International Journal of Innovative
Computing, Information and Control
Volume **8**, Number **4**, April **2012**

 ISSN 1349-4198
pp. **2855–2864**

RECOGNIZING PATHOGENIC ANTIBODIES IN SLE USING GENERAL REGRESSION NEURAL NETWORKS

Mohamad Hasan Bahari[1,4], Asad Azemi[2], Mir Mojtaba Mirsalehi[1]
Mahmoud Mahmoudi[3] and Morteza Khadime[1]

[1]Department of Electrical Engineering
Ferdowsi University of Mashhad
Azadi Square, Mashhad 9177948974, Iran
m.h.bahari@ieee.org; { mirsalehi; khademi }@um.ac.ir

[2]Department of Engineering
Penn State University
Delaware, USA
Azemi@psu.edu

[3]Immunology Research Center
Mashhad University of Medical Science
Mashhad, Iran
mahmoudim@mums.ac.ir

[4]Centre for Processing Speech and Images
Katholieke Universiteit Leuven
Leuven, Belgium

Received November 2010; revised April 2011

Abstract. *In this paper, a new method, based on artificial neural networks (ANN), has been introduced for recognizing pathogenic antibodies in Systemic Lupus Erythmatosus (SLE). dsDNA binding antibodies have been implicated in the pathogenesis of this autoimmune disease. In order to identify these dsDNA binding antibodies, the protein sequences of 42 dsDNA binding and 608 non-dsDNA binding antibodies were extracted from Kabat database and coded using five different physicochemical properties of their amino acids. Coded antibodies were used as the training patterns for five parallel general regression neural networks (GRNNs). Comparing the results obtained by the proposed method with other published results shows the efficacy of proposed approach.*
Keywords: Anti-dsDNA, Antibody, General regression neural network, Systemic lupus erythematosus, Physicochemical properties

1. **Introduction.** Systemic Lupus Erythematosus (SLE or 'lupus') is a major autoimmune rheumatic disease where autoantibodies are frequently targeted against intracellular antigens of the cell nucleus (double and single stranded DNA) [1]. Unfortunately, the cause of lupus is unknown. The Lupus Foundation of America estimates that approximately one million Americans have SLE [2]. Lupus can occur at any age and in either sex, although it is more common in women of childbearing ages [3]. SLE can affect almost any organ or system of the body. In most cases, the disease affects the kidneys, lungs and central nervous system. Infection, lupus flares and cardiovascular disease are cause of most deaths [1,4]. In general, people suffering from SLE have periods of illness and wellness with varying signs. Some have just a few signs of the disease while others have more. This variation in clinical characteristics has made diagnosis of SLE very challenging. According to the result of an experiment on more than five million U.S. Armed Forces personnel, in 115 of the 130 patients with SLE (88 percent), at least one SLE autoantibody tested was

present 3.3 years before the diagnosis [5]. This fact suggests that SLE can be predicted several years before the diagnosis.

In patients with SLE, a wide variety of antibodies against nuclear antigens can be found, including antibodies to nucleic acids, histones and non-histone nuclear proteins. However, our interest is in anti-dsDNA. These antibodies were identified, for the first time, nearly forty years ago in the serum of patients with SLE [6]. Based on similar findings, many researchers have concluded that there is a close relationship between disease activity and levels of anti-dsDNA antibodies. In other words, dsDNA binding antibodies have been implicated in the pathogenesis of this autoimmune disease [7].

Anti-dsDNA antibodies seem highly likely to cause tissue damage in patients with SLE and they are present in the serum of patients several years before their diagnosis. Therefore, their identification and analysis may be useful in the identification of patients who would benefit from early diagnosis, as well as patients who do not require further evaluation. In addition, a powerful method of their analysis may lead us to the design of new drugs that interfere with antibody-DNA interactions, which might have therapeutic applications.

Recently, bioinformatics, including machine learning and statistical techniques, have been successfully applied to various problems about protein structures [8-10], protein families' classification [11], predictions of secondary structure [12], tertiary structure [13], relative solvent accessibility [14] and number of contacts between amino acids [15]. Moreover, techniques using computer-based analysis of autoantibody have received some attention [16-19]. Among those, artificial neural networks (ANNs) have received special attention due to their interesting features including ability to deal with ill-defined and noisy real signals and providing a robust and accurate pattern recognition scheme [20-25,28].

These new approaches, by analyzing a huge number of data, have tried to obtain a better grasp of the DNA-binding structures and utilize them in an early prediction of autoimmune diseases such as SLE. Although these kinds of works have been able to explain many biological phenomenons leading to the development of mathematical models for prediction and reduction in the cost of experimental research, there is still no accurate mathematical method for identifying dsDNA binding antibodies [26]. We can trace the roots of this difficulty to such problems as the long amino acid sequence of antibodies, lack of suitable mathematical tools, and more importantly the vast diversity of antibodies.

In this paper, we extend the computer-assisted autoantibody analysis methods by proposing a new approach with the capability of identifying dsDNA binding antibodies. The proposed identification system is constructed using the General Regression Neural Networks (GRNNs) [27]. The results are compared with analysis using the Multi Layer Perceptron (MLP) neural network and Radial Basis Function (RBF) neural network, instead of GRNN.

In the proposed method, we first encode the amino acid sequence of antibodies using physiochemical properties of amino acids. Next, to obtain better accuracy, each antibody is coded into five discrete signals using five different physicochemical properties of amino acids. Finally, these five signals are fed to five parallel GRNNs. In order to better identify the role of different parts of the anti-dsDNA antibodies, we have investigated the roles of light and heavy chains and their bindings to DNA. The proposed method not only provides an efficient and accurate method to identify the dsDNA binding to antibodies, but also gives insight to the importance of different physiochemical properties in regard to the identification process. This will be shown in the simulation section.

The remainder of this paper is organized as follows. Section 2 includes a brief overview of the coding scheme, which is used in the identification process. Section 3 presents the

proposed method. In Section 4, we compare our results with similar neural networks. The paper ends with conclusions in Section 5.

2. **Coding.** An antibody or immunoglobulin is a large Y-shaped protein used by the immune system to identify and neutralize foreign objects like bacteria and viruses. Each antibody recognizes a specific antigen unique to its target [29]. The basic Y-shaped body of an antibody consists of four polypeptide chains; two identical heavy chains and two identical light chains connected by disulfide bonds [30,31]. Schematic of an antibody is shown in Figure 1.

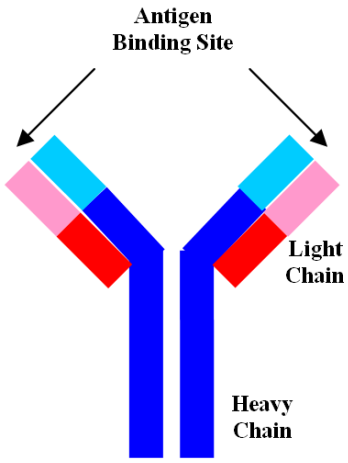

FIGURE 1. Structure of an antibody

Immunoglobulins are heavy plasma proteins, where proteins are complex molecules assembled from amino acids. There are over 300 naturally occurring amino acids on earth, but the number of different amino acids in proteins is only 20 [32-35]. From a one-dimensional point of view, a protein sequence contains characters from these 20 amino acids. In another words, a protein's function depends on its specific amino acid sequence called the primary sequence structure [36-38].

In order to use ANN to identify dsDNA binding antibodies the training set needs to include a set of anti-dsDNA, as the positive set, and a set of other antibodies, as the negative set. Our training data set was constructed using the Kabat database. Kabat antibody sequence database contains primary structure and sequence information on antibodies and other proteins of immunological interest [39]. There are 42 sequences of anti-dsDNA (both heavy and light chains) available in the Kabat database, which we used as positive patterns, and we took 608 of the other antibody sequences (both heavy and light chains) as negative patterns. It is important while there are many antibodies in Kabat database, only 42 anti-dsDNA were available.

An important issue in applying computer-based systems to identify dsDNA binding antibodies is how to encode protein sequences; i.e., how to represent the protein sequences as the input of a system. This is crucial to the success of neural network learning process [40]. Several different methods have been used for encoding proteins. For example, Wu [41] has utilized the 2-gram encoding method, which extracts various patterns of two consecutive amino acid residues in a protein sequence and counts the number of occurrences of the extracted residue pairs. Huang [42] has employed another encoding technique using six physiochemical properties (attributes) of amino acids, namely, composition, predicted secondary structure, hydrophobicity, normalized Van Der Waals volume, polarity, and

polarizability. It should be mentioned that we have used similar coding method as [40], but with very different implementation scheme.

In this work, we have used five different physiochemical properties of amino acids for encoding the protein sequence, resulting in five different coding for each one. The five physiochemical properties namely, isoelectric pH, surface area, hydrophilicity, polarity and the ability of reaction between the ion and the electron (EIIP) were chosen based on the previous studies [43-47].

For simulation, a protein sequence was represented by normalizing the corresponding numbers for each of its amino acids properties. For example, in coding using the first property, the normalized number of isoelectric pH of the amino acid is used instead of each amino acid in the protein sequence. Likewise, in coding using the second property, the normalized number of the surface area of the amino acid is used instead of each amino acid in the protein sequence. Finally, each antibody was transformed into five independent arrays, which are used as the inputs of five parallel GRNNs. An antibody that is coded into an isolated signal, using the hydrophilicity property, is shown in Figure 2. In this figure the horizontal axis shows the number of the amino acids in the protein sequence and the vertical axis shows the normalized amount of hydrophilicity for each amino acid.

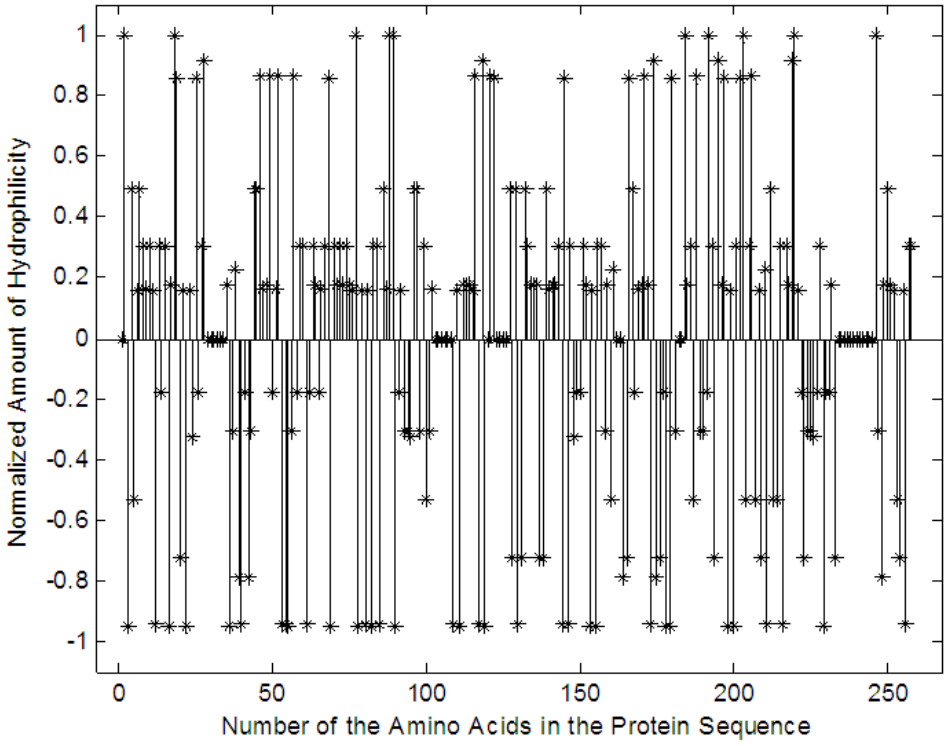

FIGURE 2. A sample of encoded antibody

3. **Simulation.** For the simulation, five parallel GRNNs were used. For training and testing of the proposed system, we have formed a database, from the KABAT database, consisting of 42 anti-dsDNAs, as the positive data set, and 608 other antibodies, as the negative data set.

3.1. **Training phase.** Figure 3 illustrates the general architecture of the system in training phase. The following steps were carried out for the training phase:

1) A training set is formed by randomly choosing 33 positive and 99 negative samples from our database.
2) In the first layer, each antibody is coded into five arrays, based on normalized quantities of the amino acids, physiochemical properties, as illustrated in Figure 3.
3) In the second layer, each neural network is trained separately by one of the arrays obtained from the previous layer so that +1 is assigned as the output for positive samples and −1 is assigned as the output for negative samples.

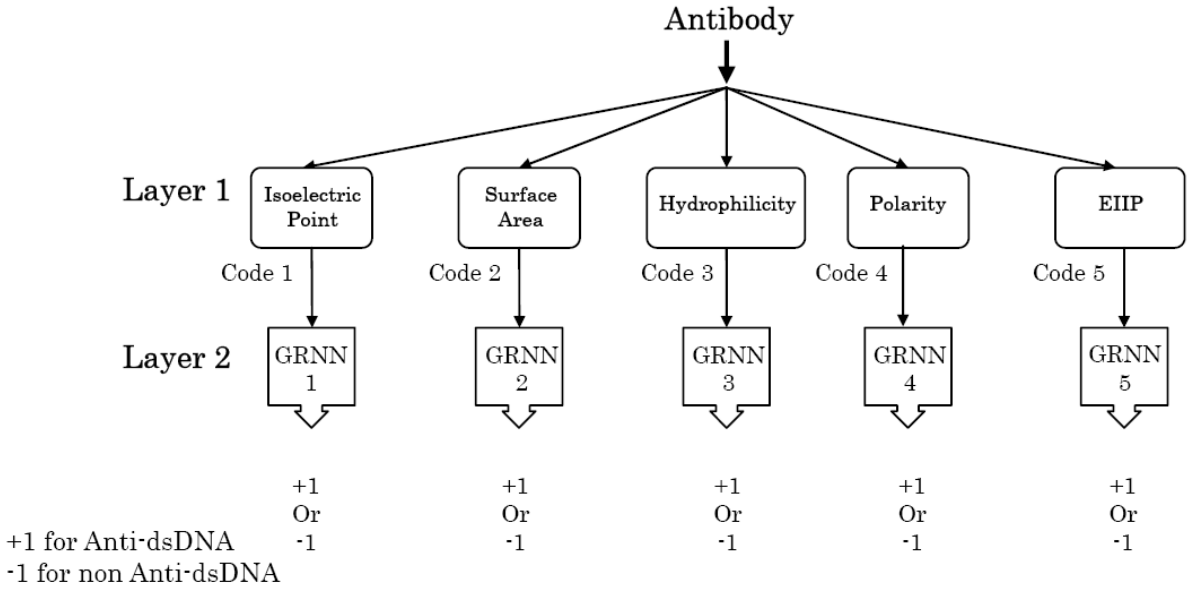

FIGURE 3. Architecture of dsDNA binding antibodies identification system in training phase

3.2. **Testing phase.** Architecture of dsDNA binding antibodies identification system in testing phase is illustrated in Figure 4. The following steps are performed during the testing phase:

1) A testing set is formed by randomly choosing 9 positive and 27 negative samples from the part of our database that was not used in the training set.
2) Like the training phase, each antibody is coded into five discrete signals, based on normalized quantities of the amino acids physiochemical properties, as illustrated in Figure 4.
3) Signals from the previous step are fed separately to one of the five neural networks. The output of each network will be a number between −1 and +1. By considering a threshold number between −1 and +1, we can separate the anti-dsDNA's from other antibodies.

The final decision about an antibody is made by using the output of the third layer. The third layer calculates the mean value of the outputs of each neural network in the second layer, and the binding or not binding decision is made by considering a threshold value in the output. The best value for the threshold, which was found by trial and error, was −0.9.

4. **Results.** In order to assure the validity of the results, the training, and testing operations of the proposed system were repeated 1000 times. The efficacy of the proposed

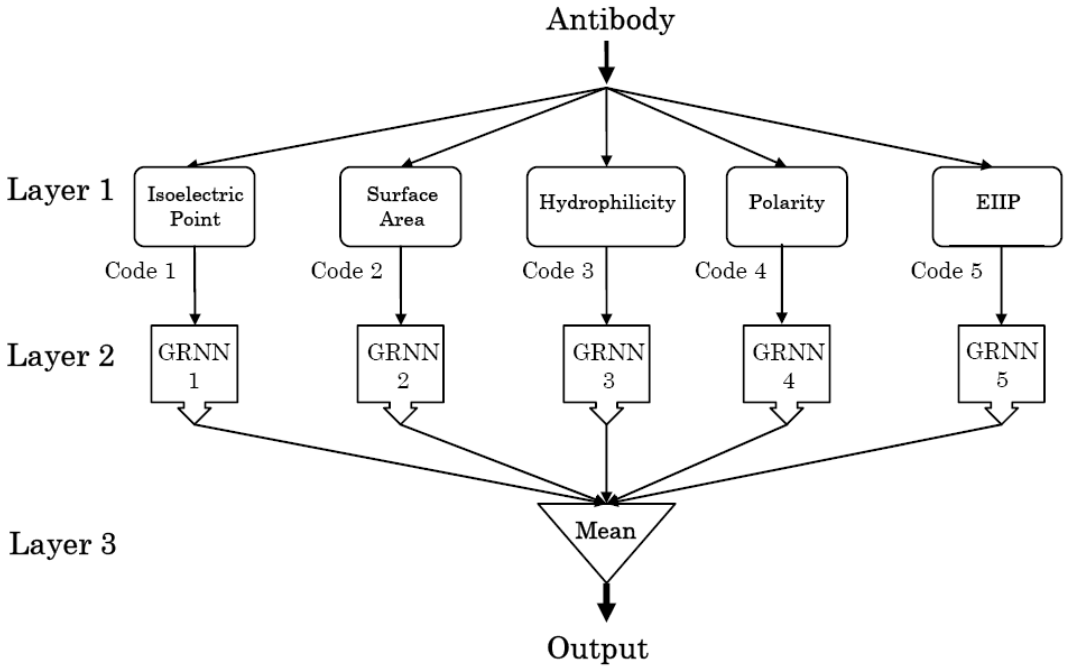

FIGURE 4. Architecture of dsDNA binding antibodies identification system

method is evaluated by calculating the following three parameters: positive accuracy, negative accuracy, and general accuracy.

$$\text{Positive Accuracy} = \frac{TP}{P} \times 100 \tag{1}$$

$$\text{Negative Accuracy} = \frac{TN}{N} \times 100 \tag{2}$$

$$\text{General Accuracy} = \frac{N}{N+P} \times \text{Negative Accuracy} + \frac{P}{N+P} \times \text{Negative Accuracy} \tag{3}$$

where P is the number of positive members of the test set and TP is the number of positive samples which are correctly found. N is the number of negative members of the test set and TN is the number of negative samples that are recognized correctly.

The results obtained from each of the five neural networks, along with the associated physiochemical properties of the antibodies, are shown in Table 1. The results obtained from the proposed method are shown on the last line of this table for comparison. As indicated in Table 1, the general accuracy is higher for the proposed method than the other networks. Moreover, the general accuracy associated with surface area only, has the highest number compared with the other physiochemical properties. This suggests that the amino acids surface area of anti-dsDNA antibodies plays a significant role in determining the binding to DNA.

Results of Table 1 also indicate higher numbers for negative accuracies. This is because we have more negative (non anti-dsDNA) samples and in practice, we are more interested in non anti-dsDNA cases.

For investigating the roles of light and heavy chains of anti-dsDNA antibodies in binding to DNA, an experiment was set up. Heavy chain and light chain of antibodies were extracted from the Kabat database. Next, we repeated the last simulation using heavy and light chains. Table 2 contains these results.

As illustrated in Table 2, the heavy chain of anti-dsDNA is more important than the light chain in binding to DNA. Our simulation results confirm the experimental studies

TABLE 1. Accuracy of each GRNN and the proposed method

	Positive Accuracy	Negative Accuracy	General Accuracy
Isoelectric Point	69.96	93.41	87.55
Surface Area	72.03	95.45	89.60
Hydrophilicity	73.23	89.80	85.66
Polarity	68.43	96.24	89.29
EIIP	73.87	90.78	86.55
Proposed Method	65.79	98.93	90.64

TABLE 2. Associated accuracies using heavy chain, light chain and the proposed method

	Positive Accuracy	Negative Accuracy	General Accuracy
Heavy chain	63.51	98.98	90.11
Light chain	40.81	99.66	84.95
Proposed method (using both heavy and light chains)	65.79	98.93	90.64

[48,49]. Results indicated in Table 2 show that the proposed method, using heavy and light chains, provides a more accurate identification of anti-dsDNA than the heavy chain or light chain alone. The reason is that by considering a more inclusive model, we are including more information and therefore, we should expect better results.

For comparison purposes, we have generated the results presented in Table 2, by using the 2-gram coding method described in [41]. These results are shown in Table 3. A comparison of these two tables indicates the overall superiority of our proposed scheme. We should mention that in this work negative accuracy (non-dsDNA binding antibodies) is more important than positive accuracy. Furthermore, the available data corresponding to the protein sequence for non-dsDNA binding antibodies are much more than their counterparts. Using unequal number of data points for non-dsDNA and dsDNA during training will introduce some bias, but as it was mentioned since the negative accuracy is more important this action is justified. Simulations using equal number of data points for non-dsDNA and dsDNA produced much lower negative and general accuracy.

TABLE 3. Result obtained using the 2-gram coding method

	Positive Accuracy	Negative Accuracy	General Accuracy
Heavy chain	71.78	89.82	85.31
Light chain	71.27	86.76	82.88
Proposed method (using both heavy and light chains)	71.30	90.88	85.99

Next, we replaced the GRNN with the multi-layer perceptron (MLP), radial basis function (RBF) and probabilistic neural networks (PNN), and constructed the proposed parallel neural networks. Table 4 shows the structures used for each of these networks. We should mention that the MLP structure reflects the one that generated the best result for the data set.

The results of using these networks are included in Table 5. We should mention that PNN results, due to its inability of handling many inputs, are not included in the table.

TABLE 4. Structure of the neural networks

	Input Layer	Hidden Layer	Output Layer
MLP	258	10	10
RBF	258	132	1
GRNN	258	132	1

TABLE 5. Result of the four different neural networks used in the proposed architecture

	Positive Accuracy	Negative Accuracy	General Accuracy	Simulation Time (Sec)
GRNN	65.79	98.93	90.64	2432.94
MLP	60.58	99.15	89.51	6526.50
RBF	61.87	87.84	81.35	3350.53

As the results of Table 5 indicate, GRNN is the most suitable neural network to identify dsDNA binding antibodies. Moreover, it takes less time to set up, train, and test as compared with the other considered networks.

In [10], a similar method for recognizing dsDNA binding antibodies is proposed. The method used in [10] is based on a single GRNN and only one physiochemical property is applied. The overall accuracy reported in [10] is 81.83% which is significantly lower than the final accuracy of proposed method in this research (90.64%).

5. **Conclusions.** In this paper, we have introduced a new method for identifying pathogenic antibodies in SLE based on a parallel implementation of GRNN. For identification of dsDNA binding antibodies, the protein sequence of 42 dsDNA binding and 608 non-dsDNA binding antibodies were extracted from the KABAT database. Next, they were coded using five different physiochemical properties of their amino acids. Coded antibodies were used to train five different parallel structured GRNNs. The simulation results indicate that the proposed method is more accurate in recognizing anti-dsDNA antibodies than MLP, RBF, or PNN networks. They also suggest that incorporating the information related to the amino acids surface structure of anti-dsDNA antibodies can improve the prediction process. We are further investigating this finding and the results will be reported in the near future. We have also investigated the roles of light and heavy chains of anti-dsDNA antibodies in binding to DNA. Our simulation results confirmed the experimental findings that the heavy chain is more important than the light chain in regard to binding to DNA.

REFERENCES

[1] R. G. Lahita, *Systemic Lupus Erythematosus*, Academic Press, San Diego, 1992.
[2] *http://www.lupus.org/education/stats.html*, 2006.
[3] F. G. Cuningham, N. F. Gant, K. J. Leveno, L. C. Gilstrap, J. C. Hauth and K. D. Wenstrom, *Williams Obstetrics*, 21st Edition, McGraw-Hill, New York, 2001.
[4] N. B. Blatt and G. D. Glick, Anti-DNA autoantibodies and systemic lupus erythematosus: A retrospective analysis of 17 unselected cases, *Pharmacology Therapeutics*, vol.83, no.2, pp.125-139, 1999.
[5] M. R. Arbuckle, M. T. McClain, M. V. Rubertone, R. H. Scofield, G. J. Dennis, J. A. James and J. B. Harley, Development of autoantibodies before the clinical onset of systemic lupus erythematosus, *The New England Journal of Medicine*, vol.349, no.16, pp.1526-1533, 2003.
[6] D. A. Isenberg, C. T. Ravirajan, A. Rahman and J. Kalsi, The role of antibodies to DNA in systemic lupus erythematosus, *Lupus*, vol.6, pp.290-304, 1997.
[7] P. E. Spronk, P. C. Limburg and C. G. Kallenberg, Serological markers of disease activity in systemic lupus erythematosus, *Lupus*, vol.4, pp.86-94, 1995.

[8] S. Gao, J. Zhang, X. Wang and Z. Tang, Multi-layer neural network learning algorithm based on random pattern search method, *International Journal of Innovative Computing, Information and Control*, vol.5, no.2, pp.489-502, 2009.
[9] Q. Zhang, R. Zhang and B. Wang, DNA sequence sets design by particle swarm optimization algorithm, *International Journal of Innovative Computing, Information and Control*, vol.5, no.8, pp.2249-2255, 2009.
[10] M. H. Bahari, A. Azemi, M. M. Mirsalehi, M. Mahmoodi and M. Khademi, Early diagnosis of systemic lupus erythematosus using artificial neural networks, *Bioinformation*, vol.5, no.2, pp.58-61, 2010.
[11] D. H. Wang, N. K. Lee and T. S. Dillon, Extraction and optimization of fuzzy protein sequence classification rules using GRBF neural networks, *Neural Information Processing – Letters and Rev.*, vol.1, no.1, pp.53-59, 2003.
[12] B. Rost, Review: Protein secondary structure prediction continues to rise, *J. Structural Biology*, vol.134, no.2, pp.204-218, 2001.
[13] I. F. Tsigelny, *Protein Structure Prediction: Bioinformatics Approach*, International University Line, La Jolla, CA, 2002.
[14] G. Pollastri, P. Baldi, P. Fariselli and R. Casadio, Prediction of coordination number and relative solvent accessibility in proteins, *Proteins*, vol.47, no.2, pp.142-153, 2002.
[15] P. Fariselli and R. Casadio, Prediction of the number of residue contacts in proteins, *Proc. of International Conf. Intelligent Systems for Molecular Biology*, vol.8, pp.146-151, 2000.
[16] S. R. Binder, M. C. Genovese, J. T. Merrill, R. Morris and A. L. Metzger, Computer-assisted pattern recognition of autoantibody results, *Clinical and Diagnostic Laboratory Immunology*, vol.12, no.12, pp.1353-1357, 2005.
[17] D. Eilat, D. M. Webster and A. R. Rees, V region sequences of anti-DNA and anti-RNA autoantibodies from NZB/NZW F1 mice, *J. Immunology*, vol.141, no.5, pp.1745-1753, 1988.
[18] J. Kalsi, C. T. Ravirajan, A. Rahman and D. A. Isenberg, Structure-function analysis and the molecular origins of anti-DNA antibodies in systemic lupus erythematosus, *Expert Rev. Molecular Medicine*, pp.1-28, 1999.
[19] W. H. Robinson, C. DiGennaro et al., Autoantigen microarrays for multiplex characterization of autoantibody responses, *Nature Medicine*, vol.8, no.3, pp.295-301, 2002.
[20] R. M. Bill, N. B. Blatt and G. D. Glick, Anti-DNA autoantibodies: The other DNA-binding proteins, *Bioorganic and Medicinal Chemistry*, vol.5, no.3, pp.467-472, 1997.
[21] T. M. Frimurer, R. Bywater, L. Naerum, L. N. Lauritsen and S. Brunak, Improving the odds in discriminating "Drug-Like" from "Non Drug-Like" compounds, *J. Chemical Information and Computer Scenic*, vol.40, pp.1315-1324, 2000.
[22] K. Gulukota, J. Sidney, A. Sette and C. DeLisi, Two complementary methods for predicting peptides binding major histocompatibility complex molecules, *J. Molecular Biology*, vol.267, pp.1258-1267, 1997.
[23] H. Bohr, S. Brunak, R. M. J. Cotterill, B. Lautrup, L. Norskov, O. H. Olsen and S. B. Petersen, Protein secondary structure and homology by neural networks, *FEBS Letters*, vol.241, pp.223-228, 1988.
[24] P. Wrede, O. Landt, S. Klages, A. Fatemi, U. Hahn and G. Schneider, Peptide design aided by neural networks: Biological activity of artificial signal peptidase I cleavage sites, *Biochemistry*, vol.37, pp.3588-3593, 1998.
[25] R. Rajimehr, S. Farsiu et al., Prediction of lupus nephritis in patients with systemic lupus erythematosus using artificial neural networks, *Lupus*, vol.11, pp.485-492, 2002.
[26] J. S. Fetrow, M. J. Palumbo and G. Berg, Patterns, structures, and amino acid frequencies in structural building blocks a protein secondary structure classification scheme, *Proteins*, vol.27, pp.249-271, 1997.
[27] D. F. Specht, A general regression neural network, *IEEE Trans. Neural Networks*, vol.2, no.6, pp.568-576, 1991.
[28] C. Christodoulou and M. Georgiopoulos, *Application of Neural Network in Electromagnetics*, Artech House, Boston, 2001.
[29] C. A. Janeway, P. Travers, M. Walport and M. Shlomchic, *Immunobiology*, 5th Edition, Garland Publishing, New York, 2000.
[30] P. Selby and M. A. Knowles, *Introduction to the Cellular and Molecular Biology of Cancer*, Oxford University Press, Oxford, 2005.

[31] E. Harlow and D. Lane, *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1988.
[32] J. Baynes and M. H. Dominiczak, *Medical Biochemistry*, Mosby, London, 1999.
[33] T. Brody, *Nutrition Biochemistry*, 2nd Edition, San Diego Academic Press, 1999.
[34] J. D. Rawn, *Biochemistry*, Neil Patterson Publishers, Burlington, 1989.
[35] L. Stryer, *Biochemistry*, 3rd Edition, W. H. Freeman and Company, New York, 1988.
[36] F. Sanger, The arrangement of amino acids in proteins, *Advance in Protein Chemistry*, vol.7, pp.1-67, 1952.
[37] H. Paulus, The chemical basis of protein splicing, *Chemical Society Rev.*, vol.27, pp.375-386, 1998.
[38] J. S. Fruton, Early theories of protein structure, *Annals of the New York Academy of Sciences*, vol.325, pp.1-18, 1979.
[39] A. C. R. Martin, Accessing the Kabat antibody sequence database by computer, *Proteins: Structure, Function, and Genetics*, vol.25, pp.130-133, 1996.
[40] H. Hirsh and M. Noordewier, Using background knowledge to improve inductive learning of DNA sequences, *Proc. of the 10th Ann. Conf. Artificial Intelligence for Applications*, pp.351-357, 1994.
[41] C. H. Wu, G. Whitson, J. McLarty, A. Ermongkonchai and T. C. Chang, Protein classification artificial neural system, *Protein Science*, vol.1, no.5, pp.667-677, 1992.
[42] C. D. Huang, C. T. Lin and N. R. Pal, Hierarchical learning architecture with automatic feature selection for multiclass protein fold classification, *IEEE Trans. Nanobioscience*, vol.2, no.4, 2003.
[43] D. L. Nelson and M. M. Cox, *Lehninger, Principles of Biochemistry*, 4th Edition, W. H. Freeman and Company, New York, 2004.
[44] A. Radzicka and R. Wolfenden, Comparing the polarities of the amino acids: Side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution, *Biochemistry*, vol.27, pp.1664-1670, 1988.
[45] T. P. Kyte and R. F. Doolittle, A simple method for displaying the hydropathic character of a protein, *J. Molecular Biology*, vol.157, pp.105-132, 1982.
[46] R. Grantham, Amino acid difference formula to help explain protein evolution, *Science*, vol.185, pp.862-864, 1974.
[47] I. Cosic, Macromolecular bioactivity: Is it resonant interaction between macromolecules? – Theory and applications, *IEEE Trans. Biomedical Eng.*, vol.41, pp.1101-1114, 1994.
[48] S. M. Barbas, H. J. Ditzel, E. M. Salomen, W. P. Yang, G. J. Silverman and D. R. Burton, Human autoantibody recognition of DNA, *Proc. of Natl. Acad. Sci. USA*, vol.92, pp.2529-2533, 1995.
[49] P. Roben, S. M. Barbas, L. Sandoval, J. M. Lecerf, B. D. Stollar, A. Solomon and J. Silverman, Repertoire cloning of lupus anti-DNA autoantibodies, *J. Clinical Investigation*, vol.98, pp.2827-2837, 1996.