets. (a) Use 5-fold crossvalidation experiment on training set 1. (b) Use 5-fold crossvalidation experiment on training set 2.

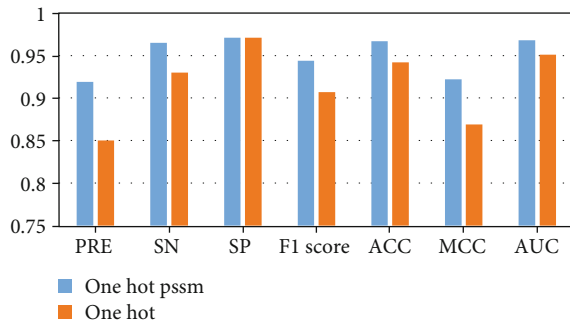


FIGURE 4: Comparison of experimental results of fusion feature and single feature under the same network model.

TABLE 1: Comparison with mainstream deep-learning methods.

Method	PRE	F1 score	ACC	MCC	AUC
ACNNT3	0.919	0.944	0.967	0.922	0.968
DenseNet	0.850	0.907	0.942	0.870	0.951
VGG16	0.846	0.892	0.934	0.847	0.937
ResNet	0.609	0.691	0.838	0.552	0.795
CNN	0.780	0.842	0.901	0.776	0.904
LSTM	0.875	0.933	0.959	0.909	0.961

The bold values indicate the best prediction results.

experimental results that two types of features have good compatibility with each other.

3.2. Comparison of Different Deep-Learning Methods. We compared the results from different popular network models using the independent dataset with the same feature input, as

TABLE 2: Comparison of ACNNT3 and DeepT3, Effective T3, BPBAac, and BEAN2 on an independent dataset.

Method	PRE	SN	SP	F1 score	ACC	MCC	AUC
ACNNT3-1	0.919	0.971	0.965	0.944	0.967	0.922	0.968
ACNNT3-2	0.711	0.914	0.849	0.800	0.868	0.716	0.882
DeepT3-1	0.825	0.943	0.919	0.880	0.926	0.830	0.974
DeepT3-2	0.643	0.771	0.825	0.701	0.810	0.569	0.896
Effective T3	0.542	0.839	0.741	0.658	0.767	0.521	0.803
BPBAac	0.944	0.548	0.988	0.694	0.871	0.656	0.902
BEAN2	0.674	0.935	0.835	0.784	0.862	0.706	0.865

The bold values indicate the best prediction results.

TABLE 3: Comparison of ACNNT3 and DeepT3, Effective T3, BPBAac, and BEAN2 on a *P. syringae* dataset.

Method	PRE	SN	SP	F1 score	ACC	MCC	AUC
ACNNT3-1	0.900	0.976	0.357	0.936	0.887	0.452	0.667
ACNNT3-2	0.872	0.988	0.143	0.926	0.866	0.265	0.565
DeepT3-1	0.905	0.962	0.429	0.932	0.884	0.472	0.838
DeepT3-2	0.913	0.924	0.500	0.918	0.860	0.437	0.763
Effective T3	0.906	0.906	0.428	0.906	0.838	0.334	0.810
BPBAac	0.875	0.494	0.571	0.631	0.505	0.046	0.562
BEAN2	0.883	0.988	0.083	0.938	0.884	0.271	0.607

The bold values indicate the best prediction results.

shown in Table 1. For a class of sequential processing problems, the addition of an attention layer makes the network model strengthen the connection before and after the amino

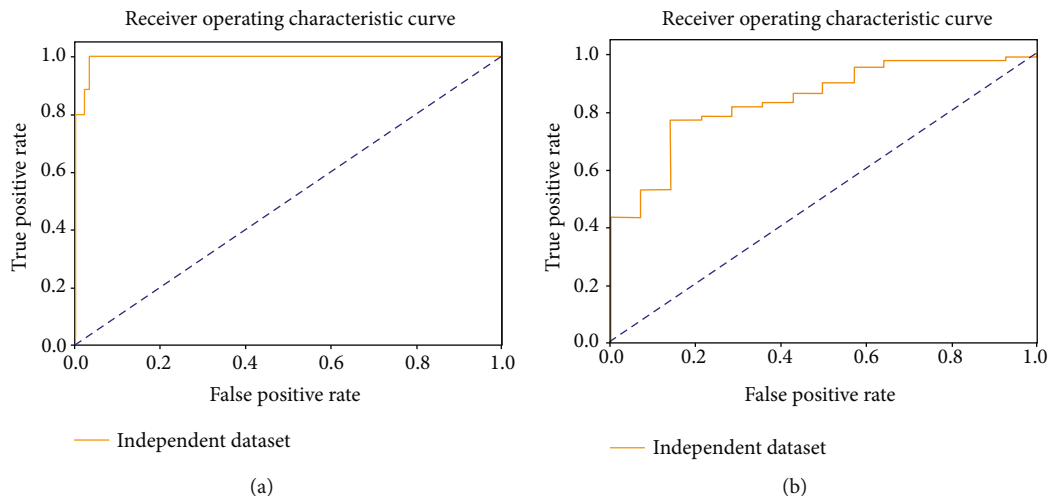


FIGURE 5: ROC curve of the best model selected from the 5-fold crossvalidation on two datasets. (a) ROC curve on a common independent dataset. (b) ROC curve on a *P. syringae* dataset.

acid and the attention of important information in the sequence. From the experimental results, it can be seen that our network model ACNNT3 is better than the existing deep-learning framework for predicting T3SEs in many indicators.

3.3. Comparison with Existing Methods. In order to evaluate the effectiveness of our method, we compared the ACNNT3 performance with four popular methods, DeepT3 [11], BPBAac [22], Effective T3 [6], and BEAN2 [34], on the same independent dataset. The parameter settings of these methods are the same as those used by Li et al. [11]. We found that our ACNNT3-1 model has a higher SN, F1 score, ACC, and MCC than the other four methods (Table 2). The results also show that our method achieved satisfactory performance in almost all indicators. For the important index of ACC, the accuracy of ACNNT3-1 is 0.967, which is 9.9%, 4.1%, 15.7%, 20.0%, 9.6%, and 10.5% higher than ACNNT3-2, DeepT3-1, DeepT3-2, Effective T3, BPBAac, and BEAN2, respectively. In another *P. syringae* dataset, our model still performed better than the existing methods on the index of ACC (Table 3). The accuracy of ACNNT3-1 is 0.887. We selected the best model in the fivefold crossvalidation and used the independent and *P. syringae* datasets to test it. We also obtained the ROC curves of the model on two test sets (Figure 5). Overall, our method has been shown to be superior to all the latest methods in T3SE prediction and is reliably stable.

4. Conclusion

We have proposed a new prediction model for Gram-negative bacteria type III secreted proteins based on a deep neural network. In order to better learn the feature preference of type III secreted proteins, we integrated the one-hot encoding and PSSM extracted from the protein primary sequence as the feature input and embedded the attention layer into CNN to improve the model's prediction ability. This method outperforms other existing methods on most

indicators, and using feature and network model comparisons, we have shown its advantages. In comparison with other popular methods, ACNNT3 is more accurate at predicting and recognizing T3SEs in the independent test set, which reflects its advantages and effectiveness. However, we found that ACNNT3's performance using the *P. syringae* dataset is not particularly obvious and was only slightly higher than the previous methods in terms of ACC and MCC. Our work in the future will focus on achieving better results in other experimental indicators and on applying this model for prediction using other large-scale datasets.

For easy implementation, all data used in this work and the source code for feature computing can be accessible at <https://github.com/Lijiesky/ACNNT3>.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant Nos. 11671009 and 61762035, Zhejiang Provincial Natural Science Foundation of China under Grant Nos. LZ19A010002 and LY18F020027.

References

- [1] M. Desvaux, M. Hébraud, R. Talon, and I. R. Henderson, "Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue," *Trends in Microbiology*, vol. 17, no. 4, pp. 139–145, 2009.
- [2] G. R. Cornelis, "The type III secretion injectisome," *International Journal of Medical Microbiology*, vol. 4, no. 11, pp. 811–825, 2006.

- [3] S. Y. He, K. Nomura, and T. S. Whittam, "Type III protein secretion mechanism in mammalian and plant pathogens," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1694, no. 1-3, pp. 181–206, 2004.
- [4] G. N. Schroeder and H. Hilbi, "Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by type III secretion," *Clinical Microbiology Reviews*, vol. 21, no. 1, pp. 134–156, 2008.
- [5] J. Engel and P. Balachandran, "Role of *Pseudomonas aeruginosa* type III effectors in disease," *Current Opinion in Microbiology*, vol. 12, no. 1, pp. 61–66, 2009.
- [6] R. Arnold, S. Brandmaier, F. Kleine et al., "Sequence-based prediction of type III secreted proteins," *PLoS Pathogens*, vol. 5, no. 4, article e1000376, 2009.
- [7] M. Löwer and G. Schneider, "Prediction of type III secretion signals in genomes of gram-negative bacteria," *PLoS One*, vol. 4, no. 6, article e5917, 2009.
- [8] X. Dong, Y.-J. Zhang, and Z. Zhang, "Using weakly conserved motifs hidden in secretion signals to identify type-III effectors from bacterial pathogen genomes," *PLoS One*, vol. 8, no. 2, article e56632, 2013.
- [9] S. Xie, Z. Li, and H. Hu, "Protein secondary structure prediction based on the fuzzy support vector machine with the hyperplane optimization," *Gene*, vol. 642, no. 5, pp. 74–83, 2018.
- [10] X. Yang, Y. Guo, J. Luo, X. Pu, and M. Li, "Effective identification of gram-negative bacterial type III secreted effectors using position-specific residue conservation profiles," *PLoS One*, vol. 8, no. 12, article e84439, 2013.
- [11] L. Xue, B. Tang, W. Chen, and J. Luo, "DeepT3: deep convolutional neural networks accurately identify gram-negative bacterial type III secreted effectors using the n-terminal sequence," *Bioinformatics*, vol. 35, no. 12, pp. 2051–2057, 2019.
- [12] H. Hu, Z. Li, A. Elofsson, and S. Xie, "A Bi-LSTM based ensemble algorithm for prediction of protein secondary structure," *Applied Sciences*, vol. 9, no. 17, article 3538, 2019.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [14] Z. Li, J. Zhu, X. Xu, and Y. Yao, "RDense: a protein-RNA binding prediction model based on bidirectional recurrent neural network and densely connected convolutional networks," *IEEE Access*, vol. 8, pp. 14588–14605, 2020.
- [15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," <http://arxiv.org/abs/1409.1556>.
- [16] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology," *Molecular Systems Biology*, vol. 12, no. 7, p. 878, 2016.
- [17] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018.
- [18] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," <https://www.aclweb.org/anthology/N16-1174.pdf>.
- [19] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2204–2212, 2014.
- [20] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," <http://arxiv.org/abs/1409.0473>.
- [21] P. Verga, E. Strubell, and A. McCallum, "Simultaneously self-attending to all mentions for full-abstract biological relation extraction," <https://arxiv.org/pdf/1802.10569>.
- [22] Y. Wang, Q. Zhang, M. Sun, and D. Guo, "High-accuracy prediction of bacterial type III secreted effectors based on position-specific amino acid composition profiles," *Bioinformatics*, vol. 27, no. 6, pp. 777–784, 2011.
- [23] D. M. M. Tay, K. R. Govindarajan, M. A. Khan et al., "T3SEdb: data warehousing of virulence effectors secreted by the bacterial type III secretion system," *BMC Bioinformatics*, vol. 11, Suppl 7, p. S4, 2010.
- [24] Y. Huang, B. Niu, Y. Gao, L. Fu, and W. Li, "CD-HIT suite: a web server for clustering and comparing biological sequences," *Bioinformatics*, vol. 26, no. 5, pp. 680–682, 2010.
- [25] D. A. Baltrus, M. T. Nishimura, A. Romanchuk et al., "Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates," *PLoS Pathogens*, vol. 7, no. 7, article e1002132, 2011.
- [26] Y. An, J. Wang, C. Li et al., "Comprehensive assessment and performance improvement of effector protein predictors for bacterial secretion systems III, IV and VI," *Briefings in Bioinformatics*, vol. 19, no. 1, pp. 148–161, 2018.
- [27] J. D. Bendtsen, L. J. Jensen, N. Blom, G. von Heijne, and S. Brunak, "Feature-based prediction of non-classical and leaderless protein secretion," *Protein Engineering Design and Selection*, vol. 17, no. 4, pp. 349–356, 2004.
- [28] C. Casper-Lindley, D. Dahlbeck, E. T. Clark, and B. J. Staskawicz, "Direct biochemical evidence for type III secretion-dependent translocation of the AvrBs2 effector protein into plant cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 8336–8341, 2002.
- [29] Y. Yang, J. Zhao, R. L. Morgan, W. Ma, and T. Jiang, "Computational prediction of type III secreted proteins from gram-negative bacteria," *BMC Bioinformatics*, vol. 11, article S47, 2010.
- [30] K. Schesser, E. Frithz-Lindsten, and H. Wolf-Watz, "Delineation and mutational analysis of the *Yersinia pseudotuberculosis* YopE domains which mediate translocation across bacterial and eukaryotic cellular membranes," *Journal of Bacteriology*, vol. 178, no. 24, pp. 7227–7233, 1997.
- [31] M. P. Sory, A. Boland, I. Lambermont, and G. R. Cornelis, "Identification of the YopE and YopH domains required for secretion and internalization into the cytosol of macrophages, using the *cyaA* gene fusion approach," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 26, pp. 11998–12002, 1996.
- [32] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.
- [33] H. B. Shen and K. C. Chou, "Virus-mPLOC: a fusion classifier for viral protein subcellular location prediction by incorporating multiple sites," *Journal of Biomolecular Structure and Dynamics*, vol. 28, no. 2, pp. 175–186, 2010.
- [34] X. Dong, X. Lu, and Z. Zhang, "BEAN 2.0: an integrated web resource for the identification and functional analysis of type III secreted effectors," *Database*, vol. 2015, article bav064, 2015.