

Accurate Prediction of Protein Secondary Structure By Non-Parametric Models

İrem Ersöz Kaya¹, Turgay İbrikçi^{2*} and Ayça Çakmak³

¹*Mersin University Tarsus Technical Education Faculty,*

Department of Electronics and Computer, Tarsus, Mersin, Türkiye

²*Cukurova University Department of Electrical-Electronics Engineering,
Adana, Türkiye*

³*Kultur University Department of Computer Engineering, Istanbul, Türkiye*

**Corresponding Author E-mail: ibrikci@cu.edu.tr*

Abstract

Proteins are one of the most important parts of an organism because of their vital tasks. Consequently, it is necessary to know both the primary and secondary structure of a protein as closely related to its biological function. Artificial Neural Networks (ANNs) are a useful methodology for secondary structure prediction of proteins. In this study, a generalized regression neural network (GRNN), a probabilistic neural network (PNN) and a backpropagation algorithm (BP) were investigated with different window sizes of amino acid sequences to predict the protein secondary structure from the primary structure.

All the outputs of the networks were further combined with another GRNN for possible increase in accuracy rates. The data sets were prepared with alpha-141 and beta-146 hemoglobin chains from the Protein Data Bank. Once the window size was optimized, the overall success rate of GRNN was around 90.2-91.7% for beta chains and 85.9-87.3% for alpha chains. The PNN achieved between 91.2-92% overall accuracy for beta chains and 85.4-86.5% for alpha chains.

The BP had an overall accuracy rate of 89.9-92.5% for beta chains, and 86.5-90.3% for alpha chains.

Since the networks compared gave similar results, it was concluded that it is very important to choose the optimal window size given a particular network to achieve best accuracy rates.

Keywords: Generalized Regression Neural Network, Probabilistic Neural Network, Non-parametric Models.

Introduction

The protein molecules represent much of the bulk of an organism and accomplish almost all of its biochemical activities. To understand the life process of an organism, it is necessary to know the structure of a protein since it is closely related to its biological function. Routine and reliable methods to determine the structure of a protein are laborious techniques such as X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR). Since biological techniques are quite expensive and time consuming, scientists have used computer and statistical modeling for protein secondary structure prediction in recent years. In this study, the aim is to investigate the success rate of neural network algorithms as a function of window size and type of neural network in predicting the secondary structure of a protein as a first step towards determination of the overall structure.

All proteins are composed of linear unbranched combinations of twenty different amino acids. The structural and functional properties of the proteins primarily depend on this linear sequence of amino acids. Protein structures are described through four main hierarchical levels; *Primary*, *Secondary*, *Tertiary*, *Quaternary*. There are three common secondary structures in proteins, namely, alpha helices (H), beta sheets (E), and turns or coils (C) [1], [2].

ANN's are a useful approach for prediction of secondary structures of proteins [3], [4]. They have the capability of mapping the relationships between the protein sequences. In this paper, we focus on improving the prediction accuracy with non-parametric, thus fast, neural network methods.

Studies on the protein secondary structure prediction by computational and statistical methods started with Krigbaum and Kuntton as pointed out by Cai et al. [5]. They used multiple linear regression algorithms to predict the content of protein secondary structure elements that was based on pair-coupled amino acid composition. This attempt continued with the Chou-Fasman method that achieved a three-state (Q3) accuracy of 52% [6]. This method is a well-known empirical statistical algorithm that is based on the frequencies of the secondary structure types.

At the end of 1980's, the researches about the secondary structure prediction were developed considerably along with the initial use of neural network methods. The average accuracy rate reached 64.3% on three types of secondary structure conformation (α -helix, β -sheet and coil) of a non-homologous data set by using a backpropagation neural network model [1]. Qian and Sejnowski applied the sliding window strategy to preprocess the data, similarly to Chou and Fasman [6]. Within the field of protein secondary structure prediction, the idea of combining different prediction methods is also well-established. For example, Zhang et al. used a final neural network to combine predictors obtained from another neural network, nearest neighbor and naive Bayes algorithms with 66.4% accuracy [7].

In the last decade of the 20th century, the accuracy percentage level rose over 70%. During this period, Rost et al. extensively studied the prediction of secondary structure [8], [9]. The accuracy rate of the network, which trained a set of profiles of multiple alignment sequences, was about 70%. They also used an ensemble of networks, and combined predictions from 10 distinct neural network systems to

improve the accuracy as part of their PHD predictor using about 10000 parameters to the top of CASP2 [10], [11].

Salamov and Solovyev applied their NNSSP (nearest neighbor secondary structure prediction) method on non-membrane non-homogenous proteins with 71%-72% overall three-state accuracy [12]. The NNSSP is based on nearest neighbor rule, which carries out classification for a test sample according to the classifications of neighbor training examples from a database of known structure. In 1998, one of the combination methods, Jpred, was applied by Cuff [13]. This method combined the outputs of several specific methods such as PHD, NSSP, etc. by using a consensus method, and achieved 72.9% accuracy rate. Chandonia and Karplus obtained a 74.8% accuracy rate on non-homologous protein data that is considerably larger than the other data sets [14]. Jones proposed a predictor, named PSIPred which was the top performer at CASP3. The PSIPred predictor was the highest scoring neural network technique which achieves an overall Q3 of about 77% [15].

In 2000, Petersen et al. achieved 77.2%-80.2% accuracy by structure filtering neural networks [16]. More recently, three and eight classes of protein secondary structures were predicted with the performance of about 78% accuracy rate by using recurrent neural networks and profiles in the study of Pollastri et al. [17]. Using about 2000 parameters, they employed bidirectional recurrent Networks in SSPro.

Recently, Support Vector Machines (SVMs) and Hidden Markov Models (HMMs) have also been used as standalone tools for secondary structure prediction. Nguyen and Rajapakse presented an integrated SVM classifier and Bayesian predictor framework for secondary structure prediction [18]. The Q_α score was 74.1% for the alone and 77.0% for the Bayesian/SVM hybrid. Nguyen and Rajapakse reported an overall Q3 of 79.5% on the PSIPred database [19]. A neural network has been followed by an HMM, resulting in a simple and fast system [20].

Pollastri and McLysaght developed a server for protein secondary structure prediction. The server, Porter, also uses Bidirectional Recurrent Neural Networks as in SSpro predictor [21]

Data Sets and Methods

Data Collection and Preprocessing

Hemoglobin protein family is one of the most important proteins and is frequently investigated because the various types of disorders in its structure cause many diseases in the human body [22]. The hemoglobin molecule is made up of four polypeptide subunits that are known as globins and four heme groups for each globin chain, as its name states [23]. The four globins of normal adult hemoglobin consist of two identical alpha chains, each with 141 amino acids and two beta chains each with 146 amino acids. In this study, the dataset was taken from the Protein Data Bank, USA and consisted of 2562 alpha-141 and 2540 beta-146 hemoglobin chains [24].

The networks are trained to predict three categories, helix, strand and coil, on a secondary structure assignment reduced from the eight-category assignment produced by the DSSP program [25]. In this dataset, there is no strand (E) structure, and thereby only helix and coil represent the secondary structure. 1890 helix and 672 coils exist in

the alpha-141 set, and 1995 helix and 545 coils exist in the beta-146 set. The training files contain a primary structure and its corresponding secondary structure. These structures are manipulated by determining a certain size of sliding window that consists of contiguous amino acid residues. Since there are 20 known amino acids in nature, each of them is coded as a 20-bit string with all but one bit turned on, and each output is coded as a 1-bit string that is chosen 0 for helix and 1 for coil.

Based on a previous study, the value of the window size was chosen as 19 [26]. The centering technique is based on the assumption that the central amino acid has a large influence in the structural classification of that window [5]. Prediction is applied by labeling the secondary structure corresponding to the input pattern. Starting with the first residue, the window is moved by sliding one residue to the right until the end of the sequence. Then, all residues are classified.

General Regression Neural Networks

The General Regression Neural Networks (GRNN) introduced by Donald Specht in 1990 is a memory-based feed forward neural network based on the estimation of the probability density function (pdf) from observed samples using Parzen-window estimation [27]. It approximates any arbitrary function between input and output vectors. This approach removes the necessity to specify a functional form of estimation.

The method utilizes a probabilistic model between an independent random vector \mathbf{X} (input) and a dependent scalar random variable Y (output). Let \mathbf{x} and y be the particular measured values of \mathbf{X} and Y , respectively, and $g(\mathbf{X}, Y)$ be the joint continuous probability density function of \mathbf{X} and Y . A good choice for a non-parametric estimate of the probability density function g is the Parzen window estimator as proposed by Parzen and performed for multidimensional cases by Cacoullos [28]. Given a sample of n real D dimensional \mathbf{x}_i vectors and corresponding scalar y_i values, the estimate of joint probability density is given by

$$\hat{g}(x, y) = \frac{1}{(2\pi)^{(d+1)/2} \sigma^{(d+1)}} \frac{1}{n} \sum_{i=1}^n \left[\exp\left(-\frac{(x-x_i)^T(x-x_i)}{2\sigma^2}\right) \exp\left(-\frac{(y-y_i)^2}{2\sigma^2}\right) \right] \quad (1)$$

where σ is the window width of a sample probability, called the smoothing factor of the kernel [29].

The expected value of Y given \mathbf{x} (the regression of Y on \mathbf{x}) is given by

$$E[Y/x] = \frac{\int_{-\infty}^{+\infty} Y \cdot g(x, Y) dY}{\int_{-\infty}^{+\infty} g(x, Y) dY} \quad (2)$$

Using Eq. (1), Eq. (2) becomes

$$\hat{y}(x) = E[Y/x] = \frac{\sum_{i=1}^n [y_i \exp(d_i)]}{\sum_{i=1}^n \exp(d_i)} \quad (3)$$

where d_i is the distance function between the input vector and the i^{th} training vector, and is given by

$$d_i^2 = -\frac{(x-x_i)^T(x-x_i)}{2\sigma^2} \quad (4)$$

The estimate $\hat{y}(x)$ is thus a weighted average of all the observed y_i values where each weight is exponentially proportional to its Euclidean distance from \mathbf{x} .

The structure of the GRNN consists of 4 layers; the input layer, the hidden (pattern) layer, the summation layer and the output layer. As a preprocessing step, all input variables of the training data are scaled. Then, they are copied as the weights into the pattern units. The summation layer has two units that can be denoted as the numerator and the denominator of Eq. (3). Lastly, the output layer gives the estimate of the expected value of $\hat{y}(x)$. If y and \hat{y} are the vector variables, the results above are generalized by adding, one summation unit for each component of \hat{y} in the output layer.

The only adjustable parameter of the network is σ , the smoothing factor for the kernel function. It is critical to decide an optimum value for σ . The larger values of this factor cause the density to be smooth and $\hat{y}(x)$ then converges to the sample mean of the observed y_i . On the other hand, when σ is chosen very small, the density is forced to have non-gaussian shapes. Then, the oscillatory points have a bad effect on the estimate. All values of y_i are taken into account where the points closer to \mathbf{x} are given heavier weights, if the optimum value of σ is selected [26].

Probabilistic Neural Networks

The probabilistic neural network (PNN), also introduced by Donald Specht, is a feed-forward and one-pass training algorithm primarily based on Bayes-Parzen Classification [30]. Parzen's method is used for estimating the probability density function (pdf) of random variables from a set of training examples as in GRNN.

To understand the Bayesian technique, consider a pattern vector \mathbf{x} with dimension d taken from a set of examples that belong to k number of distinct classes (1,2,...,k). Let p_i be the prior probability of the i^{th} class. For all classification classes, assume that $F_i(\mathbf{x})$ is the conditional probability density functions of the i^{th} class given \mathbf{x} . According to Bayes' decision rule, \mathbf{x} belongs to the i^{th} class if

$$p_i F_i > p_j F_j \quad (5)$$

for all categories $j \neq i$ [31].

Typically, the class densities of an unknown input vector are calculated from all the training pattern vectors in the PNN by using the multivariate case of Parzen's pdf estimator as for the GRNN. However, PNN requires the conditional pdf for each class instead of the joint pdf between input and output. The estimate of the conditional probability density function for the i^{th} class is expressed as

$$\hat{f}_i(x) = \frac{1}{(2\pi)^{(d+1)/2} \sigma^{d+1}} \frac{1}{m_i} \sum_{j=1}^{m_i} \exp\left(-\frac{(x-x_{ij})^T(x-x_{ij})}{2\sigma^2}\right) \quad (6)$$

where m_i is the number of training patterns that belong to class i , x_{ij} is the j^{th} training pattern of this class and x is the i^{th} test vector [32].

Subsequently, a classification of the unknown input vector is performed according to the highest pdf as directed by Bayes' decision rule expressed by the equation 5.

Results

The algorithms of GRNN, PNN and Backpropagation (BP) are applied to the datasets of 141 alpha and 146 beta chains constructed by sliding windows for the prediction of the secondary structure of hemoglobin [33]. The GRNN and PNN are accepted as a type of nonparametric approximation whereas the PB is notably a parametric model.

Suggestions of Selections

As mentioned above, the optimum value of the window size was chosen as 19 according to the evaluation of the accuracy results that were given in the study of Ersoz et al. [26]. Eight different window sizes were used in the study to obtain comparative accuracy rates of secondary structure prediction for two classes.

The BP algorithm was applied on a multi-layer perceptron that consists of one hidden layer with 50 hidden units. The chosen value of learning rate parameter was 0.01 as in references [34], [35]. The BP was controlled with the minimum error value that was chosen as 0.001.

The only parameter, sigma (σ) of the algorithms of GRNN and PNN were selected as 1.2 and 0.3, respectively. The sigma values have been chosen between 0.1 and 2 by testing with an increase of 0.1 for both methods and the values are superior to the others for the improvement of accuracy and efficiency. When the sigma value approached 2, the success rate of helix prediction approached 100% and the coil success rate approached zero. This indicates that the larger values of sigma cause a convergence of the density to the mean of the sample as mentioned before [27].

Evaluation Measures

By using the optimum window size, successes of the methods were discussed in our study. The accuracy rates for helix (H) and coil (C) were separately denoted by Q_1 . Then, all results were calculated for overall accuracy rates in which the overall accuracy Q_2 is computed with the following equation:

$$Q_2 = \frac{H+C}{N} \% \quad (7)$$

where N shows the total number of predicted residues.

To evaluate the generalization performance of the algorithms, leave-one-out type of cross validation was used in this study. In this type of cross-validation, a network of sample size n is trained using $n-1$ cases and tested on the single remaining case, then repeated for each case in turn [36], [37].

The quality performance of the prediction was popularly measured by sensitivity (S), specificity (Sp) and Matthews correlation coefficient (Cc) defined by the following equations;

$$\text{Sensitivity } (S) = \frac{TP}{TP + FN} \quad (8)$$

$$\text{Specificity } (Sc) = \frac{TN}{TN + FP} \quad (9)$$

$$\text{Correlation Coefficient } (Cc) = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP + FP) \times (FP + TN) \times (TN + FN) \times (FN + TP)}} \quad (10)$$

where TP (true positive) is the number of correctly classified coil structured residues, TN (true negative) is the number of correctly classified residues which have the helix structure, FP (false positive) is the number of helix structured residues incorrectly classified as coil and FN (false negative) is the number of coil structured residues incorrectly classified as helix.

Therefore sensitivity (S) and specificity (Sp) represent the fraction of correctly identified residues as coil and helix, respectively [38]. The difference value between the specificity and sensitivity of approximately more than 30% shows an unbalanced tendency of prediction.

For instance, a sensitivity of less than 50% but a specificity of more than 80% demonstrates an under-prediction of a predictor which has the tendency of predicting helix more than coil.

The correlation coefficient, introduced by Matthews, gives 1, 0 and -1 for perfect, random and completely wrong predictions, respectively [39]. Although the Matthews correlation coefficient provides a much more balanced evaluation of the prediction than the percentages, all three measures are critically affected by the relative frequency of the target and they are not suitable for the isolated evaluation. Therefore probability excess was recommended as an independent measure for evaluating the performance of prediction by Yang et al. [6].

Probability excess is given as sensitivity + specificity – 1 concisely and defined by the equation 11;

$$\text{Probability Excess } (ProbEx) = \frac{TP \times TN - FP \times FN}{(TP + FN) \times (TN + FP)} \quad (11)$$

Performances of Testing

The performance comparisons of testing the data sets 141 alpha and 146 beta are provided in Tables 1 and 2, respectively. Since all three methods have the probability excess of more than 0.5 for both data sets, it can be accepted that all methods provide suitable predictions. However, when compared with the results of the BPN, non-parametric models perform the best in all measures.

Table 1: Testing results with the 141 alpha set.

141	Q ₂	S	Sp	Cc	ProbEx
GRNN	90.04	0.719	0.965	0.733	0.684
PNN	90.00	0.713	0.967	0.732	0.680
BP	86.15	0.655	0.934	0.625	0.589

Table 2: Testing results with the 146 beta set.

146	Q₂	S	Sp	Cc	ProbEx
GRNN	92.17	0.754	0.967	0.759	0.721
PNN	92.20	0.752	0.968	0.760	0.720
BP	90.94	0.709	0.966	0.724	0.675

All measures also show that the PNN and GRNN results for both the 141 alpha and the 146 beta data sets resemble each other, excessively and especially based on the probability excess values, quite good accuracies were obtained with the non-parametric methods.

The multilayer perceptron (MLP) with BP algorithm has distinguishing low performance on testing the 141 alpha set according to the value of Matthews correlation coefficient (Cc); even it can be claimed that the Cc value of 0.641 is close to random prediction. Furthermore BP algorithm on testing 141 alpha set roughly approaches the probability excess value of 0.5. Although the measures of testing data 146 with BPN are relatively better when compared with the performance on testing the other data set, it can be said that BPN has the tendency of under-prediction. The non-parametric methods, GRNN and PNN outperform the parametric learning without either under-predicting or over-predicting.

The results reveal that the GRNN and the PNN appear to be more robust in the prediction of the class with lower probability of occurrence than the multilayer perceptron algorithm. Hence, resampling procedures are less necessary with the GRNN and the PNN methods [40]. This is probably due to the simplified estimation of the density function allowing a single parameter sigma (σ) to be determined from available data. As discussed previously, sigma is often determined heuristically as well and this allows less dependence on available data.

While the study is currently limited to an alpha helix prediction due to the unbalanced helix/coil distribution, for this task it performs better sensitivity ($Q_\alpha = 75.4\%$ and 71.9%) than many of the studies described above; furthermore, it does so fast (more or less 180 seconds) without a parameter.

Discussion and Conclusion

The problem of predicting a protein, hemoglobin, secondary structure is successfully handled by the networks. As we can observe from the results obtained, the non-parametric models, GRNN and PNN are more successful than the MLP network, especially in the prediction of unbalanced distributions. GRNN and PNN allow much more reliable prediction of the coil category because of their probabilistic approach and simplicity, thereby less dependence on available data. Furthermore, there are some other advantages of these non-parametric probabilistic neural networks. Because of their one-pass learning algorithm, GRNN and PNN have the advantage of fast training time as compared with the BP training time. Another advantage of the GRNN and the PNN is the fact that adding new samples to the training set does not require

re-calibrating the model. Furthermore, GRNN and PNN have only one parameter, whereas BP has many parameters to be learned. Although PNN and GRNN have totally different algorithms, their pros and cons show similar qualities in terms of speed of learning, data variation, adding new samples, choosing the smoothing factor, and so on. It is possible with the BP algorithm to get different results for the same data set at each training process because of starting with random parameters and reaching a different local minimum during learning. Thus, it is typical that the success rate with the same data set may not be same for the next time of training.

Acknowledgements

This research was supported by Cukurova University Research Foundation FBE2002-YL328 and TUBITAK-EEEAG-100E037. We thank Prof. Dr. Kiyem Aksoy and Erdinc A. Yalin, our colleagues at the Medical School of Cukurova University for their help in understanding the hemoglobin structures and their contribution to the discussions.

References

- [1] Denniston K J, Topping J J, Caret R L (2001) General, Organic and Biochemistry. Third International Edition, McGraw-Hill Companies Inc., New York.
- [2] Solomons T W G (1995) Organic Chemistry. Fifth Edition. John Wiley & Sons Inc, New York.
- [3] Qian N, Sejnowski T J (1988) Predicting the Secondary Structure of Globular Proteins Using Neural Network Models. *J. Mol. Biol.*, Vol.202, 865-884.
- [4] Ibrikci T (2000) Phd Thesis. Cukurova University
- [5] Cai Y, Liu X, Xu X, Chou K (2002) Artificial Neural Network for Predicting Protein Secondary Structure Content. *Computers and Chemistry*, 26, 347-350.
- [6] Chou P Y, Fasman G D (1978) Empirical Predictions of Protein Conformation. *Annual Review Biochemistry*, Vol. 47, 251-276.
- [7] Zhang X, Mesirov J P, Waltz D L (1992) Hybrid System for Protein Secondary Structure Prediction. *J. Mol. Biol.*, 225, 1049–1063.
- [8] Rost B, Sander C (1993a) Improved Prediction Of Protein Secondary Structure By Use Of Sequence Profiles And Neural Networks. *Proc. Natl. Acad. Sci., U S A*, 90, 7558-62.
- [9] Rost B, Sander C (1993b) Prediction of Protein Secondary Structure at Better Than 70% Accuracy. *J. Mol. Biol.*, 232, 584-599.
- [10] Rost B, Sander C (1994) Combining Evolutionary Information and Neural Networks to Predict Protein Secondary Structure. *Proteins*, 19, 55-72.
- [11] Rost B, Sander C, Schneider R (1994) Redefining the Goals of Protein Secondary Structure Prediction. *J. Mol. Biol.*, 235, 13-26.
- [12] Salamov A A, Solovyev V V (1997) Protein Secondary Structure Prediction Using Local Alignments. *J. Mol. Biol.*, 268, 31-6.

- [13] Cuff J A, Clamp M E, Barton G J (1998) Jpred: A Consensus Secondary Structure Prediction Server. *Bioinformatics*, 14, 892-893.
- [14] Chandonia J M, Karplus M (1999) New Methods for Accurate Prediction of Protein Secondary Structure. *Proteins*, 35, 293-306.
- [15] Jones D.T. and Ward J.J (2003) Prediction of Disordered Regions in Proteins from Position Specific Scoring Matrices. *Proteins*, 53, 573-578.
- [16] Petersen T N, Lundegaard C, Nielsen M, Bohr H, Bohr J, Brunak S, Gippert G P, Lund O (2000) Prediction of Protein Secondary Structure at 80% Accuracy. *Proteins: Structure, Funct. and Genetics*, Vol. 41, 17-20.
- [17] Pollastri G, Przybylski D, Rost B, Baldi P (2002) Improving the Prediction of Protein Secondary Structure in Three and Eight Classes Using Recurrent Neural Networks and Profiles. *Proteins*, 47, 228-235.
- [18] Nguyen M N and Rajapakse J C (2002) Prediction of protein secondary structure using bayesian method and support vector machines. In 9th International Conference on Neural Information Processing, volume 2, pages 616–620.
- [19] Nguyen M N and Rajapakse J C (2003) Multi-Class Support Vector Machines for Protein Secondary Structure Prediction. *Genome Informatics*, 14:218–227.
- [20] Lin K, Simossis V A, Taylor W R, Heringa J (2005). A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics*, 21(2):152–159.
- [21] Pollastri G and McLysaght A (2005) Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics*, 21(8):1719–1720, 2005.
- [22] Pauling L (1952) The Hemoglobin Molecule In Health And Disease. *Proceedings of the American Philosophical Society*, Vol. 96, No: 5, October, 555-565.
- [23] Rinsho N (1998) Human Hemoglobin Structure and Respiratory Transport. *Department Of Human Genetics. National Institute Of Genetics*, 9, 54, 2320-2325.
- [24] Protein Data Bank, Chemistry Department, Brookhaven National Laboratory. Upton, Ny 11973 Usa, www.rcsb.org/pdb/
- [25] Kabsch W, Sander C (1983) Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers*, 22: 12, 2577-637.
- [26] Ersoz I, Ibrikci T, Cakmak A, Ersoy O K (2004) Secondary Structure Prediction of Hemoglobin by Neural Networks. ANNIE 2004 Artificial Neural Networks in Engineering, University of Missouri-Rolla, Missouri, USA.
- [27] Specht D (1991) A General Regression Neural Network. *Ieee Trans. Neural Networks*, Vol. 2, No 6, 568-576.
- [28] Cacoullos T (1966) Estimation of a Multivariate Density. *Annals of Inst. Stat. Math.*, Vol. 18, 178-189.
- [29] Yeung D, Chow C (2002) Parzen Window Network Intrusion Detectors. *Proceedings of the Sixteenth International Conference on Pattern Recognition. Quebec City, Canada, Vol.4, 385-388.*

- [30] Specht D (1990) Probabilistic Neural Networks. *Neural Networks*, Vol. 3, 109-118.
- [31] Goh A T C (2002) Probabilistic Neural Network for Evaluating Seismic Liquefaction Potential. *Canadian Geotechnical Journal*, Vol. 39, 219-232.
- [32] Niwa T (2003) Using General Regression and Probabilistic Neural Networks to Predict Human Intestinal Absorption with Topological Descriptors Derived from Two-Dimensional Chemical Structures. *J. Chem. Inf. Computer Sci.*, 43, 113-119.
- [33] Ersoz, I (2003). Master Thesis, Cukurova University.
- [34] Duda R O, Hart P E, Stork D G (2000) *Pattern Classification*. Wiley, Second Edition, 654.
- [35] Bishop C M (1995) *Neural Networks for Pattern Recognition*. Oxford University Press Inc., New York, 482.
- [36] Benediktsson J, Swain P H, Ersoy O K (1997) Parallel Consensual Neural Networks, *Ieee Tran. Neural Networks*, Vol. 8, No. 1, 54-64.
- [37] Cuff J A, Barton G J (1999) Evaluation and Improvement of Multiple Sequence Methods for Protein Secondary Structure Prediction. *Proteins: Structure, Function, and Genetics* 34:508–519.
- [38] WU C H, McLarty J W (2000) *Neural Networks and Genome Informatics*. Elsevier Science Ltd., Vol. 1, UK, 97-99.
- [39] Matthews B W (1975) Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme. *Biochim Biophys Acta* 405, 442-451.
- [40] Baldi P, Brunak S (2001) *Bioinformatics*. Massachusetts Institute of Technology Press, Cambridge, Massachusetts, London, England, 126, 139.