IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 1, 2018, accepted July 31, 2018, date of publication August 10, 2018, date of current version September 5, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2864665

Prediction of Protein Secondary Structure With Clonal Selection Algorithm and Multilayer Perceptron

BURCU ÇARKLI YAVUZ[1], NİLÜFER YURTAY[2], AND OZHAN OZKAN[3]
[1]Department of Information Systems Engineering, Sakarya University, 54050 Sakarya, Turkey
[2]Department of Computer Engineering, Sakarya University, 54050 Sakarya, Turkey
[3]Department of Electrical and Electronics Engineering, Sakarya University, 54050 Sakarya, Turkey

Corresponding author: Burcu ÇarklıYavuz (bcarkli@sakarya.edu.tr)

This work was supported by Research Fund of the Sakarya University under Project 2013-50-02-029.

ABSTRACT The recent studies indicate that the protein secondary structure provides very important advantages in determining the function of a protein, treating numerous diseases and drug design. Determining the secondary structure in the laboratory environment is both costly and challenging. Therefore, the prediction of protein secondary structure has been an important study field of bioinformatics and computational biology for many years. The aim of this paper was to provide a contribution to the prediction of protein secondary structure using the nature-inspired methods. The data in the first phase were trained with clonal selection algorithm (CSA) which was modeled by being inspired by the live immune system. The classification was then performed with multilayer perceptron which is one of the deep learning methods modeled by being inspired by the biological nervous system. The results obtained indicated that training of the data with CSA prior to classification contributed positively to classification success.

INDEX TERMS Clonal selection algorithm, deep learning, hemoglobin protein, multilayer perceptron, prediction of protein secondary structure.

I. INTRODUCTION

Proteins are found in all living systems such as single-celled eukaryotes, plants, fungi, bacteria, animals. Proteins function in all kinds of biological processes in living organisms [1]. The three-dimensional configuration of a protein and specific chemical characteristics of amino acid side chains determine the function of the protein. As a result of this, proteins serve a number of important functions which are important for the intracellular and extracellular vital activities or which provide an evolutionary advantage to the cell and organism [2].

Proteins can be found in primary, secondary, tertiary and quaternary structures. Amino acids constitute the primary structure of a protein by coming together via peptide bonds. The second level in the hierarchical structure of the protein is secondary structure. The most common structural patterns in the secondary structure are α-helices, β-sheets, and coils. Tertiary structure is called the integral conformation of a polypeptide chain, ie. the three-dimensional arrangement of all amino acids. A functional protein is composed of one or more polypeptides constituting the quaternary structure.

Proteins take place in every intracellular process, therefore, they have a vital importance for the organism. Dysfunctional proteins can cause devastating consequences. With aging, mutations and some external factors lead to the misfolding of proteins. Nowadays, the cause of Alzheimer's, Parkinson's and Type 2 diabetes diseases, which are frequently encountered especially in people with advanced age is the misfolded proteins [3].

The determination of misfolded proteins will provide important advantages in the fight against many genetic disorders and diseases. It is of great importance to know the three-dimensional structures of proteins in order to be able to treat genetic disorders and diseases and to be able to develop new treatments.

The information necessary for correct protein folding was proven to exist in the amino acid sequence by the in vitro studies. Under denaturation conditions, deteriorations were observed in the foldings of proteins. Some of the denatured pure polypeptides regained their natural conformation spontaneously when normal conditions were returned. This condition indicates that the information necessary for proteins to

fold properly is present in their primary structure. The protein sequence determines its three-dimensional structure whereas this three-dimensional structure determines the function of the protein [2].

Identification of the three-dimensional structures of proteins is a very costly and challenging process in the laboratory environment because of their very small structures. To make proper identification of proteins three major methods/technologies shall be used: X-ray diffraction of protein crystals, nuclear magnetic resonance, electron crystallography. Even these methods are applied still it is a very hard process to reach the exact identification since every other protein crystallize under different conditions which seems like an impossible process in laboratory environment for some kind of proteins. As a result academicians studying in the field of bioinformatics aim to find new solutions to this problem by making use of computerized computation methods.

The studies on the prediction of protein secondary structure first started in the 1970s. The methods of Chou-Fasman [4] and GOR [5] were founded on the statistical analysis of single residues. The prediction successes at that time remained between 50% and 60%. After the successful adaptation of artificial neural networks in the prediction of protein secondary structure by Qian & Sejnowski [6], machine learning algorithms were preferred more often and an increase was observed in prediction successes [7]. Artificial neural networks [8], [9], Hidden Markov Model [10], [11], support vector machines [12], [13], bee colony [14] are the frequently used methods. The hybrid methods obtained through the collocation of the methods with high prediction success were observed to be used frequently in the studies conducted recently [15]–[17].

The aim of this study was a more successful prediction of protein secondary structure by combining the advantageous aspects of CSA and deep learning methods used in the prediction of protein secondary structure. In this context, the data were improved with the CSA modeled by being inspired by the immune system in the first phase whereas classification was carried out with the deep learning methods modeled by being inspired by, one of the most successful methods in the classification problem solving, the neural networks in the second phase.

There are not many studies in which artificial immunity algorithms were applied to the problem of protein structure prediction. Cutello *et al.* [18], [19] performed protein structure prediction with artificial immunity algorithms in the lattice models. In parallel with the increasing popularity of deep learning in recent years, their number of applications in the prediction of protein secondary structure has also increased day by day. Deep learning networks [20], deep convolutional neural fields [21], deep recurrent encoder-decoder networks [22], the combination of convolutional and supervised generative stochastic networks [23] have been the main deep learning methods used in protein structure prediction.

This article was organized as follows: The data and methods used in the study were introduced in the second part. The experimental results were shared in the third part. Finally, the results of the study were discussed in the fourth part.

II. DATA AND METHOD

In this study, a 3336 amino acids long data set composed of 22 different hemoglobin proteins was used. The data were derived from the Protein Data Bank (PDB).

The primary structure of the proteins originated from the sequences of 20 different amino acids in different orders and lengths. Each of these 20 amino acids is represented by a letter (*A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y*).

Different assignment methods are available for assigning the secondary structures of proteins. The Dictionary of Protein Secondary Structure (DSSP), which is the most preferred method in the literature, was used in this study. DSSP describes eight different secondary structures considering the hydrogen bond patterns. These structures are represented by *H, G, I, E, B, T, S*, and *C*. Since the prediction of eight structures is difficult, these eight structures can be reduced to three structures with reduction methods. In this study, the reduction was done by transforming $\{H, G\}$ to $\{H\}$ (helix), $\{E, B\}$ to $\{E\}$ (sheet) and the rest to $\{C\}$ (coil).

It is not possible to apply the prediction algorithms recommended to the proteins in the form of the amino acid sequence. All protein data must be scanned and transformed into a row-column format with the sliding window method. As a result of the experiments carried out with 11, 13, 15, 17, 19, 21 window sizes, 15 window size was decided to be an appropriate approach for solving the problem [24].

Success ratio of deep learning methods increase with the bigger amount of data. The data trained firstly with MLP. Then the same data improved with CSA using the cloning and mutation processes before MLP classification stage to have the mentioned bigger amount of data above. Gathering the results of both processes, the effect of CSA on the classification success with MLP method is investigated (Fig.1).

A. CSA

The first studies on Artificial Immunity appeared at the end of the' 90s. In the period starting with the studies by Dasgupta [25], De Castro *et al.* [26] and coming up to today, Artificial Immune System (AIS) has been applied in different fields such as computer security [27], optimization [28], swarm robotics [29], disease diagnosis [30] and bankruptcy prediction [31].

CSA, one of the most studied algorithms of AIS, was modeled by being inspired by the principle of biological clonal selection. The aim of the clonal selection principle is to provide the antibody diversity that can fight against the antigens. Whenever a new antigen is encountered, the immune network is updated according to these antigens, thereby increasing the identifiability of the antigens.

Antigens (*Ag*) are the elements of the set of the problems which are expected to be solved whereas antibodies (*Ab*) are the elements of the solution set. Each *Ab* and *Ag* is composed of 20 types of amino acids found in the protein structure.

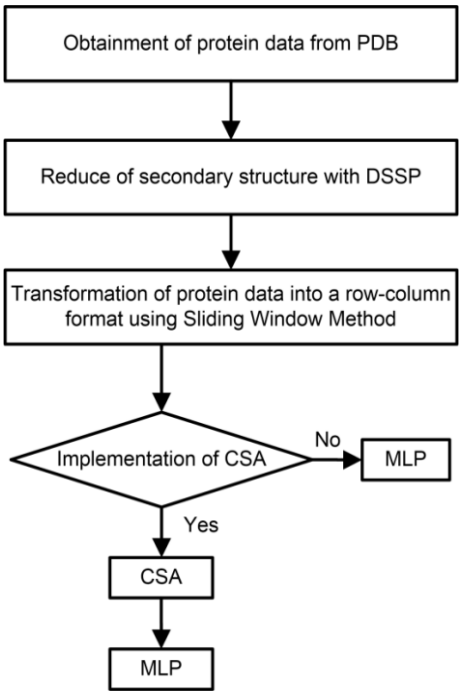

FIGURE 1. Flow chart of the whole study.

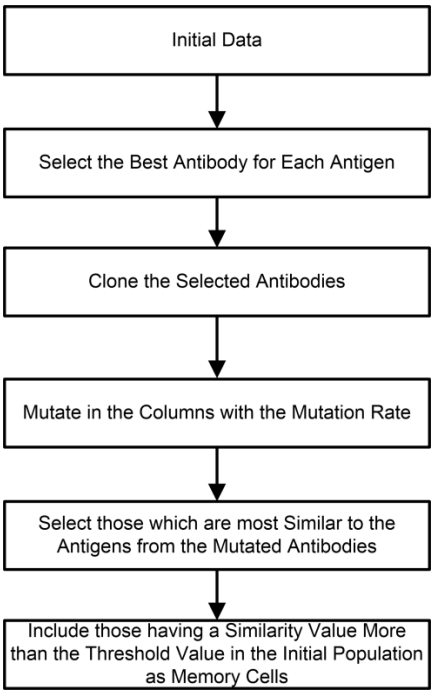

FIGURE 2. Flow chart of CSA.

Each Ab and Ag had a size of lx15 since the size of the window was 15 in this study. On the other hand, the antibody population (P) consists of n antibodies.

$$Ab = \{a_1, a_2, a_3, \ldots a_k\} \quad (1)$$

$$Ag = \{a_1, a_2, a_3, \ldots a_k\} \quad (2)$$

$$P = \{Ab_1, Ab_2, Ab_3, \ldots Ab_n\} \quad (3)$$

The affinity between antibody and antigen is determined by the Hamming distance measure. The motivation behind preferring the Hamming distance measure is to represent the amino acids constituting Ab and Ag with letters.

$$Hamming\,(Ag, Ab) = \sum_{i=1}^{n} \left|Ag_i - Ab_i\right| \quad (4)$$

The process steps of the CSA whose flow chart was given in Figure 2 are as follows [26];

Step 1: Initialization: Create antibody population from the data set.

Step 2: Selection: Select the closest antibody to each antigen using Hamming distance measure.

Step 3: Cloning and mutation: Clone the selected antibodies and randomly mutate them into a set of candidate antibodies.

Step 4: Re-Selection: Compute the similarities of candidate antibodies and antigens. Include the antibodies with similarity over the threshold value in the initial population.

Step 5: If the number of iteration is completed, go to step 6, otherwise go to step 2.

Step 6: Training with CSA was completed. The data are ready for classification phase.

For the training phase with CSA, the software was written with the C# programming language under the framework of NET. MS-SQL Server was used as the database engine.

B. MLP

Although deep learning is quite popular in recent years, actually it is not a new topic. It is possible to solve a problem with increasing complexity via Artificial Neural Networks (ANN) by increasing the number of layers. Although the increasing number of layers refers to a deeper decision mechanism, it also means a more processor power. Technological development in recent years enabled being able to study with deep learning.

MLP is a feedforward artificial neural network model that is trained with stochastic gradient descent (SGD) using backpropagation.

The output of each neuron is defined as:

$$f(\alpha) = f(\sum_{i=1} w_i x_i + b) \quad (5)$$

where x_i and w_i represent the firing neuron's input values and their weights, respectively; the function f represents the nonlinear activation function used throughout the network and the bias b represents the neuron's activation threshold.

As the activation function of MLP tanh, rectifier and maxout were given in Equation 6, 7 and 8 respectively. In this paper, the highest classification success was obtained with the rectifier activation function.

$$\tanh(\alpha) = \frac{e^{\alpha} - e^{-\alpha}}{e^{\alpha} + e^{-\alpha}} \quad (6)$$

$$rectifier(\alpha) = max(0, \alpha) \quad (7)$$

$$maxout(\alpha_1, \alpha_2) = max(\alpha_1, \alpha_2) \quad (8)$$

Three different layers exist in the structure of MLP. At the input layer, incoming data are transmitted to the intermediate layer. The hidden layer can be one or more than one in number. The data received from the input layer are processed in this layer. In the output layer, output values are obtained according to the information received from the intermediate layer. The processes are executed in the network step by step from the input layer to the output layer. The structure of MLP is shown in Figure 3.

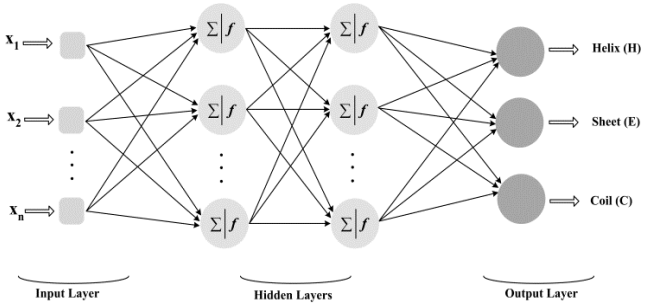

FIGURE 3. MLP structure recommended in the study.

The process of minimizing the loss function $L(W, B|j)$ is a parallelized version of SGD. It has a structure in which many neurons with nonlinear activation function are bound to each other hierarchically. Weights and biases are updated in each epoch by Equation 9 and Equation 10 where α is a constant parameter called the learning rate.

$$w_{jk} := w_{jk} - \alpha \frac{\partial L(W, B|j)}{\partial w_{jk}} \tag{9}$$

$$b_{jk} := b_{jk} - \alpha \frac{\partial L(W, B|j)}{\partial b_{jk}} \tag{10}$$

MLP is often used alone or in combination with other methods in the medical area [32]–[35].

The classification and deep learning phase of the study were carried out in the RapidMiner Studio Educational 8.1 platform. RapidMiner executes the deep learning algorithms using open-source software H2O.

While working with supervised learning models in machine learning such as Support Vector Machines (SVM) or the methods like ANN, categorical input data need to be transformed into numerical data. After this transformation, normalization may be required. These pre-treatments may bring the disadvantages of divergence from the original data and extra process. Since deep learning algorithms enable working with categorical data in the RapidMiner platform, there is no need for the digitalization and normalization of the protein data. The ability to work with the original data directly without making data transformation provided an ease of application as well as time-saving. These advantages were taken into consideration while selecting the classification algorithm.

III. EXPERIMENTAL RESULTS

A two-phase method was recommended in the study aiming to predict the secondary structure of hemoglobin protein. The protein data were improved with CSA in the first phase whereas the trained data were classified with MLP in the second stage. Because of the low success ratio obtained in iterations which are smaller than 5 iterations, 5, 6, 7 and 8 iterations training were executed with CSA. For each iteration group, data trained with CSA, untrained data, and independent data were classified with MLP separately.

Dropout is an effective regularization technique to avoid overfitting in the neural networks. It reduces the memorization of the data by neglecting 50% of neurons in each hidden layer from the network for a given training sample.

Ten-fold cross validation technique was applied in the classification phase.

The results were assessed according to the Q_3 success criteria, the percentage total number of residues correctly for helices (q_H), strands (q_E) and coils (q_C) assigned to the three secondary structure states [36].

$$Q_3 = \frac{q_H + q_E + q_C}{n} * 100 \tag{11}$$

The results were observed by changing the number of hidden layers from 1 to 6 and the number of neurons in each hidden layer from 1 to 200 in this study.

The number of hidden layers each of which is composed of 50 neurons from 1 to 6 and the results obtained for 10, 50, 80, and 200 epochs MLP were given in the Tables 1 to 4.

TABLE 1. Success change based on the number of hidden layers and epoch after 5 iterations with CSA.

Epochs	1 Hidden Layer	2 Hidden Layers	3 Hidden Layers	4 Hidden Layers	5 Hidden Layers	6 Hidden Layers
10	89.96	89.6	89.82	89.85	89.71	89.41
50	91.74	92.18	91.88	91.83	92.02	92.05
80	91.85	91.94	92.53	92.53	92.05	92.01
200	92.12	92.56	92.50	92.18	92.24	91.94

TABLE 2. Success change based on the number of hidden layers and epoch after 6 iterations with CSA.

Epochs	1 Hidden Layer	2 Hidden Layers	3 Hidden Layers	4 Hidden Layers	5 Hidden Layers	6 Hidden Layers
10	90.16	89.92	90.08	90.32	90.18	90.26
50	91.89	91.73	92.3	91.89	92.17	92.12
80	91.79	92.47	92.47	91.97	92.28	92.25
200	92.25	92.38	92.21	92.42	92.09	91.31

As seen in Table 1, when the data iterated 5 times with CSA was classified with MLP, the percentage of classification success occurred between 89.41% and 89.96% for 10 epochs, between 91.74% and 92.18% for 50 epochs, between 91.85% and 92.53% for 80 epochs and between 91.94% and 92.56% for 200 epochs.

TABLE 3. Success change based on the number of hidden layers and epoch after 7 iterations with CSA.

Epochs	1 Hidden Layer	2 Hidden Layers	3 Hidden Layers	4 Hidden Layers	5 Hidden Layers	6 Hidden Layers
10	93.23	93.03	93.17	92.97	92.93	93.08
50	94.45	94.76	94.85	94.69	94.69	94.37
80	94.71	94.74	94.85	94.83	94.96	94.73
200	94.92	95.06	95.06	94.99	95.02	94.94

TABLE 4. Success change based on the number of hidden layers and epoch after 8 iterations with CSA.

Epochs	1 Hidden Layer	2 Hidden Layers	3 Hidden Layers	4 Hidden Layers	5 Hidden Layers	6 Hidden Layers
10	94.81	94.89	94.84	94.76	94.47	94.38
50	96.02	96.21	96.24	96.36	95.96	96.02
80	96.29	96.47	96.44	96.36	96.3	96.15
200	96.36	96.61	96.46	96.49	96.47	96.45

As seen in Table 2, as a result of the classification of the data, which were iterated 6 times with CSA, with MLP, the percentage of classification success was observed to occur between 89.92% and 90.32% for 10 epochs, between 91.73% and 92.30% for 50 epochs, between 91.79% and 92.47% for 80 epochs and between 91.31% and 92.42% for 200 epochs.

When the data iterated 7 times with CSA was classified with MLP, the classification success occurred between 92.23% and 93.17% for 10 epochs, between 94.37% and %94.85 for 50 epochs, between 94.71% and %94.96 for 80 epochs and between 94.92% and 95.06% for 200 epochs as seen in Table 3.

Table 4 shows the percentages of classification success when the data iterated 8 times with CSA was classified with MLP. Accordingly, the success percentage occurred between 94.38% and 94.89% for 10 epochs, between 95.96% and 96.36% for 50 epochs, between 96.15% and 96.47% for 80 epochs and between 96.36% and 96.61% for 200 epochs.

IV. CONCLUSIONS

In this study, the prediction of protein secondary structure was made from the amino acid sequence using MLP, which is one of the deep learning techniques. The classification is examined in two stages: direct MLP and MLP with CSA improvement.

The experimental results obtained indicate that there is a significant increase in the prediction success when the data is improved with CSA before MLP classification. The direct MLP classification success is recorded 84.01%. The MLP with CSA classification success is examined for different number of iterations and hidden layers. It is observed that the MLP with CSA classification success is highest with 2 hidden layers and 200 epochs, 96.61%.

In conclusion applying CSA prior to classification is recommended for a higher prediction success.

ACKNOWLEDGMENT

This article is an extension paper from its conference version [37].

REFERENCES

[1] L. A. Allison, *Fundamental Molecular Biology*, 2nd ed. Hoboken, NJ, USA: Wiley, 2012.

[2] H. F. Lodish, *Molecular Cell Biology*, 7th ed. New York, NY, USA: Freeman, 2013.

[3] E. Reynaud, "Protein misfolding and degenerative diseases," *Nature Educ.*, vol. 3, no. 9, p. 28, 2010.

[4] P. Y. Chou and G. D. Fasman, "Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins," *Biochemistry*, vol. 13, no. 2, pp. 211–222, Jan. 1974.

[5] J. Garnier, D. J. Osguthorpe, and B. Robson, "Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins," *J. Mol. Biol.*, vol. 120, no. 1, pp. 97–120, Mar. 1978.

[6] N. Qian and T. J. Sejnowski, "Predicting the secondary structure of globular proteins using neural network models," *J. Mol. Biol.*, vol. 202, no. 4, pp. 865–884, Aug. 1988.

[7] W. Pirovano and J. Heringa, "Protein secondary structure prediction," in *Data Mining Techniques for the Life Sciences* (Methods in Molecular Biology), vol. 609. Clifton, NJ, USA: Humana Press, 2010, pp. 327–348.

[8] L. H. Holley and M. Karplus, "Protein secondary structure prediction with a neural network," *Proc. Nat. Acad. Sci. USA*, vol. 86, no. 1, pp. 152–156, Jan. 1989.

[9] S. Kaur and B. Pandey, "Protein secondary structure prediction using feed forward artificial neural network and perceptron," *Indian J. Sci. Technol.*, vol. 10, no. 18(1), 2017.

[10] K. Asai, S. Hayamizu, and K. Handa, "Prediction of protein secondary structure by the hidden Markov model," *Comput. Appl. Biosci.*, vol. 9, no. 2, pp. 141–146, Apr. 1993.

[11] J. Martin, J.-F. Gibrat, and F. Rodolphe, "Analysis of an optimal hidden Markov model for secondary structure prediction," *BMC Struct. Biol.*, vol. 6, p. 25, Dec. 2006.

[12] J. J. Ward, L. J. McGuffin, B. F. Buxton, and D. T. Jones, "Secondary structure prediction with support vector machines," *Bioinformatics*, vol. 19, no. 13, pp. 1650–1655, Sep. 2003.

[13] H. Shah, "Protein secondary structure prediction using Support Vector Machines (SVMs)," in *Proc. Int. Conf. Mach. Intell. Res. Adv.*, Dec. 2013, pp. 594–598.

[14] B. Li, Y. Li, and L. Gong, "Protein secondary structure optimization using an improved artificial bee colony algorithm based on AB off-lattice model," *Eng. Appl. Artif. Intell.*, vol. 27, pp. 70–79, Jan. 2014.

[15] X. Zhang, J. P. Mesirov, and D. L. Waltz, "Hybrid system for protein secondary structure prediction," *J. Mol. Biol.*, vol. 225, no. 4, pp. 1049–1063, Jun. 1992.

[16] K.-P. Wu, H.-N. Lin, J.-M. Chang, T.-Y. Sung, and W.-L. Hsu, "HYPROSP: A hybrid protein secondary structure prediction algorithm—A knowledge-based approach," *Nucl. Acids Res.*, vol. 32, no. 17, pp. 5059–5065, Sep. 2004.

[17] H. Hasic, E. Buza, and A. Akagic, "A hybrid method for prediction of protein secondary structure based on multiple artificial neural networks," in *Proc. 40th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2017, pp. 1195–1200.

[18] V. Cutello, G. Nicosia, M. Pavone, and J. Timmis, "An immune algorithm for protein structure prediction on lattice models," *IEEE Trans. Evol. Comput.*, vol. 11, no. 1, pp. 101–117, Feb. 2007.

[19] V. Cutello, G. Morelli, G. Nicosia, M. Pavone, and G. Scollo, "On discrete models and immunological algorithms for protein structure prediction," *Natural Comput.*, vol. 10, no. 1, pp. 91–102, Mar. 2011.

[20] M. Spencer, J. Eickholt, and J. Cheng, "A deep learning network approach to *ab initio* protein secondary structure prediction," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 12, no. 1, pp. 103–112, Jan./Feb. 2015.

[21] S. Wang, J. Peng, J. Ma, and J. Xu, "Protein secondary structure prediction using deep convolutional neural fields," *Sci. Rep.*, vol. 6, Jan. 2016, Art. no. 18962.

[22] Y. Wang, H. Mao, and Z. Yi, "Protein secondary structure prediction by using deep learning method," *Knowl.-Based Syst.*, vol. 118, pp. 115–123, Feb. 2017.

[23] J. Zhou and O. G. Troyanskaya, "Deep supervised and convolutional generative stochastic network for protein secondary structure prediction," in *Proc. 31st Int. Conf. Mach. Learn.*, vol. 32, 2014, pp. 745–753.
[24] B. Ç. Yavuz, C. Sertkaya, and N. Yurtay, "Prediction of secondary structures of hemoglobin using clonal selection algorithm," in *Proc. 7th Int. Work. Comput. Sci. Eng.*, 2017, pp. 1387–1391.
[25] D. DasGupta, "An overview of artificial immune systems and their applications," in *Artificial Immune Systems and Their Applications*. Berlin, Germany: Springer, 1993, pp. 3–21.
[26] L. N. De Castro and F.J. Von Zuben, "Artificial immune systems. Part I—Basic theory and applications," FEEC/UNICAMP, Campinas, Brazil, Tech. Rep. RT DCA 01/99, 1999.
[27] P. K. Harmer, P. D. Williams, G. H. Gunsch, and G. B. Lamont, "An artificial immune system architecture for computer security applications," *IEEE Trans. Evol. Comput.*, vol. 6, no. 3, pp. 252–280, Jun. 2002.
[28] H. S. Bernardino and H. J. C. Barbosa, "Artificial immune systems for optimization," in *Nature-Inspired Algorithms for Optimisation*. Berlin, Germany: Springer, 2009, pp. 389–411.
[29] J. Timmis, A. R. Ismail, J. D. Bjerknes, and A. F. T. Winfield, "An immune-inspired swarm aggregation algorithm for self-healing swarm robotic systems," *Biosystems*, vol. 146, pp. 60–76, Aug. 2016.
[30] C. Liang and L. Peng, "An automated diagnosis system of liver disease using artificial immune and genetic algorithms," *J. Med. Syst.*, vol. 37, no. 2, p. 9932, Apr. 2013.
[31] R. Singh and R. N. Sengupta, "Bankruptcy prediction using artificial immune systems," in *Artificial Immune Systems*. Berlin, Germany: Springer, 2007, pp. 131–141.
[32] F. Ahmad, N. A. M. Isa, Z. Hussain, and M. K. Osman, "Intelligent medical disease diagnosis using improved hybrid genetic algorithm-multilayer perceptron network," *J. Med. Syst.*, vol. 37, no. 2, p. 9934, Apr. 2013.
[33] M. B. Rodrigues *et al.*, "Health of things algorithms for malignancy level classification of lung nodules," *IEEE Access*, vol. 6, pp. 18592–18601, 2018.
[34] O. Ozkan *et al.*, "A study on the effects of sympathetic skin response parameters in diagnosis of fibromyalgia using artificial neural networks," *J. Med. Syst.*, vol. 40, no. 3, p. 54, Mar. 2016.
[35] M. R. Nalluri and D. S. Roy, "Hybrid disease diagnosis using multiobjective optimization with evolutionary parameter optimization," *J. Healthcare Eng.*, vol. 2017, Jul. 2017, Art. no. 5907264. [Online]. Available: https://www.hindawi.com/journals/jhe/2017/5907264/
[36] P. Ghanty, N. R. Pal, and R. K. Mudi, "Prediction of protein secondary structure using probability based features and a hybrid system," *J. Bioinf. Comput. Biol.*, vol. 11, no. 5, p. 1350012, Oct. 2013.
[37] B. Ç. Yavuz, N. Yurtay, and O. Ozkan, "Estimation of secondary structure of hemoglobin protein by multilayer feeding artificial neural aggregates," in *Proc. 3rd Int. Congr. Eng., Archit. Design*, Kocaeli, Turkey, 2018, pp. 65–66.

BURCU ÇARKLI YAVUZ was born in Sakarya, Turkey. She received the bachelor's degree in mathematics-computer from İstanbul Kültür University in 2005, and the M.Sc. degree in computer and information engineering from Sakarya University, Turkey, in 2010, where she is currently pursuing the Ph.D. degree. She is also a Research Assistant with the Faculty of Computer and Information Sciences, Sakarya University. Her research interests include data mining, decision support systems, artificial intelligence, and medical informatics.

NİLÜFER YURTAY was born in Eskişehir, Turkey, in 1969. She has graduated from Faculty of Science Mathematics Department, Anadolu University, in 1992, and received the M.Sc. and Ph.D. degrees from the Department of Industrial Engineering, Sakarya University in 1999 and 2004, respectively. She is currently an Associate Professor Dr. İn with the Department of Computer Engineering, Sakarya University. Her interests are basically in data mining, optimization, and medical decision support systems.

OZHAN OZKAN was born in Sakarya, Turkey. He received the B.Eng., M.Eng., and Ph.D. degrees from Sakarya University, Turkey, in 2003, 2006, and 2012, respectively. He is currently an Assistant Professor with the Department of Electrical and Electronics Engineering, Sakarya University. His research areas include biomedical engineering, electronics, and prediction systems.