



UOIuBIH
ORSinBIH
Operations Research Society in
Bosnia and Herzegovina

Southeast Europe Journal of Soft Computing
Available online: <http://scjournal.ius.edu.ba>



IUS Soft Computing
Research Group

Recurrent Neural Networks for Linear B-Epitope Prediction in Antigens

B. Akcesme, F. Akcesme, M. Adilovic, M. Can

Faculty of Engineering and Natural Sciences,
International University of Sarajevo, Sarajevo,
Hrasnicka Cesta 15, Ildža 71210 Sarajevo,
Bosnia and Herzegovina
fberat@ius.edu.ba; bcicek@ius.edu.ba;
madilovic@ius.edu.ba; mcan@ius.edu.ba

Article Info

Article history:

Article received on 17 June 2017

Received in revised form 17 August
.2017

Keywords:

Prediction of B-cell epitopes;
committee machines; recurrent
neural network

Abstract

Experimental methods used for characterizing epitopes that play a vital role in the development of peptide vaccines, in diagnosis of diseases, and also for allergy research are time consuming and need huge resources. There are many online epitope prediction tools that can help experimenters in short listing the candidate peptides. To predict B-cell epitopes in an antigenic sequence, Jordan recurrent neural network (JRNN) are found to be more successful. To train and test neural networks, 262.583 B epitopes are retrieved from IEDB database. 99.9% of these epitopes have lengths in the interval 6-25 amino acids. For each of these lengths, committees of 11 expert recurrent neural networks are trained. To train these experts alongside epitopes, non-epitopes are needed. Non-epitopes are created as random sequences of amino acids of the same length followed by a filtering process. To distinguish epitopes and non-epitopes, the votes of eleven experts are aggregated by majority vote. An overall accuracy of 97.23% is achieved. Then these experts are used to predict the linear b epitopes of antigen, ESAT6 (Tuberculosis).

1. INTRODUCTION

In a living organism, B-cells of the immune system recognize the pathogen's antigens by their membrane-bound immunoglobulin receptors and, in response, produce antibodies specific to these antigens. Antigens have the capacity to bind by either a B-cell receptor or an antibody molecule. The part of an antigen that binds to an antibody is called a B-cell epitope. If an antigen is a protein, an epitope may be either a short peptide from the

protein sequence or a patch of atoms on the protein surface in the three-dimensional structure.

In the process of antibody production, and antibody detection, one needs to design a molecule that can replace an antigen that is its epitope. If this epitope is a protein, its gene can be cloned, if it is a molecule, it can be synthesized. These designed molecules are inexpensive and noninfectious in contrast to viruses or bacteria.

A peptide is called a continuous epitope of a protein when it corresponds to a short continuous stretch from the sequence of this protein, and binds the antibody raised against this protein. However, it is observed through crystallographic studies of antibodies that, most epitopes consist of atoms from distant residues joined on the protein surface in the three-dimensional space as a patch. In that sense, they are discontinuous. A third kind is nonprotein epitopes, that are polysaccharides, DNA, glycoproteins, and other molecules, that expands the range of potential pathogens that can be targeted by vaccines and diagnostics.

Prediction of immunogenic epitopes using bioinformatics tools is a challenging task because of the inherent complexity of antigen recognition[1].At the beginning, to predict continuous B-cell epitopes, a number of algorithms have been developed based on physico-chemical properties of amino acids [2], but their accuracies were very low. The commonly used properties are hydrophilicity [3]flexibility [4], accessibility [5], and turns [6]. All the prediction calculations are based on the propensity tables for each of the 20 amino acids.

A few computer programs are also developed to make the job predicting epitopes in an antigenic sequence online. PREDITOP [7] uses 22 normalized scales, corresponding to hydrophilicity, accessibility, flexibility, and secondary structure propensities. PEOPLE [8] uses physico-chemical properties, surface accessibility, hydrophilicity, and flexibility. BEPITOPE [9] searches for patterns either in a single protein or on a complete translated genome. BcePred uses also various physicochemical properties [10]. The predictive value of algorithms varies between 40–68% [11],[12]. (Saha 2006)

In 1980s, direct prediction methods based on structural and sequential analysis of T cell epitopes are also developed [13–16]. DeLisi and Berzofsky [13] proposed that the critical requirement of T cell epitopes is its ability to form stable amphipathic structure. Based on this hypothesis, a program AMPHI was developed [14], [15]. Another algorithm SOHHA was developed based on the assumption that T cell epitopes consist of a helix of 3–5 helical turns with a narrow strip of hydrophobic residues on one side.

A number of indirect methods have also been developed that predict MHC binders instead of T cell epitopes. They are based on structure, binding motifs, propensity matrices and Artificial Neural networks (ANNs) [17–25].

In 2004, machine learning techniques like artificial neural networks (ANNs) to predict B cell epitopes started to emerge[12], [25]. In machine learning techniques, the input window length has to be fixed. 262,583 B epitopes in IEDB database, 99.9% of these epitopes have lengths in the range of 6-25 amino acids. To deal with this

abundance, Saha and Raghava [12] take a fixed length of 20 amino acids although 77% of B epitopes in IEDB database have length 15 amino acids as seen in Table 1. Furthermore, they have mistaken in assuming that adding residues taken from the parent/original antigenic sequences or removing residues at the terminals, does not change their character as being of B cell epitopes. In this research, eleven neural networks are trained for each of the 20 epitope lengths between 6 and 25.

2. RECURRENT NEURAL NETWORKS (RNN)

Variables in the data are dependent nonlinearly to the outcome, and hence recurrent neural networks (RNN) are particularly useful. They are constructed by taking a feedforward network and adding feedback connections from output and/or hidden layers to input layers. The standard backpropagation algorithm also trains these networks conditional that patterns must always be presented in time sequential order. The one difference in the structure is that there are extra neurons in the input layer that is connected to the hidden layer and/or output layer just like the other input neurons. These extra neurons hold the contents of one of the layers as it existed when the previous pattern was trained. In this way, the network takes into account previous knowledge it has about previous inputs. These extra neurons are called the context unit and it represents the network's long-term memory [26].

There are two types of RNNs: Jordan, and Elman recurrent networks. A Jordan neural network (JNN) has additional neurons in the input layer, which are fed back from output layer. While an Elman neural network (ENN) has additional neurons in the input layer, which is fed back from hidden layer.

Jordan Recurrent Neural Networks (JNN)

A Jordan neural network (JNN) has several feedback connections from the output layer to the input layer. The input layer has additional neurons, which are fed back from the output layer. Econometric interpretation of such feedback connection lies in the fact that in this way the model is expanded by lagged error terms (Figure 1).

Using JNN, the problem of convergence can be solved by a more complex model. Although this network is more complicated than a multi-layer feed-forward network, the characteristics of feeding back data to the network are similar to a GARCH model, having the previous variance in current forecasts [27].

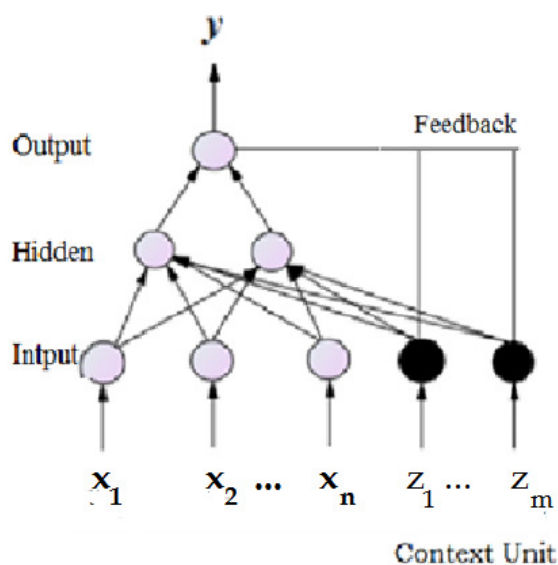


Figure 1. JNN with a single hidden layer

Coding the Peptide Data, and Architecture of the Neural Network

As seen in Figure 1, the JNN used consists of one input, one hidden, and one output layer. An amino acid sequence of length s is coded first into a sparse $20 \times s$ matrix whose 20 rows are for one amino acid each, and columns are for the s residues in the peptide. For example, if the peptide is

WVDVVLHGGCVTVM

and since the 20-letter alphabet of amino acids is

ARNDCQHGEILKMFSTWYV

the matrix representing this peptide will have “1” at the 17th row, first column. To enter into the input layer, this matrix is flattened into a $1 \times 20 \times s$ vector. For the above peptide $s=15$, therefore the vector that represents the peptide is a 1×300 vector. One feedback is taken from the output layer and added to the input. Therefore, the input layer for 15 meer peptides as in the example, input layer consists of 301 gates. Equal number of hidden neurons are taken in one hidden layer.

3. METHODS

To train neural networks, one needs both positive (e.g. B-cell epitopes) as well as negative (e.g. non-B-cell epitope) datasets. We downloaded the positive B-cell epitope data set from the iedb.org. database. Since there is no proven non-epitopes database, peptides with desired number of amino acids are taken randomly. Since these random peptides may also have B-cell epitopes, they are filtered according to a similarity threshold.

Jordanian recurrent neural networks (JRNN)s are used in the present study to predict the B-cell epitopes in an antigenic sequence. For each window length, in the domain 6–25, committee machines with eleven JRNNs in each committee are trained and votes of experts are aggregated by simple majority voting. It is seen that committee machine voting boosts accuracy at least 10%.

3.1 The Data Set

B-cell epitopes are downloaded from IEDB database that contains 365 076 continuous B-epitopes. After elimination of identical peptides, a data with 262.563 non-redundant entries resulted. The distribution of frequencies of epitopes in their lengths are shown in Table 1.

Table 1. Frequencies of B-Epitopes of given lengths

| W Size | Freq | W Size | Freq |
|--------|--------|--------|------|
| 6 | 3399 | 16 | 2352 |
| 7 | 3147 | 17 | 1043 |
| 8 | 7949 | 18 | 1402 |
| 9 | 4739 | 19 | 1160 |
| 10 | 9851 | 20 | 5914 |
| 11 | 1154 | 21 | 618 |
| 12 | 6724 | 22 | 426 |
| 13 | 2221 | 23 | 468 |
| 14 | 2620 | 24 | 528 |
| 15 | 202363 | 25 | 690 |

3.2 Committee Machines

For each of 20 different lengths, committee machines with eleven members are trained. Votes of eleven experts are aggregated by majority voting. In committee machines accuracy is boosted. For example, committee machine for length 17, although accuracy of individual experts are all less than 90%, the accuracy in distinguishing epitopes and non-epitopes is boosted to 99.5% in committee voting.

Table 2. Individual accuracies of committee members and accuracy of committee voting

| Peptide Length 17 | | | |
|-------------------|----------|---------|----------|
| Expert | Accuracy | Expert | Accuracy |
| 1 | 78.5 | 7 | 87.5 |
| 2 | 84.0 | 8 | 81.0 |
| 3 | 78.5 | 9 | 87.0 |
| 4 | 81.5 | 10 | 76.0 |
| 5 | 87.0 | 11 | 88.0 |
| 6 | 84.0 | Overall | 99.5 |

3.3 Accuracy on Test Sets

For each length, from the epitope and non-epitope data, 100 epitopes and equal number of non-epitopes are randomly chosen for disjoint train and test sets.

Performance Measure

It is accustomed to calculate the performance of an estimator by five parameters

- (1) Sensitivity (SENS) is the percent of epitopes that are correctly predicted as epitopes;
- (2) Specificity (SPEC) is the percent of epitopes correctly predicted as non-epitopes;
- (3) Accuracy (ACC) is the proportion of correctly predicted peptides;
- (4) Positive prediction value (PPV) is the probability that a predicted epitope is in fact an epitope;
- (5) Matthew’s correlation coefficient (MCC) [12]. These parameters are calculated by the formulas (1-5).

$$SENS = \frac{TP}{TP+FN}; \tag{1}$$

$$SPEC = \frac{TN}{TN+FP} \tag{2}$$

$$ACC = \frac{TP+TN}{TP+FP+TN+FN}; \tag{3}$$

$$PPV = \frac{TP}{TP+FP} \tag{4}$$

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{5}$$

where TP and FN refer to true positive and false negatives, TN and FP refer to true negatives and false positives.

After training, experts achieved a performance to distinguish epitopes from non-epitopes with an average accuracy of 97.22. Sensitivity and specificity averages are also high, 96.95%, and 97.50% respectively. Positive prediction value (PPV), the probability that a predicted epitope is in fact an epitope is 97.50% in average, and Matthew’s correlation coefficient (MCC) is 0.94 in average. These five performance measures for 20 windows are given in Table3.

Table 3. Sensitivity, specificity, and accuracy of the committee machines in distinguishing epitopes and non-epitopes in testing sets.

| Window | SENS | SPEC | AC% | PPV% | MCC |
|--------|-------|------|------|-------|------|
| 6 | 100. | 100. | 100. | 100. | 1. |
| 7 | 100. | 100. | 100. | 100. | 1. |
| 8 | 91. | 95. | 93.0 | 94.79 | 0.86 |
| 9 | 90. | 90. | 90.0 | 90. | 0.8 |
| 10 | 100. | 100. | 100. | 100. | 1. |
| 11 | 91. | 92. | 91.5 | 91.92 | 0.83 |
| 12 | 99. | 100. | 99.5 | 100. | 0.99 |
| 13 | 96. | 99. | 97.5 | 98.97 | 0.95 |
| 14 | 97. | 100. | 98.5 | 100. | 0.97 |
| 15 | 100. | 100. | 100. | 100. | 1. |
| 16 | 100. | 100. | 100. | 100. | 1. |
| 17 | 99. | 100. | 99.5 | 100. | 0.99 |
| 18 | 89. | 89. | 89.0 | 89. | 0.78 |
| 19 | 94. | 93. | 93.5 | 93.07 | 0.87 |
| 20 | 100. | 99. | 99.5 | 99.01 | 0.99 |
| 21 | 100. | 100. | 100. | 100. | 1. |
| 22 | 100. | 99. | 99.5 | 99.01 | 0.99 |
| 23 | 100. | 100. | 100. | 100. | 1. |
| 24 | 94. | 100. | 97.0 | 100. | 0.94 |
| 25 | 99. | 94. | 96.5 | 94.29 | 0.93 |
| Mean | 96.95 | 97.5 | 97.2 | 97.50 | 0.94 |

3.4 Accuracy in identifying Epitopes as Epitopes.

In practice, these experts will be asked to identify a given peptide as epitope or non-epitope. When all epitopes in Table 1. are sent to experts, the accuracies in identification of these epitopes as epitopes, and non-epitopes as non-epitopes are 90.56, and 91.22 respectively. The difference between Table 3, is due to the fact that, in Table 3, the test set consist of only 100 epitopes, and 100 non-epitopes, while in the next, trained experts are tested on the whole of the large data.

4. A BLIND DATA

To evaluate performances of experts in committee machines, they are tested on the antigen ESAT-6, whose B-Epitopes are already explored in wet laboratories. ESAT-6 protein, a low-molecular weight protein secreted by virulent Mycobacterium tuberculosis, induced strong antibody response in experimentally infected monkeys. The epitopes were determined using synthesis of overlapping peptides spanning of ESAT-6 protein and by measuring antibody response to ESAT-6 peptides by ELISA in serum samples from monkeys [28-33].

ESAT6 has 104 residues in its sequence:

ADPMTEQQWNFAGIEAAASAIQGNVTSIHSLLDDEGK
 QSLTKLAAAWGGSGSEAYQGVQQKWDATATELNN
 ALQNLARTISEAGQAMASTEAGVTGMFAHHHHHH

Kanaujia et. al. in [28] reported eight epitopes

- 1) DPMTEQQWNFAGIEAAASAIQGNV
- 2) EAAASAIQGNVTSIHSLLEDEGKQS
- 3) VTSIHSLLEDEGKQSLTKLAAAWGG
- 4) GKQSLTKLAAAWGGSGSEAYQGVQ
- 5) AWGGSGSEAYQGVQKWDATATEL
- 6) QGVQKWDATATELNNALQNLART
- 7) ATELNNALQNLARTISEAGQAMAS
- 8) LARTISEAGQAMASTEAGNVTGMFA

all have 24 amino acids. Experts in committee machines reported ten epitopes of length 20.

- 1) QWNFAGIEAAASAIQGNVTSIHSL
- 2) AAASAIQGNVTSIHSLLEDEGKQSL
- 2) QGNVTSIHSLLEDEGKQSLTKLAAA
- 2) HSLLDEGKQSLTKLAAAWGGSGSE
- 3) VTSIHSLLEDEGKQSLTKLAAAWGG
- 4) SLLDEGKQSLTKLAAAWGGSGSEA
- 5) GSEAYQGVQKWDATATELNNALQ
- 6) EAYQGVQKWDATATELNNALQNL
- 6) QGVQKWDATATELNNALQNLART
- 7) ATELNNALQNLARTISEAGQAMAS

The amino acids colored in red are common residues in wet lab report, and predicted peptides. The ratio of the correctly predicted amino acids is $124/160 = 77.50\%$.

When the sequence of ESAT6 is uploaded to the best online predictor ABCpred, the following peptides as B-epitopes for ESAT6 are returned:

- 1) KWDATATELNNALQNL
- 2) GGGSGSEAYQGVQKWD
- 3) ASAIQGNVTSIHSLD
- 4) ARTISEAGQAMASTEAG
- 5) AGQAMASTEAGNVTGMF
- 6) PMTEQQWNFAGIEAAA
- 7) AGIEAAASAIQGNVTS
- 8) TSIHSLLEDEGKQSLTK

The amino acids colored in red are common residues in wet lab report, and predicted peptides by ABCpred

- 1) DPMTEQQWNFAGIEAAASAIQGNV
- 2) EAAASAIQGNVTSIHSLLEDEGKQS
- 3) VTSIHSLLEDEGKQSLTKLAAAWGG
- 4) GKQSLTKLAAAWGGSGSEAYQGVQ
- 5) AWGGSGSEAYQGVQKWDATATEL
- 6) QGVQKWDATATELNNALQNLART
- 7) ATELNNALQNLARTISEAGQAMAS
- 8) LARTISEAGQAMASTEAGNVTGMFA

The percentage of the correctly predicted amino acids is $110/160 = 68.75\%$.

5. DISCUSSION

During the last three decades, many algorithms have been developed to predict the location of continuous B cell epitopes in antigen proteins but their rate of successful prediction remained low.

One of the major problems faced in developing B-cell epitope prediction is the variable length, mostly in the range of 6-25, of epitopes. Since machine learning techniques require fixed length of

peptide, we trained 20 different RANNs for each peptide length. To improve the accuracy, indeed eleven experts are trained for each length, 220 RANNs in total. Using majority voting in the committee, accuracy of 97.20 % is achieved in distinguishing epitopes from non-epitopes.

All the existing methods rely on residue properties based on hydrophilicity, flexibility, charge, mass, etc. Summing up the scores of residues from propensity vectors related to these physicochemical properties, finally they select the regions in an antigen protein, which have peaked scores and assign them as B-cell epitopes. These methods are subjective in nature because they heavily depend on the thresholds used, and one can never be sure about the boundaries of epitopes.

The performance of the method that relies on committees of RNNs in this work is the best in literature in distinguishing epitopes and non-epitope peptides of lengths in the region 6-25. The performance of committee machines is also tested on a blind dataset where we compare the predicted and experimentally determined epitopes in antigen ESAT6 protein. It is seen that our results are better than B-cell epitope prediction is made by the famous online tool ABCpred server for detecting potential B-cell epitopes in an antigen.

6. CONCLUSION

Although high 97.20 % prediction accuracy in distinguishing B cell epitopes from non-epitope peptides is achieved in this study, committee machines is not an alternate to existing methods that rely on physicochemical properties. Instead, it may be a complementary to these methods. It is assumed that there is a relation between the compositions of peptides as a whole, and their properties of being an epitope, and RANNs are able to capture these properties. To implement ANN techniques in B-cell epitopes prediction firstly one needs to decide about the length of the peptide. Although committee machines decide about the epitopes with consensus, they propose B cell epitopes in all lengths in the range 6-25. Mostly small ones are nested in larger ones. Then the user needs to distinguish the largest peptides which are not nested in larger ones. Therefore, users are advised to validate the

predicted B-cell epitopes in an antigen by the use of all other existing methods.

REFERENCES

- [1] Flower DR. Towards in silico prediction of immunogenic epitopes. *Trends Immunol*;24:667–674; 2003.
- [2] Pellequer JL, Westhof E, Regenmortel MHV. Predicting location of continuous epitopes in proteins from their primary structures. *Methods Enzymol*;203:176–201; 1991.
- [3] Parker JMD, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*;25: 5425–5432; 1986.
- [4] Karplus PA, Schulz GE. Prediction of chain flexibility in proteins: a tool for the selection of peptide antigen. *Naturwissenschaften*;72:212,213; 1985.
- [5] Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol*;55:836–839; 1985.
- [6] Pellequer J-L, Westhof E, Regenmortel MHV. Correlation between the location of antigenic sites and the prediction of turns in proteins. *Immunol Lett*;36:83–99; 1993.
- [7] Pellequer JL, Westhof E. PREDITOP: A program for antigenicity prediction. *J MolGraphics*;11:204–210; 1993.
- [8] Alix AJ. Predictive estimation of protein linear epitopes by using the program PEOPLE. *Vaccine*;18:311–314; 1999.
- [9] Odorico M, Pellequer JL. BEPITOPE: predicting the location of continuous epitope and patterns in proteins. *J MolRecognit*;16:20–22; 2003.
- [10] Blythe MJ, Flower DR. Benchmarking B cell epitope prediction: underperformance of existing methods. *ProtSci*;14:246–248; 2005.
- [11] VanRegenmortel MHV, de Marcillac GD. An assessment of prediction methods for locating continuous epitopes in proteins. *Immunol Lett*;17:95–107; 1988.
- [12] Saha, S., and Raghava G. P. S. Prediction of Continuous B-Cell Epitopes in an Antigen Using Recurrent Neural Network, *PROTEINS: Structure, Function, and Bioinformatics* 65:40–48; 2006.
- [13] DeLisi C, Berzofsky JA. T-cell antigenic sites tend to be amphipathic structures. *Proc Natl Acad Sci USA*;82:7048; 1985.
- [14] Cornette JL, Margalit H, DeLisi C, Berzofsky JA. The amphipathic helix as a structural feature involved in T cell recognition. In: Epand RM, editor. *The amphipathic helix*. Boca Raton: CRC Press; 1993.
- [15] Spouge JL, Guy HR, Cornette JL, Margalit H, Cease K, Berzofsky JA, et al. Strong conformational propensities enhance T cell antigenicity. *J Immunol*;138:204; 1987.
- [16] Stille CJ, Thomas LJ, Reyes VE, Humphreys RE. Hydrophobic strip of helix algorithm for selection of T cell-presented peptides. *Mol Immunol*; 24:1021; 1987.
- [17] Brusica V, Rudy G, Harrison LC. Prediction of MHC binding peptides by using artificial neural networks. In: *Complex mechanism of adaptation*. Amsterdam: IOS Press. p. 253–60; 1994.
- [18] Gulukota K, Sidney J, Sette A, DeLisi C. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J MolBiol*; 267:1258; 1997.
- [19] Rammensee HG, Friede T, Stevanović S. MHC ligands and peptide motifs: first listing. *Immunogenetics* 41:178 ;1995.
- [20] Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J Immunol*;152:163; 1994.
- [21] Adams HP, Koziol JA. Prediction of binding to MHC class I molecules. *J Immunol Methods*;185:181; 1995.
- [22] Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. *Bioinformatics* 2001;17:1236.
- [23] Singh H, Raghava GPS. ProPred1: prediction of promiscuous MHC class I binding sites. *Bioinformatics*;19:1009; 2003.
- [24] Bhasin M, Raghava GPS. SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence. *Bioinformatics*;20:421; 2003.
- [25] Bhasin, M., and Raghava, G.P.S. Prediction of CTL epitopes using QM, SVM and ANN techniques, *Vaccine* 22, 3195–3204; 2004.
- [26] Balkin, S.D. Using recurrent neural networks for time series forecasting, Working Paper Series number 97–11, International Symposium on Forecasting, Barbados; 1997.
- [27] Dechpichai, P. Nonlinear neural network for conditional mean and variance forecasts, Doctor of Philosophy thesis, University of Wollongong, School of Mathematics and Applied Statistics, University of Wollongong, Dubai ; 2010.
- [28] Kanaujia GV, Motzel S, Garcia MA, Andersen P, Gennaro ML. Recognition of ESAT-6 sequences by antibodies in sera of tuberculous nonhuman primates. *ClinDiagn Lab Immunol*; 11:222–226; 2004.
- [29] Harboe M, Malin AS, Dockrell HS, Wiker HG, Ulvund G, Holm A, Jorgensen MC, Andersen P. B-cell

epitopes and quantification of the ESAT-6 protein of *Mycobacterium tuberculosis*. *Infect Immun*;66:717–723; 1998.

[30] Aagaard CS, Hoang TTKT, Lundberg CV, Dietrich J, and Andersen P, Quality and Vaccine Efficacy of CD4 T Cell Responses Directed to Dominant and Subdominant Epitopes in ESAT-6 from *Mycobacterium tuberculosis*1, *The Journal of Immunology*, 183: 2659–2668; 2009.

[31] Lindestam Arlehamn CS, Sidney J, Henderson R, Greenbaum JA, James EA, Moutaftsi M, Coler R, McKinney DN, Park D, Taplitz R, Kwok WW, Grey H, Peters B, and Sette A, Dissecting Mechanisms of Immunodominance to the Common Tuberculosis Antigens ESAT-6, CFP10, Rv2031c (hspX), Rv2654c (TB7.7), and Rv1038c (EsxJ), *J Immunol*, 188 (10) 5020-5031; 2012.

[32] Tiwari D, Haque S and Chandra R, Vaccine development for Tuberculosis: Past, Present and Future Challenges, *Int. J. Drug Dev. & Res.*, April-June, 3 (2): 75-84; 2011.

[33] Yang H, Chen H, Liu Z, Ma H, Qin L, Jin R, Zheng R, Feng Y, Cui Z, Wang J, Liu J, and Hu Z, A Novel B-Cell Epitope Identified within *Mycobacterium tuberculosis* CFP10/ESAT-6 Protein, *PLOS ONE*, V. 8, No 1, 1-9, 2013.