American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-10, Issue-05, pp: 163-169
www.ajer.org
Research Paper Open Access

Recognition of Antimicrobial Peptides by Neural Networks

Sophio Barnovi
Georgian Technical University, s.barnovi@gtu.ge
Mariam Chkhaidze
Georgian Technical University, m.chkhaidze@gtu.ge
Nino Mchedlishvili
Georgian Technical University, n.mchedlishvili@gtu.ge

Abstract
Frequent use of traditional antibiotics develops resistance against bacteria thus making it more difficult to develop new antibiotics. Fast spreading of medicine resistant infections has become a challenge for anti-microbe therapies. It is much more difficult for bacteria to develop resistance against antimicrobial peptides. Therefore, antimicrobial peptides can become a good alternative for antibiotics. Scientists hope that they will be able to create a medicine with anti-microbe action that will strengthen the immune system of an organism in case of acute bacterial and virus diseases. Neural networks as one of the important directions of machine learning are used for AMP's identification. It is significant for artificial neural network researchers to develop effective algorithms for learning. Therefore, the researches aim at refining identification methods to achieve the highest reliability of identification.This work reviews various methods of artificial neural networks, represents and describes the process of AMP identification via selected artificial neural networks. The identification procedures, such as neural network learning and identification, are performed in MatLab. We have chosen a feed-forward, hierarchical and recurrent neural networks for neural network learning. This learning is implemented with several algorithms. We reviewed the respect results of the research for each algorithm; also, we compare the algorithms. From all the classification methods the Ensemble Method gives us the best result – 93,8% identification accuracy, the second one is the Support Vectors Method (SVM) – 92,3% accuracy. The rest of the methods give us good results – 80 – 90%.
KeyWords: *Antimicrobial peptides, artificial neural networks, recognition.*

Date of Submission: 27-04-2021 Date of acceptance: 11-05-2021

I. INTRODUCTION

Antimicrobial peptides (AMPs) are short size molecules (12-50 amino acids) [1, 3]. They are widespread in living organisms and are the defense peptides that destroy pathogenic microorganisms. Studies have revealed a wide range of antimicrobial peptides that show activity against the pathogens resistant to the modern antibiotics. Antimicrobial peptides have broad-spectrum activity against gram-positive, gram-negative bacteria, fungi, viruses and the simplest microorganisms. Besides this, antimicrobial peptides reveal their antimicrobial action against antibiotic resistant bacteria strains. Accordingly, it is much more difficult for bacteria to develop resistance against antimicrobial peptides. Thus, antimicrobial peptides can be a good alternative to antibiotics. In order to implement projects on AMPs-based antibiotics it is necessary to study the dependency between their structure and activity. The first stage in the new medicine project is to determine the antimicrobial activity of a peptide. The goal of this work is to identify the antimicrobial peptides based on the current peptide pool.

A Database of Antimicrobial Activity and Structure of Peptides (DBAASP) was created at the bioinformatics laboratory of Tbilisi Experimental Medicine Center (Republic of Georgia) [1]. It contains information about more than 4,500 peptides. The database allows us to form experimentally proven AMP (positive) and non-AMP (negative) multiplicities.

1.1. Literature Overview

According to the worldwide data, antimicrobial peptides identification researches are carried out via the methods of deep learning, genetic algorithms, neural networks etc., though the required precision has not been reached yet. Researchers try to use artificial neural networks to identify AMPs. This method allows us to identify AMPs with fewer expenses. For example:

In 2015, Malaysian scientists carried out research to develop a new method of antimicrobial peptides prognosis. The researchers are interested in developing a new alternative preparation based on AMPs, because it is proved that a big number of bacteria strains are resistant to antibiotics. They encountered many difficulties during the studies because the experiments (retrieving AMPs from the protein sequence) are expensive and take a long time. The scientists suggested the algorithm that may allow researchers to identify peptides from unknown proteins with higher sensitivity [18].

Scientists of Georgia Technical University, in the research carried out by Michael Youmans and others (2017) described how to identify antimicrobial peptides via machine learning. They used Long Short-Term Memory (LSTM) to classify AMPs. This method uses the Random Forest and the K-nearest classification algorithms [7].

American scientists Daniel Veltri and others used the convolutional deep learning neural network (2018) to identify antimicrobial peptides [6].

In 2019, American researchers used a multilayer convolutional neural network with filters. This architecture of DNN model is normally used in Natural Language Programming (NLP) and it is not suitable for AMP identification. They developed a multilayer convolutional network that contains filters of different sizes. Based on the multilayer convolutional network they created a DNN model to improve AMP identification [9].

A stage of data preparation for an artificial neural network contains creation of a database from the signs characteristic for the task under question. Then, the database is divided on learning and testing multiplicities. Sometimes, it also requires validation multiplicity.

Antimicrobial peptides have their own characteristic signs that help us in formation of a sign space: Hydrophobic moment; Hydrophobicity; Charge; Isoelectric point; Depth; Tilt angle; Disordering; Linear moment; Aggregation; Helical Wheel; ppII Propensity; Amphiphilicity index.

We carried out the first research via a feed-forward backpropagation neural network and various learning algorithms that are recommended for difficult problems solutions. These are: Algorithm of Levenberg-Marquardt (trainLM), Algorithm of Gradient Descent backprop (train GD), Bayesian regularization (train BR), Scaled Conjugate gradient (train SCG); Random order incremental (train R);Quasi – Newton backpropagation (train BFG) [1, 15].

To build the network we select the number of layers and neurons. It is known that the number of the compact clusters must be equal to that of the formal neurons or types; activation function – hyperbolic tangent; the work was carried out via the feed-forward network though with different learning algorithms. The results for each of them are given in the table N1. Based on these results we evaluate the network performance quality according to the respect criteria [4].

feed - forward backprop		
№	***Learning Method***	***Identification Reliability, %***
1.	Levenberg-Marquardt	76% reliability
2.	Gradient algorithm	77% reliability
3.	Bayesian Regularization	75% reliability
4.	Scale conjugate backpropagation	75% reliability
5.	Random order incremental	74% reliability
6.	BFG - Quasi-Newton backprop	70% reliability

Table I. The percentage of the results received from the learning of feed-forward backprop.

Based on the primary results we can conclude that we have 77% percent of the identification reliability. As a primary conclusion, we notice that the identification results vary between 70-77%. Considering that we

have not used any additional procedures during the network learning (additional procedures imply signs marking, clustering) the results are quite good.

II. NEURON NETWORKS LEARNING PROCESS

A cluster is a compact one if it aggregates only one type of realizations. Accordingly, if a cluster aggregates more than one type of realizations, then it is a non-compact cluster. Thus, it is desirable to evaluate a degree of insulation. In case of neural networks, we could use Net multiplicities, namely, Net_i and Net_j values.

Net_i and Net_j values are positive and one dimensional, i. e. their values are located on $[0; \infty]$ axis. Accordingly, clusters, if they exist, are located in the same range. One dimensional compact clusters are presented in the figure below [12].

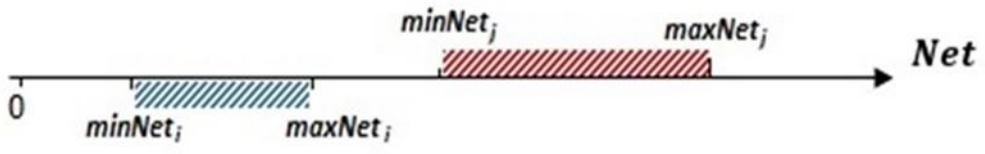

Figure. 1. Compact Clusters

Insulation, i. e. compactness means that the inequality:

$$minNet_j - maxNet_i = \Delta NET_{ij} > 0$$

ΔNET_{ij} value is a degree of compactness, namely, the more is the ΔNET_{ij} value, the more is a degree of A_i and A_j clusters and vice versa – the less are A_i and A_j clusters insulated (see the figure. 2).

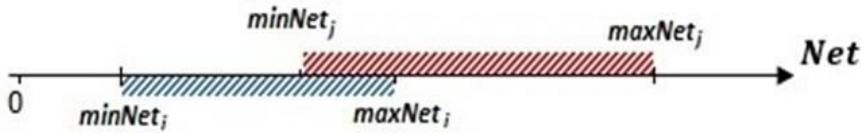

Figure. 2. Non-Compact Clusters

There are such non-compact clusters that contain realizations of two or more types. It is clear that for non-compact clusters $\Delta NET_{ij} < 0$. Therefore, the goal of the learning process is to make such changes in the weight coefficients of the neurons at the clusters intersection zone that will turn non-compact clusters into the compact ones.

In case we do not get any insulated clusters in the characteristic signs space, then the high reliability identification process is limited or almost impossible.

As we have noticed, the signs space consists of empirically measurable signs that are peptide characteristic, descriptive signs. First, we made up compact and non-compact antimicrobial peptides to analyze the signs. The results for each of the signs reflect the range of signs changes. We can use them on the identification and clustering (or preliminary selection) stages as well. In our instance, we have 12 characteristic signs. For each of these signs we make up the signs changes range that will result in selection of characteristic signs [15, 17].

In the given signs, non-compact clusters prevail. On this basis, we derive the characteristic signs that have less impact on the identification process. We shall take out three signs from the database. We presented them in different colors in the table below. They are as follows: Depth, Linear moment and Aggregation. These signs imped the identification process instead of improving the network learning process [18].

From the data given in table 2, we can clearly see that we have only two classes to identify. Via the realizations (type descriptions), we get the multiplicities of the types that are difficult to insulate (non-compact). It makes the task more complicated.

The most important component of an artificial neural network architecture is its learning algorithm. Artificial neural networks can be built via some programming language as well as via standard libraries and applications. We chose Matlab as a working environment for our research. We carried out experimental research of various neural networks. Namely, we built feed-forward backprop, cascade-forward backprop and Layer Recurrent. We teach the networks using various learning algorithms. Accordingly, we described the research results and compared them to each other [2].

AMPs	*Hydrophobic Moment*	*Hydrophobicity*	*Charge*	*Isoelectric Point*	*Depth*	*Tilt angle*	*Disordering*	*Linear Moment*	*Aggregation*	*Helical Weel*	*ppII Propensity*	*Amphiphilicity index*
max	1.51088	1.63	11	14	30	145	0.76334	0.56142	659.32397	250	4.55909	1.26615
min	0.03612	-0.46294	-0.46294	9.58006	6	5	-1.57445	0.1575	0	20	0.11154	0.75562
NAMPs	*Hydrophobic Moment*	*Hydrophobicity*	*Charge*	*Isoelectric Point*	*Depth*	*Tilt angle*	*Disordering*	*Linear Moment*	*Aggregation*	*Helical Weel*	*ppII Propensity*	*Amphiphilicity index*
max	1.25539	1.78	12	14	30	172	0.88303	0.55565	1004.05	330	3.35615	2.17
min	0.00446	-0.47615	-4	2.79	2	5	-1.32678	0	0	0	0	0.766

Table II. Selection of the Antimicrobial Peptides Types Descriptive Signs

As we know, building a neural network implies its structure selection. Next, we evaluate the network learning and its quality. A network building process also implies setting all the necessary parameters that are important for its proper performance. These parameters are the following: number of layers, number of neurons in the layer that must be equal to the number of given classes and weights, entering the values of the weights and displacements, selection of the respect algorithm and activation function. Finally, we selected the recurrent network [5, 10].

In backpropagation neural networks signals are transmitted to the next layer as well as to the previous one. Teaching a neural network via a recurrent function provides important signs and characteristics for the given type that take non-zero and great values compared to other signs. From the identification point of view, the greater the difference between the various types of important signs the more effective is the process.

Recurrent neural networks (Layer Recurrent), i.e. Back-Propagation Networks. In this type of network, the signal coming out of the neuron is partially transmitted to the incoming neurons. It consists of three hidden layers. The number of neurons is two. Performance function – MSE (Mean Square Error), adaptation learning function – Learn GD (Network Learning Function that renews the values of the weight and displacement (Bias) via the gradient descent method), learning algorithm Levenberg-Marquardt (trainLM) – Levenberg Training Algorithm. The structural scheme is given in figure 3 and the network learning process is shown in figure 4.

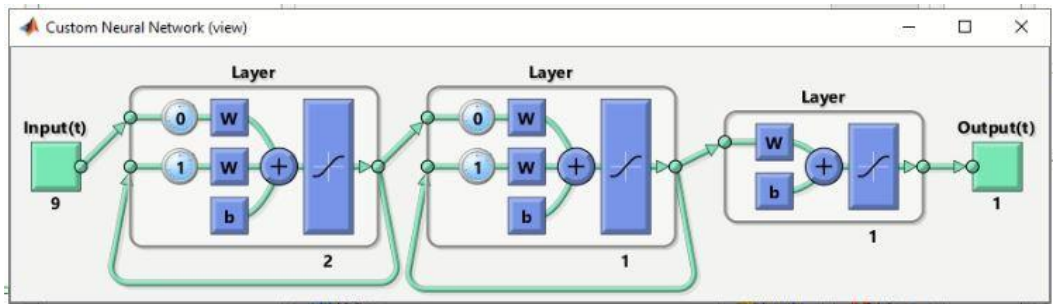

Figure. 3. A Structural Scheme of the Recurrent Network.

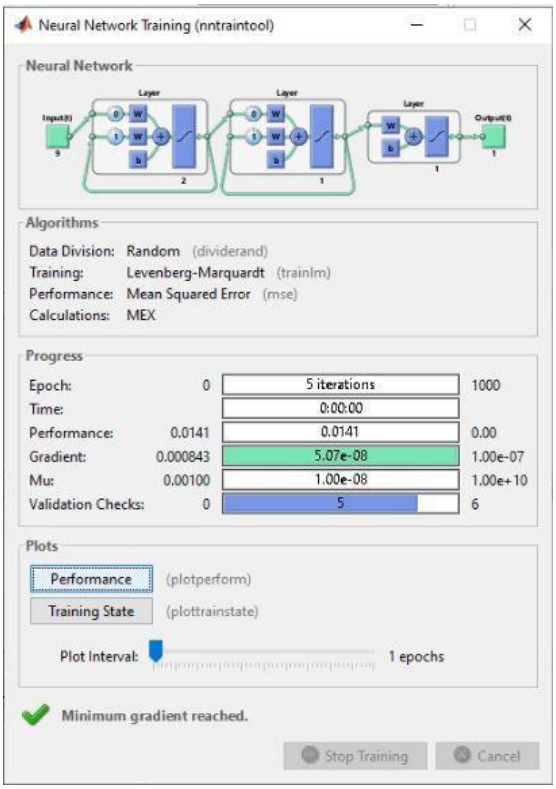

Figure. 4. The Network Learning Process

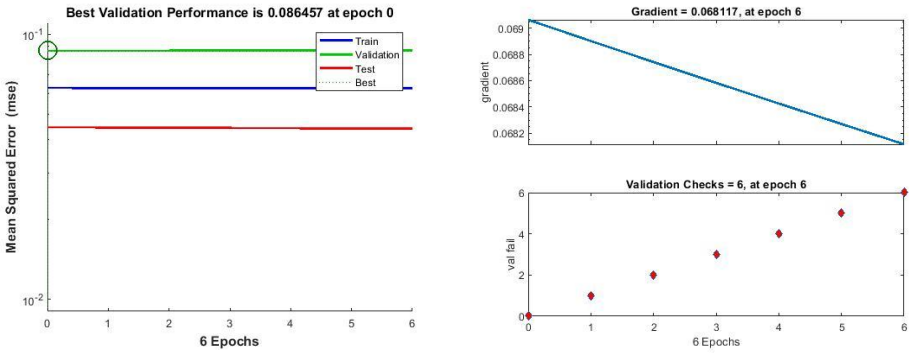

Figure. 5. Levenberg Training Method Graphs.

The results of the recurrent network learning research allow us to conclude that the identification is quite high and there are almost no errors.

During the neural network learning process, we identified the antimicrobial peptides via various classification methods. We can use the cleaned base for various methods of classification: Decision Trees, Discriminant Analysis, Logistic Regression Classifiers, Naive Bayes Classifiers, Support Vector Machines, Nearest Neighbor Classifiers, Ensembles Classifiers [10, 11].

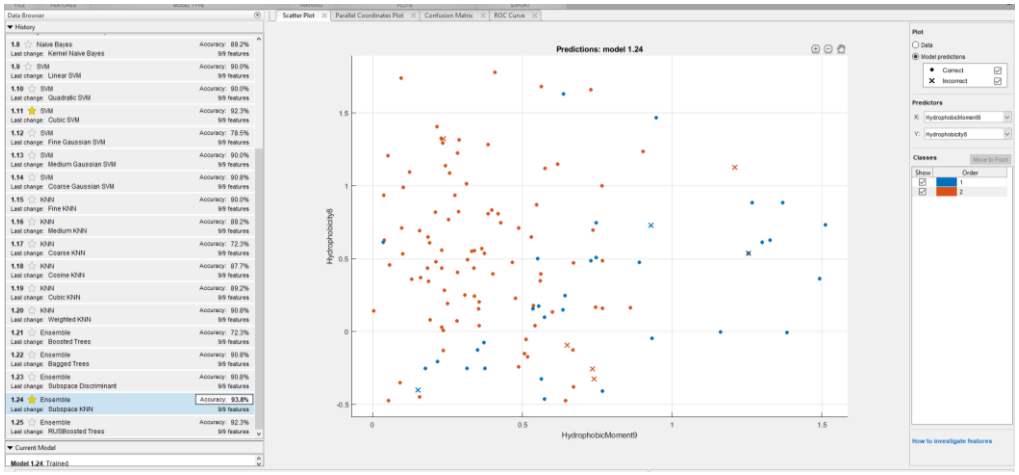

Figure. 6. The Methods of Classification of the Processed Base.

As we can see, the base processing improved the results. Namely, the Ensemble Method gave the best result with 93,8% of accuracy and Support Vectors Method – 92,3%.

Ensemble Method is a machine-learning algorithm that combines several basic methods to create one optimal predictable model. The goal of any machine learning problems is to find such a model that will provide the best prediction for our desirable results. The Decision Trees are not the only form of the Ensemble Method. They are the most popular method in science according to the fresh data [23].

The rest of the classification methods give us a positive result, too. It varies between 80-90%. However, the recurrent neural networks identify with even higher accuracy that is up to 95%.

III. CONCLUSION

Based on the current peptides base that allowed formation of experimentally proved AMP and non-AMP types we implemented the following:

Analyzed and evaluated the peptide descriptive signs in the multiplicities (types) of the amino acid sequences; Processed the data preliminarily and selected the respect signs space for the task; Analyzed the characteristic signs; Used the clustering procedure in order to evaluate the signs space; Made up compact and non-compact clusters for the types of antimicrobial and non-antimicrobial peptides. The results for each type reflect the range of sign changes. We could use it on the identification (preliminarily selection) and clustering stages; Carried out the procedures of normalization of the given data in numeric form; Divided the database on the learning and control (test) multiplicities and formed validation multiplicity; Developed a method to recognize antimicrobial peptides using an artificial neural network for such features [15], with which we obtained many difficult-to-isolate types [14]. Thus, we refined the identification methods to get the highest reliability of identification; Built a neural network and selected its structure and typology (number of the layers, number of neurons in the layer, the connection between the neurons); Chose the learning algorithm for the neural network, neuron activation functions and classification methods; Selected the architecture for the neural network that provides the best results; Made a final evaluation of the network performance quality that is the most important and very difficult task; Reviewed the methods of learning, identification, classification and clustering for the task solution; Defined the network parameters and developed its reliability [16].

The research was carried out for various types of neural networks and we finally selected recurrent neural networks that brought us the best results. We used various classification algorithms for identification. The most important methods turned out to be Support Vector Method and Ensemble Method. Among the classification methods the best results were given by Ensemble Method with 93,8% of accuracy and Support Vector Method with 92,3% of accuracy. The rest of the methods bring good results that vary between 80-90%. After we reselected the base the recurrent neural network showed even the better result. Thus, the identification reliability increased up to 95%. Thus, considering that the identification was implemented with a quite good result, it is possible to solve similar tasks with the respect methods.

REFERENCES

[1]. Vishnepolsky B., Pirtskhalava M. Prediction of Linear Antimicrobial Peptides Based on Characteristics Responsible for Their Interaction with the Membranes. Journal of Chemical Information and Modeling, 2014, 54, 5, 1512–1523.
[2]. Martin T. Hagon, Howard B. Demuth Neural Network Toolbox, 2016 by The MathWorks, Inc.
[3]. Lee Y., Lee W., Fulan M., Ferguson L., Wong GCL. What can machine learning do for antimicrobial peptides, and what can antimicrobial peptides do for machine learning? Interface Focus 7, 2017.

[4]. Fang C., Moriwaki Y., Li Caihong L., Shimizu K. prediction of Antifungal peptides by Deep Learning with Character Embedding. IPSJ Transactions on Bioinformatics Vol.12, 21-29, 2019.
[5]. Hamid M., Friedberg I., Identifying antimicrobial peptides using word embedding with deep recurrent neural networks. Bioinformatics, Vol. 35, Issue 12, 2019, 2009-2016 p.
[6]. Veltri D., Kamath U,Shehu A. Deep learning improves antimicrobial peptide recognition. Bioinformatics, Vol 34, Issue 16, 2018, 2740-2747 p.
[7]. Youmans M., Spainhour C., Qui P. Long short term memory recurrent neural networks for antibacterial peptide identification, IEEE,multi 2017.
[8]. Guangshun Wang, Antimicrobial Peptides, Boston, USA, 2017.
[9]. Su X., Xu J., yin Y., Quan X. Antimicrobial peptide identification multi-scale convolutional network. BMC, 2019.
[10]. Koivo H.N. Neural Networks: Basics using MATLAB. 2000, 54 p.
[11]. Paluszek M., Thomas S. Matlab Machine Learning. SSBM Inc. USA, 2017, 335 p.
[12]. Teodoridis S., Koutroumbas K. An Introduction to Pattern Recognition. Matlab Approach. Elsevier Inc, USA, 2010, 216 p.
[13]. Chuncai Z,Chuncai Z. Design, Synthesis and Applications of Antimicrobial Peptides and Antimicrobial Peptide-Mimetic Copolymers[J]. Progress in Chemistry, 2018, 30(7): 913-920.
[14]. Wang G., Li X., Wang Z. The Antimicrobial peptides database as a tool for research and education. Nucleic Acids Research, 2015, Vol. 44, pp 1087-1093.
[15]. Gogoladze G., Grigolava M., Vishnepolsky B., Chubinidze M., Duroux P., Lefranc M., Pirtskhalava M. DBAASP: Database of Antimicrobial Activity and Structure of Peptides. FEMS MicrobioL Lett, 2014.
[16]. Zhang L., Gallo R. Antimicrobial peptides. Current Biology 26, 2016, R1–R21.
[17]. Wang S., Zeng X., Yang Q., Qiao S. Antimicrobial Peptides as Potential Alternatives to Antibiotics in Food Animal Industry. Int Mol Sci, V.17(5), 2016.
[18]. Xin Yi Ng, Bakhtiar Affendi Rosdi, Shahriza Shahrudin, "Prediction of Antimicrobial Peptides Based on Sequence Alignment and Support Vector Machine-Pairwise Algorithm Utilizing LZ-Complexity", BioMed Research International, vol. 2015, Article ID 212715, 13 pages, 2015.
[19]. R.P.Lippmann, "An Introduction to Computing with Neural Nets", IEEE ASSP Magazine, Vol.4, No.2, Apr. 1987.
[20]. Sebastian R., Vahid M. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow, 2nd Edition, Packt Publishing, 2017, 595p.
[21]. Géron A. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc, 2019, 799p.
[22]. Goodfellow I, Bengio Y. Deep Learning (Adaptive Computation and Machine Learning), National Bureau of Economic Research, Inc, 2019, 625p.
[23]. Murty M.N., Raghava R. Support Vector Machines and Perceptrons: Learning, Optimization, Classification, and Application to Social Networks, 2016.

Sophio Barnovi, et. al. "Recognition of Antimicrobial Peptides by Neural Networks."*American Journal of Engineering Research (AJER)*, vol. 10(5), 2021, pp. 163-169.