

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Optimization ACE inhibition activity in hypertension based on random vector functional link and sine-cosine algorithm



Mohammed Abd Elaziz^{a,c,*}, Ahmed Abdelmonem Hemedan^b, Marek Ostaszewski^b, Reinhard Schneider^b, Songfeng Lu^{a,d,**}

^a School of Computer Science & Technology, Huazhong University of Science and Technology, Wuhan 430074, China

^b LCSB - Luxembourg Centre for Systems Biomedicine, Bioinformatics core unit, Luxembourg University, Luxembourg

^c Department of Mathematics, Faculty of Science, Zagazig University, Egypt

^d Shenzhen Huazhong University of Science and Technology Research Institute, Shenzhen, 518063, China

ARTICLE INFO

Keywords:

Bioactive peptides

Angiotensin-converting enzyme (ACE)

Random vector functional link (RVFL)

Sine-cosine algorithm (SCA)

ABSTRACT

Bioactive peptides from protein hydrolysates with antihypertensive properties have a great effect in health, which warrants their pharmaceutical use. Nevertheless, the process of their production may affect their efficacy. In this study, we investigate the inhibitory activities of various hydrolysates on angiotensin-converting enzyme (ACE) in relation to the chemical diversity of corresponding bioactive peptides. This depends on the enzyme specificity and process conditions used for the production of hydrolysates. In order to mitigate the uncontrolled chemical alteration in bioactive peptides, we propose a computational approach using the random vector functional link (RVFL) network based on the sine-cosine algorithm (SCA) to find optimal processing parameters, and to predict the ACE inhibition activity. The SCA is used to determine the optimal configuration of RVFL, improving the prediction performance. The experimental results show that the performance measures of the proposed model are better than the state-of-the-art methods.

1. Introduction

Functional properties of peptides with antihypertensive effect are directly related to their chemical structure [1]. Therefore, the structure-function relationship of peptides from protein hydrolysates needs to be analyzed in depth to allow optimal use of their chemical diversity by the chemical industry. In this study, we measured the concentration of bioactive amino acids in different chemical structures of peptides, and examined the effect of different temperature levels on their ACE inhibition activity, with the aim to control the chemical alteration of hydrolysates. The uncontrolled chemical alteration in hydrolysates may lead to adverse health effects in hypertensive patients [2]. A part of that, ACE inhibition activity is affected by both amino acid type and chain length [3]. Furthermore, the negative chemical charge of acidic amino acids (aspartic and glutamic) reduces ACE inhibition activity [4]. In turn, the presence of arginine, phenylalanine, valine, proline, lysine, tryptophan, leucine, isoleucine and tyrosine in ACE inhibitory peptides increases the strength of ACE binding [5]. Therefore, the drug industry

struggles with the challenge of reducing chemical diversity of the production of hydrolysates and evaluating its effects on functional properties of peptides. Thus, it is important to develop fast, and accurate statistical models to predict and account for the influences of chemical alteration on ACE inhibition activity of bioactive peptides.

The relationship between the chemical structure of a peptide and its ACE inhibitory activity still remains to be fully elucidated [6]. Previous studies used the quantitative structure-activity relationship (QSAR) model to predict ACE inhibition activity [7]. However, this model is limited to the structure, and does not consider recently identified features that affect the structure itself. Noteworthy, our model highlights the impact of process conditions on the chemical alteration in the peptide composition, allowing to account for this alteration and further optimize ACE inhibition activity.

Several machine learning methods were applied to a range of regression problems. These include artificial neural networks (ANN) with one or more hidden layers and the backpropagation training [8,9] and the Support Vector Regression (SVR) [10,11]. In an interesting study, an

* Corresponding author. School of Computer Science & Technology, Huazhong University of Science and Technology, Wuhan 430074, China.

** Corresponding author. School of Computer Science & Technology, Huazhong University of Science and Technology, Wuhan 430074, China.

E-mail addresses: abd_el_aziz_m@yahoo.com (M.A. Elaziz), dr_hemedan@yahoo.com (A.A. Hemedan), marek.ostaszewski@uni.lu (M. Ostaszewski), reinhard.schneider@uni.lu (R. Schneider), lusongfeng@hust.edu.cn (S. Lu).

<https://doi.org/10.1016/j.chemolab.2019.05.009>

Received 22 December 2018; Received in revised form 10 May 2019; Accepted 20 May 2019

Available online 26 May 2019

0169-7439/© 2019 Elsevier B.V. All rights reserved.

adaptive neuro-fuzzy inference system (ANFIS) model have been used to predict the bioactive peptides in relation to climate change [1].

An interesting variant of neural network paradigm, the random vector functional link (RVFL) network, introduces the direct connection between the input and output neurons [12,13]. The RVFL, is an ANN with a single layer with the input-output connection, and the nodes of its hidden layer are called enhancement nodes [14]. In the RVFL, the weights between input and enhancement nodes are randomly generated from a suitable interval. This interval should constrain the activation functions away from the saturation region. The weights between input and output nodes, and between enhancement and output nodes, are updated during the training process by optimization techniques such as a conjugated backpropagation, or least squares method.

To date, the RVFL has been applied to various applications, including short-term electricity load demand forecasting [15], remote sensing applications [16], forecasting the distribution of the temperature [17], and Big Data processing [18]. In order to improve the performance of the RVFL network, several methods have been applied. Chen and Wan [19] introduced two versions of RVFL to determine the optimal weights and update them on-the-fly. Scientists presented another improvement to RVFL to solve nonlinear dynamic systems [20] by improving the input pattern with nonlinear functional expansion. In another line of research, Scardapane et al. proposed a distributed RVFL algorithm to improve the efficiency of the RVFL [21]. Noteworthy is the RVFL approach, using the Bayesian inference for robust data modeling [22].

However, in the previous studies, the optimization of several parameters in RVFL was a challenge, as it is an NP-hard problem. To overcome this challenge, Zhang and Suganthan [23] presented a comprehensive study to optimize the parameters of RVFL and to improve its performance in classification and regression tasks. But their method is time-consuming as it requires manual permutation [24,25]. This motivated us to propose an automated method for selecting the optimal configuration of RVFL parameters in order to optimize the ACE inhibition activity prediction. In order to determine this configuration, we used the sine-cosine algorithm (SCA) [26].

For the first time, to the best of our knowledge, we propose a descriptive machine learning paradigm, based on RVFL and SCA, to investigate the chemical diversity of bioactive peptides in order to predict their ACE inhibition activity in hypertensive patients. The proposed method, SCA-RVFL, starts by generating a random population where the dimension of each solution represents the number of parameters used for building the RVFL networks. The dataset is divided into training and testing set randomly. The fitness function of each solution is computed by the RVFL built using the parameters of the current solution. The fitness is calculated as root mean square error between the original ACE and predicted ACE. Overfitting of the predicted inhibition activity is evaluated using the testing set. The next step is to determine the best solution, best configuration, and then update the solutions using the sine and cosine functions. These steps are repeated until the stop conditions are met.

The rest of this paper is organized as follows: in Section 2, materials and methods are discussed, including the preparation of the dataset, the RVFL and SCA methods. In Section 3, the proposed method is introduced. Section 4 presents the experimental results and the discussion. Finally, the conclusion and the future works are introduced in Section 5.

2. Materials and methods

2.1. Dataset collection

Byproducts of one hundred and twenty *Tilapia Nilotica* fish were collected for twelve months from April 2016 to March 2017 at different localities in Egypt, one of the most important areas of tilapia production worldwide. Measured parameters of the process include but not limited to temperature, humidity, weight, length, sex, ration and water quality.

The samples were prepared follow the enzymatic hydrolysis protocol

[27]:

- Thaw fish byproducts in 4 °C for 12 h.
- Mix 15% of the byproducts' volume with 50 ml of phosphate buffer saline (pH 7.5).
- Pre-incubate at 60 °C for 20 min.
- Initiate the enzymatic hydrolysis reaction with the alcalase enzyme (2.5%)
- Heat in a water bath at 90 °C for 15 min then cool on ice 2–3 °C.
- Centrifuge for 20 min at 10000 rpm then measure the hydrolysis degree
- Extract then frozen supernatant and analyze.

2.2. Preparation phase

2.2.1. Analysis of tilapia byproducts and its hydrolysates

Tilapia byproducts and its hydrolysates were analyzed according to AOAC protocol [28]. The pure protein percentage was measured using the Kjeldahl method [28]. Moisture content was estimated using the drying method [28]. Ash percentage was calculated using muffle furnace. The proteins from byproducts were hydrolyzed with an alcalase enzyme in an estimated PH = 5 and temperature = 30.

2.2.2. Amino acid sequence analysis

The gel buffer was used to prepare the stacking and separating gel in 4% and 16% respectively. Then, the mixture was heated with the buffer to 90 °C for 10 min, then specific wells were loaded with the heated mixture. Standard measures of protein (1.06–26.6 kDa) were also considered. After electrophoresis, fixing, staining and destaining, solutions were prepared and mixed with gel, and the resulted protein bands were compared with the standard ones [29].

2.2.3. Measuring the ACE inhibition activity

We measured the ACE inhibition activity using the Angiotensin-I As Substrate [30], as follows:

- Prepare 20 µl assay buffer in a microplate, at 37 °C for 5 min.
- Initiate the reaction by adding 30 µL of ACE in the assay buffer to 10 µL of sample, and 20 µL of angiotensin-I.
- Incubate at 37 °C for 30 min.
- Replace the inhibitor solution with assay buffer to prepare the blank samples.
- Add 100 µL of 1 M HCl to terminate the reaction.
- Quantify the liberated angiotensin-II.
- Use acetonitrile for 7 min at 25–50%, then for 1 min at 50%, then for 2 min at 25%.
- Quantify angiotensin-II by calculating peaks areas.
- Calculate the degree of ACE inhibition (%) as follows:

$$\text{ACE inhibition}(\%) = [1 - (A_{in}/A_{bl})] \times 100$$

where A_{in} and A_{bl} are peak areas corresponding to angiotensin-II for inhibitor and blank sample, respectively.

- Calculate the IC50 value, which is the concentration of the inhibitor in the sample to reduce the product peak by 50% (here, SigmaPlot 10.0 software was used).

2.3. Random vector functional link network

The random vector functional link (RVFL) was introduced by Pao et al. [14]. This method combines the characteristics of the functional link [23] with random weights. The RVFL is single layer feedforward neural network with a direct link between the input and output neurons as in Fig. 1. This link is preventing RVFL from the overfitting problem.

In the RVFL, the weights between the input layer and enhancement

nodes are initialized randomly in the interval $[-S, +S]$, where S is a scale factor to be determined during the parameter tuning stage for each dataset. These parameters are fixed during the training phase.

Only the output weights are updated during the training phase. They are computed using the ridge regression and Moore-Penrose pseudo-inverse (see below).

Consider the input neurons of RVFL receiving a set of labeled data $\{(u_i, y_i) | x_i \in R^n, y_i \in R^m, i = 1, \dots, N\}$, then the output of j th enhancement node is defined as

$$G_j(a_j x_i + b_j) = \frac{1}{1 + e^{-(a_j u_i + b_j)}} \quad (1)$$

Where $a_j \in [-S, S]$ and $b_j \in [0, S]$ represent the weight and bias between the input layer and enhancement nodes, respectively.

Then, the output of the RVFL can be defined as

$$Y = Bw \quad (2)$$

Where $w \in R^{n+p}$ represents the output weight and B is a matrix that contains the input data, and the output of the enhancement node G is defined as follows:

$$B = [B_1 B_2]$$

$$B_1 = \begin{bmatrix} u_{11} & \dots & u_{1n} \\ \vdots & \ddots & \vdots \\ u_{N1} & \dots & u_{Nn} \end{bmatrix}, B_2 = \begin{bmatrix} G_1(a_1 u_1 + b_1) & \dots & G_p(a_p u_1 + b_p) \\ \vdots & \ddots & \vdots \\ G_1(a_1 u_N + b_1) & \dots & G_p(a_p u_N + b_p) \end{bmatrix}$$

The weight w can be computed by using the Moore-Penrose pseudo-inverse [18,31] as

$$w = B^\dagger Y \quad (4)$$

or the ridge regression [25], as follows:

$$w = \left(B^T B + \frac{I}{C} \right)^{-1} B^T Y \quad (5)$$

where \dagger, C and I represents the Moore-Penrose pseudo-inverse, the

tradeoff parameter and the identity matrix, respectively.

2.4. Sine cosine algorithm

The Sine Cosine Algorithm (SCA) is a metaheuristic technique used to find the optimal solution for global optimization problems [26]. The main operators of SCA are the sine and cosine functions, which are used to update the solutions.

The SCA generates a random set of N solutions (population) X with dimension N_{par} as in the following equation [26]:

$$x_{ij} = l_j + \gamma(u_j - l_j) \quad (6)$$

Where γ, u_j and l_j represent a random number, the upper boundary and lower boundary, respectively, for each element of x_i .

The SCA updates the solution x_i using either the sine or cosine function based on the probability parameter $\delta_1 \in [0, 1]$ as in the following equation:

$$x_i = \begin{cases} x_i + 2\delta_2 \sin(\delta_3) |\delta_4 x_b - x_i| & \text{if } \delta_1 \geq 0.5 \\ x_i + 2\delta_2 \sin(\delta_3) |\delta_4 x_b - x_i| & \text{otherwise} \end{cases} \quad (7)$$

where x_b is the best solution (configuration), and δ_2, δ_3 and δ_4 are random numbers, and where δ_2 is used to find the suitable region for the next solution. It is updated during the search process, as follows:

$$\delta_2 = a - a \frac{t}{t_{max}} \quad (8)$$

where a is a constant value, t and t_{max} represent the current iteration and the maximum number of iterations, respectively.

In Equation (7), δ_3 is used to find the movement direction of the current solution (i.e., either away from or towards the best solution), while, δ_4 is used to assign a random weight to the best solution to stochastically assert ($\delta_4 > 1$) the effect of desalination in defining the distance. The final steps of the SCA are given in Algorithm 1.

Algorithm 1. The Sine Cosine Algorithm.

Algorithm 1: The Sine Cosine Algorithm.

1. Input: the size of the population N , the dimension of the solution N_{par} , and maximum number of iteration t_{max} .
2. Generate a set of N solutions with dimension N_p .
3. Calculate the fitness function for each $x_i, i = 1, 2, \dots, N$.
4. Find the best solution x_b according to the best fitness function fit_b .
5. Assign $t = 1$.
6. While $t \leq t_{max}$
7. Update the value of $\delta_1, \delta_2, \delta_3$ and δ_4 .
8. Update each solution x_i using Eq. (7) and compute fit_i .
9. Update x_b .
10. $t = t + 1$
11. End While.
12. Return x_b .

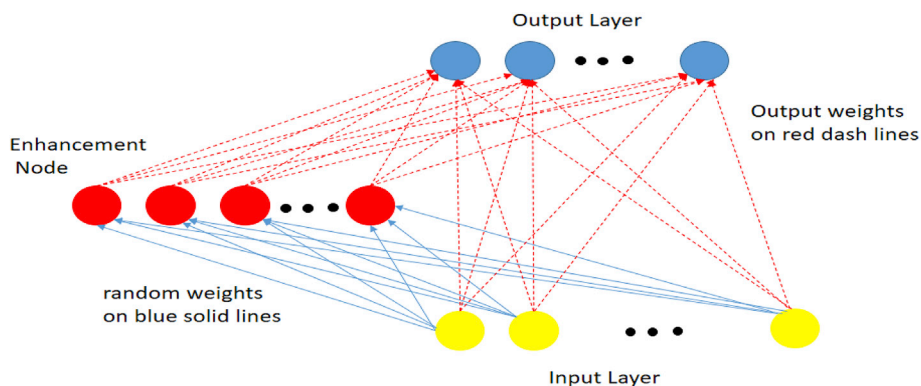


Fig. 1. The structure of RVFL network.

3. Proposed prediction method

In this section, we present the proposed method for ACE inhibition activity prediction. It based on the SCA algorithm with the RVFL network.

In general, the configuration of the RVFL network depends on several parameters and the process of selecting the optimal parameters is challenging. Therefore, the SCA is used to select the optimal configuration of these parameters. The SCA starts by building the population X that contains a set of N_{con} solutions, representing different configurations. The dimension of the solution x_1 is the number of parameters (N_{par}) used to build the suitable RVFL network.

In this study, the main parameters of RVFL are the activation function, bias, and the number of hidden neurons. For example, assume $x_i = [x_{i1}, x_{i2}, x_{i3}] = [100, 1, 3]$. First element in x_i represents the number of hidden neurons. Second element specifies if there is a bias at the output layer (it takes values 1 or 0). Third element is a categorical variable, which determines the objective function used during the training of the RVFL. Here, we used six functions, namely 1- triangular basis, 2- sign, 3- radial basis, 4- sine, 5- hard limit, and 6- sigmoid. Therefore, based on the current solution the network of RVFL is constructed using 100 hidden neurons, with bias at the output, and using the triangular basis as activation function.

The next step is to compute the fitness function for each solution and determine the solution that achieves the smallest error value. Then, the solutions are updated using the sine and cosine functions. The steps of the proposed model are described in the following with more details. The population $X \in R^{N_{con} \times N_{par}}$ is generated using the following formula:

$$x_{ij} = \text{floor}(l_j + r \text{ and } \times (u_j - l_j)), i = 1, \dots, N_{con}, j = 1, \dots, N_{par} \quad (9)$$

Where *floor* is function used to round a number to the next smaller integer. the u_j and l_j represent the upper and lower boundary of the j th parameter. Then the dataset is split into training and testing sets. Then, the RVFL is constructed based on current solution, trained using the training set, and evaluated using the testing set to predict the ACE inhibition activity. After that, the fitness function is computed overall the training and testing set using the root mean square error (RMSE) as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y - \hat{y})^2} \quad (10)$$

Then, the solutions are updated based on the sine and cosine functions as well as the best solution x_b . The solution is updated until the terminal conditions are met. The proposed SCA-RVFL is fully illustrated (Fig. 2).

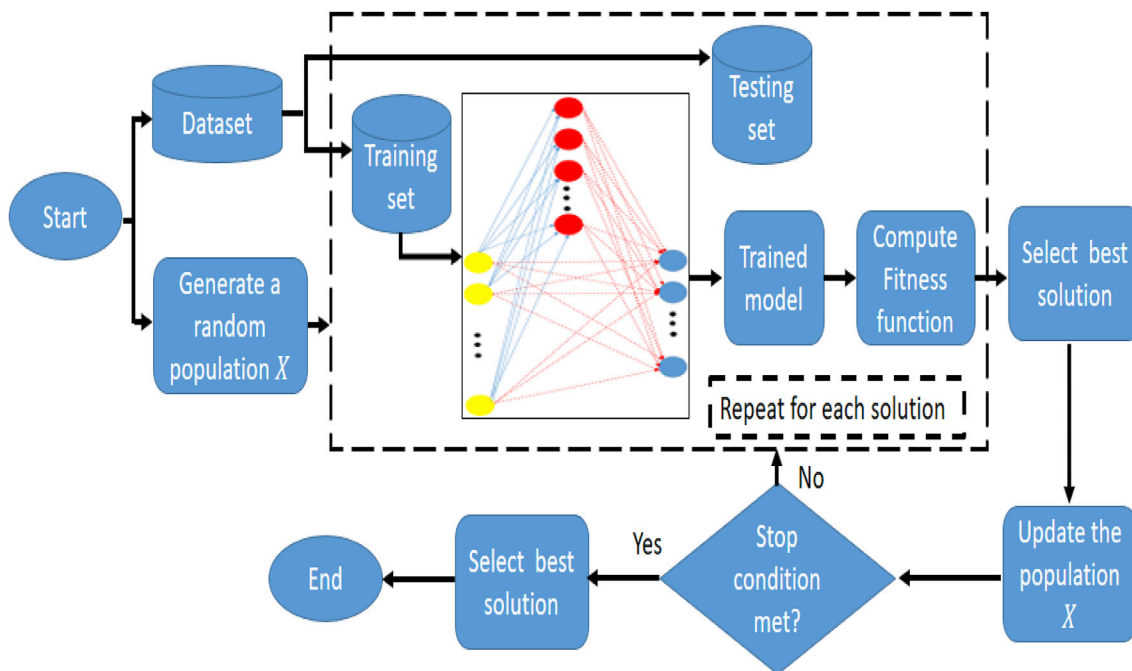


Fig. 2. The proposed SCA-RVFL method.

4. Experimental results and discussion

Our goal is to utilize the advances in machine learning to predict the effect of chemical alteration on ACE inhibition activity at different times and different localities. We investigated the effect of chemical alteration resulted from different process condition, concentration of bioactive amino acids in crude fish, then in hydrolysates using the high hydrolysis efficacy of alcalase enzyme. Although our study showed the efficacy of alcalase enzyme in producing peptides with low molecular weight, there is a great variation in ACE inhibition activity even under controlled experimental parameters.

To the best of our knowledge, we report for the first time a significant effect of chemical structure alteration on the fish proximate compositions, peptide bioactivity, and ACE inhibition activity. The highest ACE inhibition activity is at 25–27 °C, suggesting optimal chemical structure. With this temperature range, the biosynthesis of some amino acids including but not limited to tyrosine, phenylalanine, and tryptophan has the best influence on the ACE binding. Likewise, the amounts of aspartic and glutamic acids were proposed to be increased in that range. This increase of negatively charged amino acids is vital for ACE inhibition activity [4].

Higher and lower temperatures lead to a subsequent decrease in the ACE inhibition activity. Under these conditions, the biosynthesis of the positively charged and aromatic amino acids is dominant.

ACE inhibition activity varied between different samples as it is affected directly by chemical composition and structure [6]. Noteworthy, recent studies support our results and reported that temperature has the greatest effect on protein metabolism and gene expression of amino acids in fish [1,31].

Under the controlled experimental conditions and temperature fluctuations, the ACE inhibition activities can be predicted with high performance by SCA-RVFL approach.

4.1. Performance measures

To assess the quality of the predictive value, a set of performance measures were used, as follows:

Table 1

The parameters of each method and their values.

Method	Parameters and their values
RVFL	Activation function is set to radbas, bias is set to 1, link between the input and output is set to 1. Updating method is ridge regression. Number of hidden neurons is 200.
Adaptive neuro-fuzzy inference system (ANFIS)	Max. epoch = 100, Initial step = 0.01, Decrease rate = 0.9, Increase rate = 1.1
Sequential minimal optimization (SMO)	C = 9
Support Vector Machine (SVM)	C = 9 and Gamma = 7,
Instance Based Learner (IBK) (type of K-nearest neighbors)	number of nearest neighbors (K) = 5
SCA-RVFL (proposed)	a = 2, N = 10, maximum number of iterations is 20.

All the algorithms were implemented in Matlab 2018b in Windows 10 64-bit environment using a PC with 4 GB RAM and an Intel® Core™ i3-3110 M Processor.

Table 2

The Comparison Results of the proposed method and other methods for the testing set.

	RVFL	ANFIS	SMO	SVM	IBK	Proposed
MAPRE	1.213191	4.020795	8.312365	9.361185	7.749365	0.520946
RMSE	0.854426	2.903404	5.9346	6.680784	5.509736	0.388287
MRE	-0.00967	-0.04021	-0.08312	-0.09357	-0.07749	-0.00252
MAE	0.794377	2.618746	5.415131	6.098775	5.053367	0.341714

a) Mean absolute error (MAE):

$$MAE = \frac{1}{n_s} \sum_{i=1}^{n_s} |y_i - \tilde{y}_i| \tag{11}$$

b) The coefficient of determination (R^2) measures the goodness of fit between the results obtained by a method and the actual data. If the value of R^2 is closer to 1, the correlation between the experimental and predicted values is better. R^2 is defined by:

$$R^2 = 1 - \frac{\sum_{i=1}^{n_s} (y_i - \tilde{y}_i)^2}{\sum_{i=1}^{n_s} (y_i - \bar{y})^2} \tag{12}$$

c) mean relative error (MRE):

$$MRE = \frac{1}{n_s} \sum_{i=1}^{n_s} \frac{y_i - \tilde{y}_i}{y_i} \tag{13}$$

d) Mean absolute percent relative error (MAPRE):

$$MAPRE = \frac{100}{n_s} \sum_{i=1}^{n_s} \left| \frac{y_i - \tilde{y}_i}{y_i} \right| \tag{14}$$

where n_s represents the number of observations, y are the observed and \tilde{y} are the predicted values, respectively. In addition, the RMSE which defined in Eq. (10) is used.

Table 1 summarizes the methods we compared against SCA-RVFL and the parameter setting for each method used in this study.

4.2. Results: comparison with other methods

In this experimental series, the dataset is randomly split into training set with size 90% and testing set with 10%. The comparison of predictions of ACE inhibition activity on the testing set is illustrated in Table 2 and Figs. 3–8. Our proposed method shows the best performance among the compared approaches. The RMSE value achieved by our approach is the lowest, indicating that the predicted ACE inhibition activity value is the closest to the original among compared solutions. Notably, the RVFL and the ANFIS methods are second and third, respectively, in terms of their prediction RMSE. Similarly, when looking at MAPRE, MRE, and MAE values, our method has the highest scores, followed by the RVFL and the ANFIS.

Figs. 3–8 depict the overall the ACE inhibition activity values (training and testing) and it can be noticed that the performance of the proposed model is better than the other methods which obtained RMSE 0.12. Noteworthy, RVFL and ANFIS allocate the second and third rank with RMSE value 0.45 and 0.99 respectively. Likewise, SVM, SMO, and IBK obtained 2.13, 1.89, and 1.93, respectively. The histogram analysis of the error for the proposed method shows that it is closer to the normal distribution than others. Interestingly, the methods performing the best for the training set are SMO, SVM, and the proposed method as RMSE is

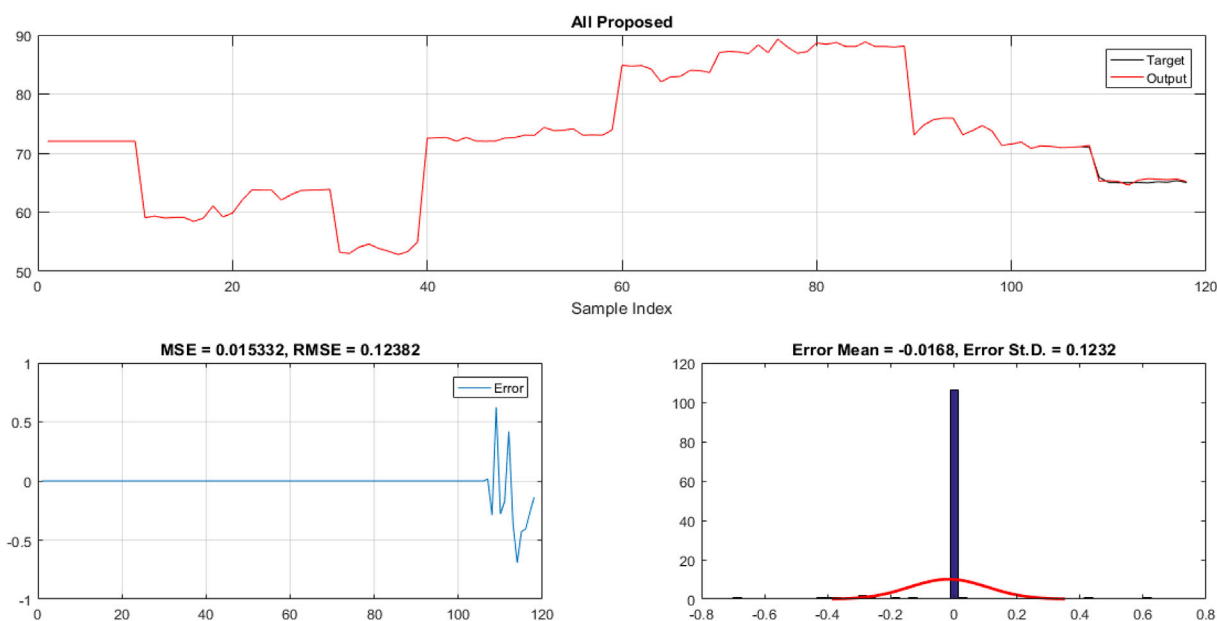


Fig. 3. The results of ACE inhibition activity based on SCA-RVFL model.

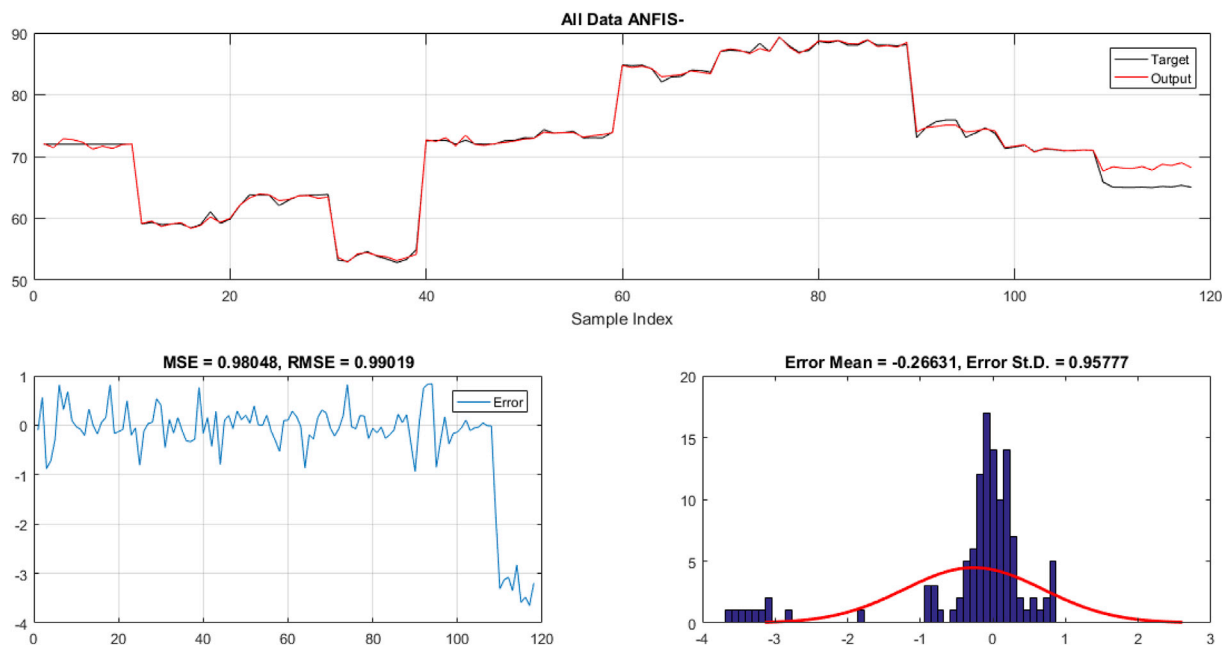


Fig. 4. The results of ACE inhibition activity based on ANFIS model.

nearly equal to zero. However, the error range of the RVFL and ANFIS is [-1, 1]. In terms of training set error analysis, the proposed method has the smallest error over all the other methods, followed by the RVFL method that has error interval [-0.5,0.5]. While, the other methods the error varies from -3 to -6.

Moreover, it can be concluded from Table 2 that the proposed model has better performance in terms of MAE, MRE, MAPRE, and RMSE using the testing set.

4.3. Results: influence of size of training set

In this section, the performance of the proposed approach, RVFL and ANFIS is evaluated for varying proportions of the training-testing set: 70%–30% and 80%–20% of the dataset. These experiments were carried

out 25 times and the average and standard deviation (STD) were calculated for each measure (discussed in Section 4.1).

The comparison results are illustrated in Table 3. One can notice that the performance of the proposed approach is better than other methods in both cases. In addition, the average value of the RMSE and R^2 for the RVFL is lower than the ANFIS for both cases. In terms of STD, one can see that the methods have the same performance. However, the proposed method has the smallest values overall except the MAE for the RVFL.

Since we demonstrated that the proposed method offers better performance than the RVFL and ANFIS, it will be used as reference to compute the relative improvement as follows:

$$RE = \frac{M_A - M_{Pro}}{M_{Pro}} \tag{15}$$

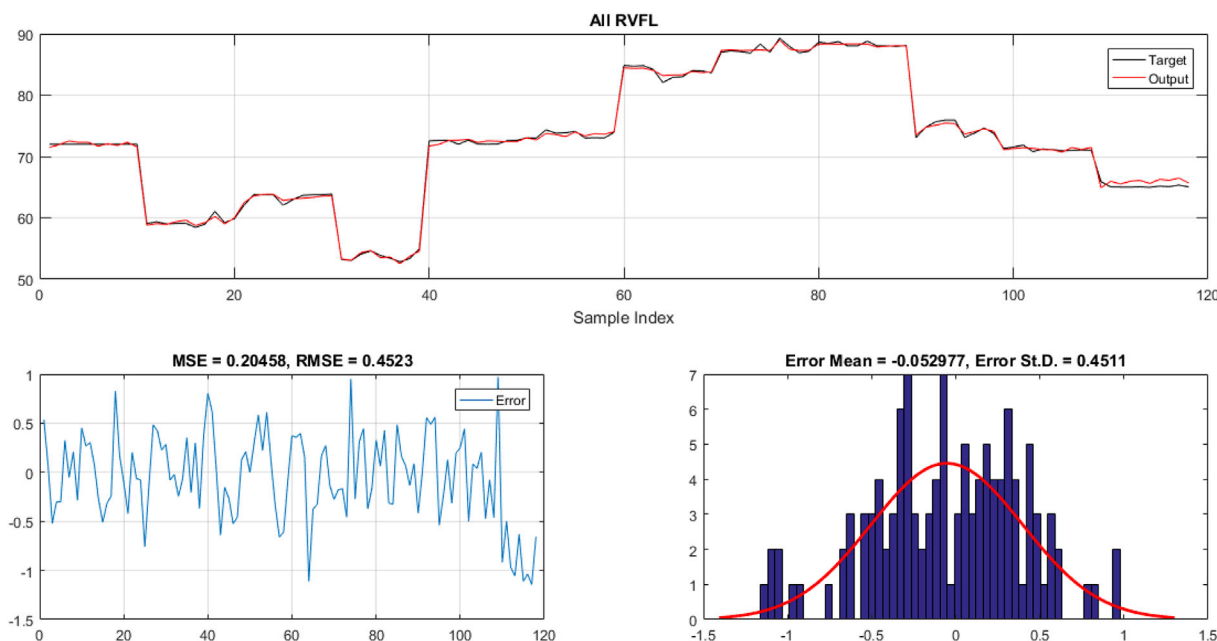


Fig. 5. The results of ACE inhibition activity based on RVFL model.

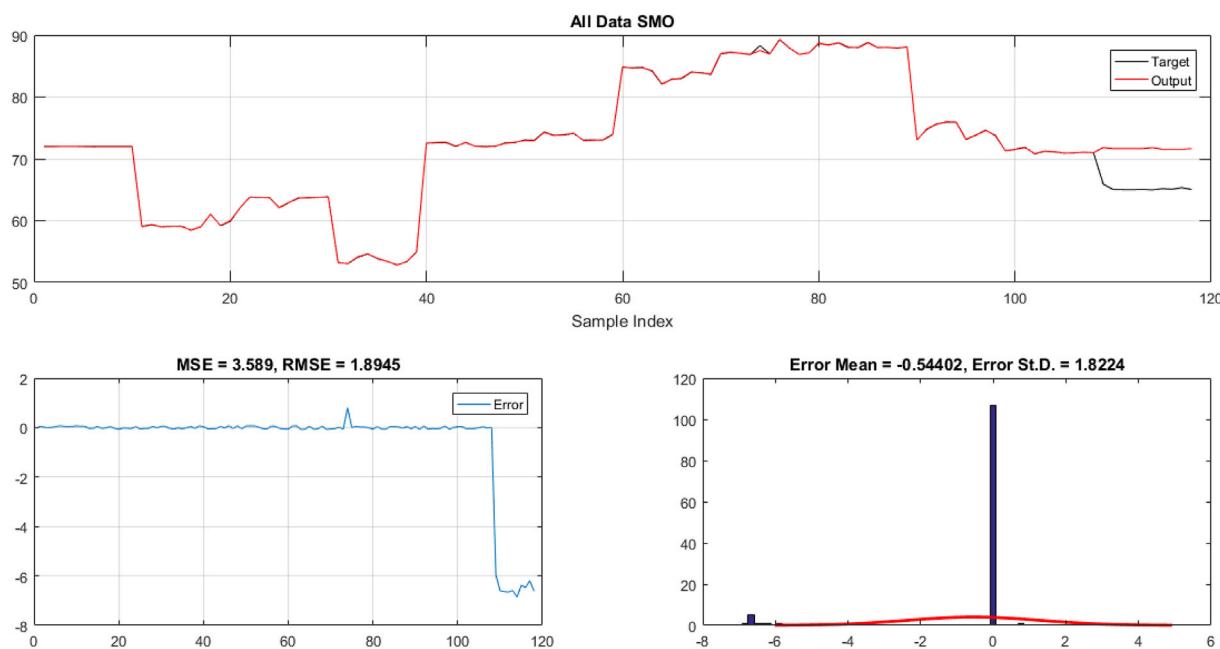


Fig. 6. The results of ACE inhibition activity based on SMO model.

where M represents the measure value (i.e., MAPRE, RMSE, MRE, MAE, and R^2) of the proposed approach (M_{pro}), and ANFIS or RVFL (M_A). Table 4 shows the value of the improvements in which it can be noticed that only R^2 yields negative values, which indicates the high performance of the proposed method for the considered measures.

According to the above results, it can be concluded that the SCA-RVFL approach is superior to other approaches in predicting the ACE inhibition activity. The proposed approach depends on the RVFL, an established variant of ANN. The main reason for its high performance results from the fact that the RVFL uses a direct link between the input and output neurons. Moreover, using the SCA to determine the optimal parameters of RVFL increases its performance and reduces the computational time to determine the optimal parameters manually.

5. Conclusions and future work

This study addresses the advances in machine learning to unveil the effects of alteration in the chemical structure of fish-derived ACE inhibitors, which is vital to predicting for the drug and food industry. The proposed machine learning method is based on the improvement of the random vector functional link (RVFL) network by using the sine-cosine algorithm (SCA). The aim of using the SCA is to find the optimal configuration of the parameters to enhance the prediction of impact of processing parameters on the resulting performance of ACE inhibitors. In order to evaluate the proposed approach, a set of experiments was performed. The results provide evidence for the superior predictive performance of the proposed method in comparison with the other approaches.

Based on the obtained results, the proposed SCA-RVFL method may

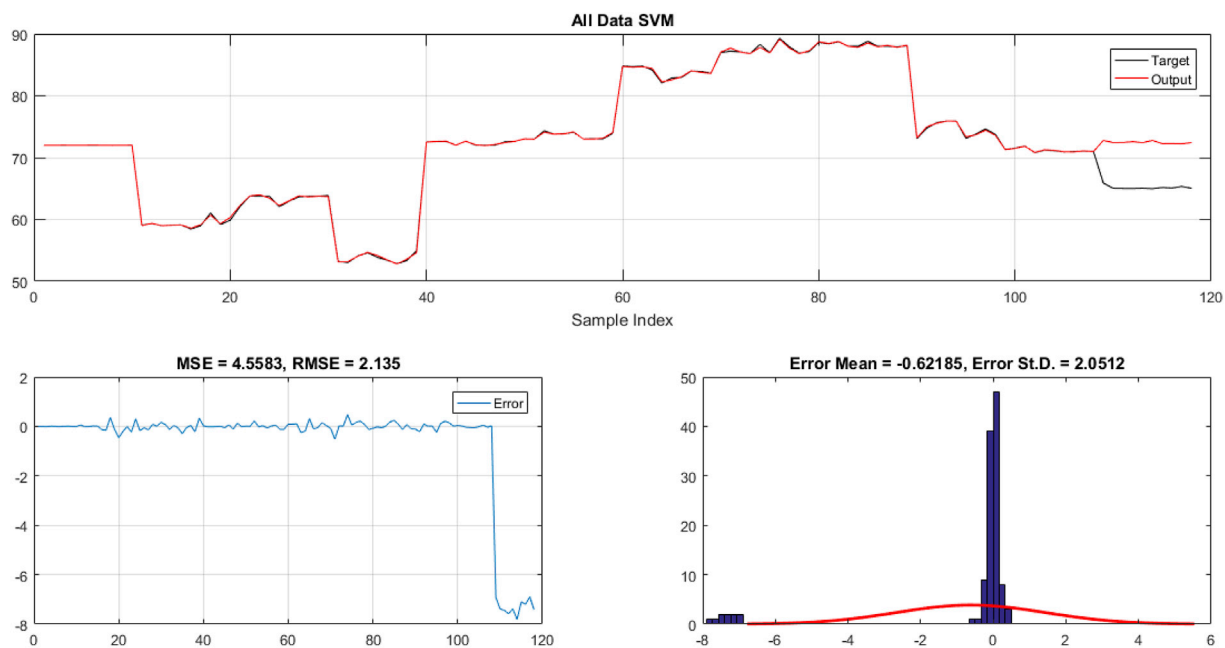


Fig. 7. The results of ACE inhibition activity based on SVM model.

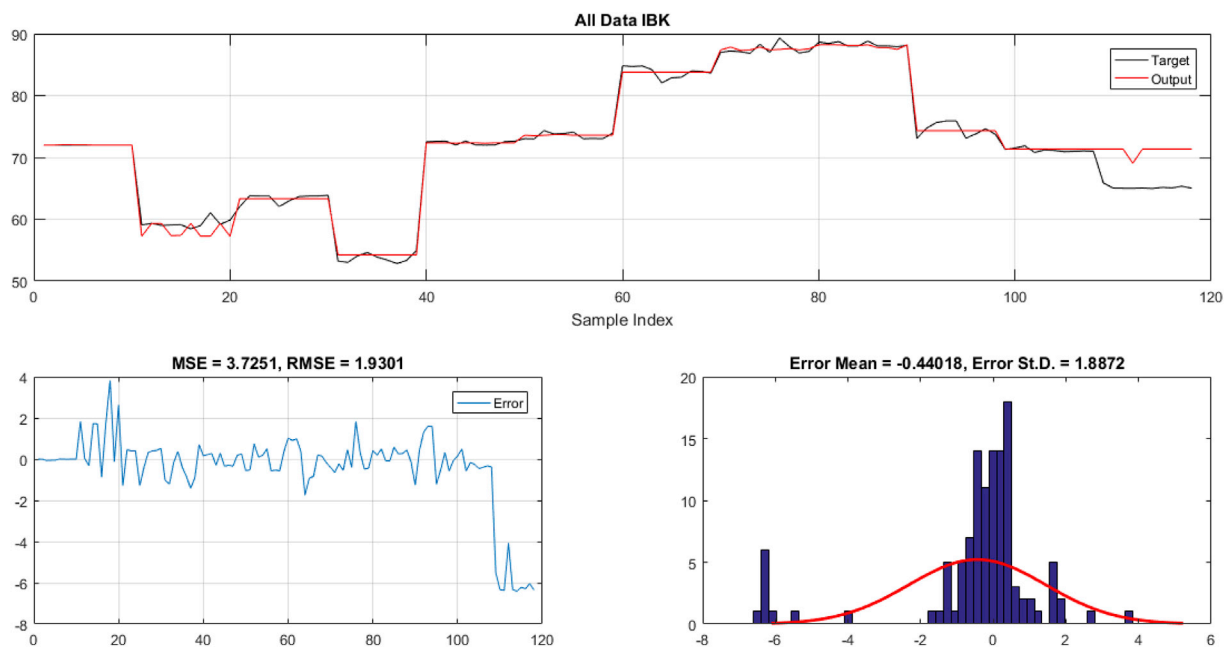


Fig. 8. The results of ACE inhibition activity based on IBK model.

Table 3
Influence of changing the size of testing set.

		80 training/20 testing			70 training/30 testing		
		RVFL	ANFIS	Proposed	RVFL	ANFIS	Proposed
MAPRE	Average	1.68664	4.38526	0.62880	2.26711	4.96841	0.91481
	STD	0.30242	0.25766	0.08013	0.32519	0.19027	0.08932
RMSE	Average	1.34455	3.33163	0.47795	2.17784	6.66220	0.88089
	STD	0.57755	0.39498	0.09427	0.35860	0.60660	0.19776
MRE	Average	-0.01162	-0.06047	-0.00292	-0.05811	-0.10417	-0.02068
	STD	0.00490	0.00266	0.00157	0.00018	0.00150	0.00060
MAE	Average	0.99138	2.97802	0.41971	1.00320	3.15085	0.60968
	STD	0.16609	0.15203	0.12041	0.12909	0.16460	0.07955
	Average	0.96999	0.95627	0.98860	0.95851	0.94934	0.97813
	STD	0.00027	0.00031	0.00013	0.00039	0.00044	0.00013

Bold indicates the best value.

Table 4
Improvement using proposed as reference.

		80 training/20 testing		70 training/30 testing	
		RVFL	ANFIS	RVFL	ANFIS
MAPRE	Average	1.68	5.97	1.48	4.43
	STD	2.77	2.22	2.64	1.13
RMSE	Average	1.81	5.97	1.47	4.29
	STD	5.13	3.19	0.81	2.07
MRE	Average	2.98	19.71	1.81	4.04
	STD	2.11	0.69	−0.70	1.51
MAE	Average	1.36	6.10	0.65	4.17
	STD	0.38	0.26	0.62	1.07
R ²	Average	−0.02	−0.03	−0.02	−0.03
	STD	1.15	1.44	2.02	2.48

be applicable to other problems, including prediction of climate change effects, challenges data mining or bioinformatics.

Conflicts of interest

The authors declare no competing interests.

Acknowledgements

This work is supported by the Science and Technology Program of Shenzhen of China under Grant Nos. JCYJ20180306124612893, JCYJ20170818160208570 and JCYJ20170307160458368.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chemolab.2019.05.009>.

References

- [1] M.A. Elaziz, A.M. Hemdan, A. Hassani, D. Oliva, S. Xiong, Analysis of bioactive amino acids from fish hydrolysates with a new bioinformatic intelligent system Approach, *Sci. Rep.* 7 (2017) 10860.
- [2] U.G. Yathisha, Ishani Bhat, Iddya Karunasagar, B.S. Mamatha, Antihypertensive activity of fish protein hydrolysates and its peptides, *Crit. Rev. Food Sci. Nutr.* (2018), <https://doi.org/10.1080/10408398.2018.1452182>.
- [3] C. Daskaya-Dikmen, A. Yucetepe, F. Karbancioglu-Guler, H. Daskaya, B. Ozelik, Angiotensin-I-Converting enzyme (ACE)-Inhibitory peptides from plants, *Nutrients* 9 (2017) 316.
- [4] R.E. Aluko, Structure and function of plant protein-derived antihypertensive peptides, *Current Opinion in Food Science* 4 (2015) 44–50.
- [5] Y. Gu, Wu, J. LC-MS/MS coupled with QSAR modeling in characterising of angiotensin I-ConvertingEnzyme inhibitory peptides from soybean proteins, *Food Chem.* 141 (2013) 2682–2690.
- [6] D. Ceren, Y. Aysun, K. Funda, D. Hayrettin, O. Beraat, Angiotensin-I-Converting enzyme (ACE)-Inhibitory peptides from plants, *Nutrients* 9 (2017) 316.
- [7] I. Sagardia, R.H. Roa-Ureta, C. Bald, A new QSAR model, for angiotensin I-converting enzyme inhibitory oligopeptides, *Food Chem.* 136 (2013) 1370–1376.
- [8] H.S. Hippert, C.E. Pedreira, R.C. Souza, Neural networks for short-term load forecasting: a review and evaluation, *IEEE Trans. Power Syst.* 16 (2001) 44–55.
- [9] D.C. Park, A time series data prediction scheme using bilinear recurrent neural network, in: *Proceedings of the IEEE International Conference on Information Science and Applications (ICISA2010)*, Seoul, 2010, pp. 1–7.
- [10] E.E. Elattar, J. Goulermas, Q.H. Wu, Electric load forecasting based on locally weighted support vector regression, *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* 40 (2010) 438–447.
- [11] Y.C. Guo, D. Xiao Niu, Y.X. Chen, Support vector machine model in electricity load forecasting, in: *Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC2006)*, 2006, pp. 2892–2896.
- [12] C.L.P. Chen, A rapid supervised learning neural network for function interpolation and approximation, *IEEE Trans. Neural Netw.* 7 (1996) 1220–1230.
- [13] C.L.P. Chen, J.Z. Wan, A rapid learning and dynamic stepwise updating algorithm for flat neural networks and the application to time-series prediction, *IEEE Trans. Syst. Man Cybern. B Cybern.* 29 (1999) 62–72.
- [14] Y.H. Pao, G.H. Park, D.J. Sobajic, Learning and generalization characteristics of the random vector functional-link net, *Neurocomputing* 6 (1994) 163–180.
- [15] Y.e. Ren, P.N. Suganthan, N. Srikanth, G. Amaratunga, Random vector functional link network for short-term electricity load demand forecasting, *Inf. Sci.* 367–368 (2016) 1078–1093.
- [16] H.M. Chi, O.K. Ersoy, A statistical self-organizing learning system for remote sensing classification, *IEEE Trans. Geosci. Remote Sens.* 43 (2005) 1890–1900.
- [17] K.K. Xu, H.X. Li, H.D. Yang, Kernel-based random vector functional-link network for fast learning of spatiotemporal dynamic processes, *IEEE Trans. Syst. Man Cybern. Syst.* 99 (2017) 1–11.
- [18] C.P. Chen, C.Y. Zhang, Data-intensive applications, challenges, techniques and technologies: a survey on big data, *Inf. Sci.* 275 (2014) 314–347.
- [19] C.P. Chen, J.Z. Wan, A rapid learning and dynamic stepwise updating algorithm for flat neural networks and the application to time-series prediction, *IEEE Trans. Syst. Man Cybern. B Cybern.* 29 (1999) 62–72.
- [20] J.C. Patra, R.N. Pal, B. Chatterji, G. Panda, Identification of nonlinear dynamic systems using functional link artificial neural networks, *IEEE Trans. Syst. Man Cybern. B Cybern.* 29 (1999) 254–262.
- [21] S. Scardapane, D. Wang, M. Panella, A. Uncini, Distributed learning for random vector functional-link networks, *Inf. Sci.* 301 (2015) 271–284.
- [22] S. Scardapane, D. Wang, A. Uncini, Bayesian Random Vector Functional-Link Networks for Robust Data Modeling, *IEEE Transaction on cybernetics*, 2017.
- [23] L. Zhang, P.N. Suganthan, A comprehensive evaluation of random vector functional link networks, *Inf. Sci.* 367–368 (2016) 1094–1105.
- [24] B. Igel'nik, Y.H. Pao, Stochastic choice of basis functions in adaptive function approximation and the functional-link net, *IEEE Trans. Neural Netw.* 6 (1995) 1320–1329.
- [25] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [26] S.S.C.A. Mirjalili, A sine cosine algorithm for solving optimization problems, *Knowl. Based Syst.* 96 (2016) 120–133.
- [27] M. Zarei, A. Ebrahimpour, A. Abdul-Hamid, F. Anwar, N. Saari, Production of defatted palm kernel cake protein hydrolysate as a valuable source of natural antioxidants, *Molecular Sciences* 13 (2012) 8097–8111.
- [28] AOAC, *Official Methods of Analysis*, sixteenth ed., 2005. Washington, DC.
- [29] M. Shamloo, J. Bakar, D. Mat Hashim, A. Khatib, Biochemical properties of red Tilapia (*Oreochromis niloticus*) protein hydrolysates, *Food Res.* 1 (2012) 183–188.
- [30] H. Yesmine, L. Anis, A. Ingrid, B. Nicolas, D. Régis, M. Thierry, P. Jean-Marie, S. Frédéric, T. Valérie, B. Stéphanie, Measuring angiotensin-I converting enzyme inhibitory activity by micro plate assays: comparison using marine cryptides and tentative threshold determinations with captopril and losartan, *J. Agric. Food Chem.* 61 (2013) 10685–10690.
- [31] P. Spiros, et al., Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions, *Nat. Commun.* 5 (2014) 4071–4080.