

# New Applications of Soft Computing in Bioinformatics: A Review

Amit Gaurav<sup>1</sup>, Vivek Kumar<sup>2</sup> and Dr. Darshika Nigam<sup>3\*</sup>

<sup>1</sup>Galgotias College of Engg. & Tech., Greater Noida (UP)

<sup>2</sup>Delhi College of Tech. & Mngment, Palwal, Haryana

<sup>3\*</sup>Deptt. of Biochemistry, SLS, Dr. B.R. Ambedker University, Agra (UP)

*Bioinformatics is a promising and innovative research field. Soft Computing is playing a crucial role as it provides techniques that are especially well suited to obtain results in an efficient way and with a good level of quality. Soft Computing can also be useful to model the imprecision and uncertainty that the Bioinformatics data and problems have. In this paper, we survey the role of different soft computing paradigms, like Fuzzy Sets (FSs), Artificial Neural Networks (ANNs), evolutionary computation, Rough Sets (RSs), and Support Vector Machines (SVMs), biologically inspired algorithm like ant colony system, swarm intelligence and others in bioinformatics systems and problems. In broader view the present review reveals the major process and functions which are handled by these computing techniques are pattern-recognition and data-mining tasks, clustering, classification, feature selection, and rule generation of Genomic sequence, protein structure, gene expression microarrays, and gene regulatory networks.*

**Keywords:** Bioinformatics, Soft computing paradigm, Fuzzy sets, Artificial neural network, Rough sets, Support vector Machines, Ant colony system.

## 1. INTRODUCTION

### 1.1. Soft computing

Soft Computing [1,2] is a discipline of methodologies that provides, in one form or another, intelligent information processing capabilities for handling real life. They are to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth. They are to achieve tractability, robustness, low solution cost, and close resemblance with human like decision-making.

The major components of Soft Computing are Fuzzy Sets (FS), Artificial Neural Networks (ANN), Evolutionary Algorithms (EAs) (including genetic algorithms (GAs), genetic programming (GP), Evolutionary Strategies (ES)), Support Vector Machines (SVM), Wavelets, Rough Sets (RS), Simulated Annealing (SA), biological inspired Swarm Optimization (SO), Memetic Algorithms (MA), Ant Colony Optimization (ACO), Tabu Search (TS) and artificial immune system.

The algorithms, and tools in bioinformatics for dedicated purposes such as reliable and parallel genome sequencing, fast sequence comparison, search in databases, automated gene identification, efficient modeling and heterogeneous data mining, protein

synthesis. The basic problems in bioinformatics like protein structure prediction, multiple alignment, phylogenetic inference etc. are NP-hard problem. Soft computing offers an effective approach to achieve efficient and robust solution. On the other side the continuous development of high quality biotechnology, e.g. micro-array techniques and mass spectrometry, which provide complex patterns for the direct characterization of cell processes, offers further promising opportunities for advanced research in bioinformatics. With concept, Computer science and statistics are the subjects by which an integrated understanding of relevant processes in systems biology may be considered.

## 1.2. Bioinformatics

Bioinformatics is the application of computer technology to the management of biological information. Computers are used to gather, store, analyze and integrate biological and genetic information which can then be applied to gene-based drug discovery and development. Bioinformatics can be defined as the application of computer technology not only to simulate biological processes and data, but includes a study of the inherent genetic information, underlying molecular structure, resulting biochemical functions, and the exhibited phenotypic properties. The discipline imposes challenges to analyze and interpret the vast amount of data that are available, involving the decoding of huge volume of human genes. *Biological data mining* is an emerging field of research and development, which addresses the challenges and the possibilities in this direction. Bioinformatics is a promising and innovative research field which has evolved in response to this deluge of information and to handle biological data by simulating. It is an interdisciplinary field involving biology, computer science, mathematics and statistics to analyze biological sequence data, genome content & arrangement, and to predict the function and structure of macromolecules. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be derived.

Different biological problems considered within the scope of bioinformatics involve the study of genes, proteins, nucleic acid structure prediction, and molecular design with docking. A broad classification of the various bioinformatics tasks is given as follows. They can be broadly classified as functions defined on proteins, DNA sequences, gene etc.

- (a) Protein structure prediction and classification [3,4,5,6,7,8].
- (b) Alignment and comparison of protein sequences [9].
- (c) Molecular design and molecular docking [8,10].
- (d) DNA structure prediction [11].
- (e) RNA structure prediction [12].
- (f) Alignment and comparison of DNA, RNA sequences [13].
- (g) Genes finding and promoter identification from DNA sequences [13].
- (h) Gene regulatory network identification [14].
- (i) Construction of phylogenetic trees for studying evolutionary relationship [15].
- (j) Gene mapping on chromosomes [16].
- (k) Interpretation of gene expression and micro-array data [17,18].

Bio-computing [19] is useful in recognition techniques to detect similarity between sequences and hence to interrelate structures and functions. Another important

application of bioinformatics is the direct prediction of protein 3-Dimensional structure from the linear amino acid sequence. It also simplifies the problem of understanding complex genomes by analyzing simple organisms and then applying the same principles to more complicated ones. This would result in identifying potential drug targets by checking homologies of essential microbial proteins. Bioinformatics is useful in designing drugs.

### 1.3. Algorithms in Bioinformatics

This discussion sheds light on algorithms that are of interest to biologists. The following are some of the most important algorithmic trends in bioinformatics:

- (a) Finding similarities among strings such as proteins of different organisms.
- (b) Detecting certain patterns within strings such as genes, introns, and  $\alpha$ -helices [16,20].
- (c) Finding similarities among parts of spatial structures such as motifs [18].
- (d) Constructing phylogenetic trees expressing the evolution of organisms whose DNA or proteins are currently known [21].
- (e) Classifying new data according to previously clustered sets of annotated data [22].
- (f) Reasoning about microarray data and the corresponding behavior of pathways [11].
- (g) Implementing various immune system algorithms like immune network, dendritic cell, clonal selection and negative selection algorithms [23].

## 2. USE OF SOFT COMPUTING IN BIOINFORMATICS

As soft computing [16,17,18] are considered to handle imprecision, uncertainty and near optimality in large and complex search spaces use of soft computing tools for solving bioinformatics problems have been gained the attention of researchers. Our literature survey of recent research papers (2006-2011) shows role of soft computing in modeling various aspects of bioinformatics, it involves genomic sequence, protein structure, gene expression microarray, and gene regulatory networks [17]. Most of the researches are woven around the tasks of pattern recognition and data mining like clustering, classification, feature selection, and rule generation, while classification pertains to supervised or unsupervised learning, clustering corresponds to unsupervised self-organization into homologous partitions. Feature selection techniques [24] aim at reducing the number of irrelevant and redundant variables in the dataset. Rule generation enables efficient representation of mined knowledge in human-understandable form. Many intangible parameters are mathematically modeled.

### 2.1. Soft Computing Techniques in Bioinformatics

- (a) An expert system is designed by collecting knowledge from specific experts. With the help of expert system, a biologist may decision [20].The expert system formulates the decision by rule selection and by deciding factors and by assessing a situation. This problem can be resolved by soft computing techniques. Soft computing mechanism can extract those factors and then fire rules that match the expert's behavior.

- (b) Systems often produce results different from the desired ones. This may be caused by unknown properties or functions of inputs during the design of systems [25]. This situation always occurs in the biological world because of the complexities and mysteries of life sciences. However, with its capability of dynamic improvement, soft computing can cope with this problem.
- (c) The molecular biology is ever changing, Those new data and concepts update or replace the old ones. Soft computing can be easily adapted to a changing environment. This benefits system designers, as they do not need to redesign systems whenever the environment changes [26].
- (d) Missing and noisy data is one characteristic of biological data. The conventional computer techniques fail to handle this. Soft computing based techniques are able to deal with missing and noisy data by deciding hedges in the data [19].
- (e) Soft computing is capable to find the unknown relation between huge volumes of ever evolving biological data. It is possible that important hidden relationships and correlations exist in the data. Soft computing methods are designed to handle very large data sets, and can be used to extract such relationships [27].

## 2.2. Relevance of Artificial Neural Network in Bioinformatics

An Artificial Neural Network (ANN) [4] is an information processing model that is able to capture and represent complex input-output relationships. The motivation the development of the ANN technique came from a desire for an intelligent artificial system that could process information in the same way the human brain. Its novel structure is represented as multiple layers of simple processing elements, operating in parallel to solve specific problems. ANNs resemble human brain in two respects: learning process and storing experiential knowledge. An artificial neural network learns and classifies a problem through repeated adjustments of the connecting weights between the elements.

An ANN learns from examples and generalizes the learning beyond the examples supplied. Artificial neural network [28,29] applications have recently received considerable attention. The methodology of modeling, or estimation [5,8] is somewhat comparable to statistical modeling. Neural networks should not, however, be heralded as a substitute for statistical modeling but rather as a complementary effort (without the restrictive assumption of a particular statistical model) or an alternative approach to fitting non-linear data [8,21].

Neural networks have been widely used in biology since the early 1990s. They can be used to:

- (a) Prediction and the translation sites initiation in DNA sequences and proteins [25,26].
- (b) Explain the theory of artificial neural networks using applications in biology [5].
- (c) Predict immunologically interesting peptides by combining an evolutionary algorithm [30].
- (d) Study human TAP transporter [31].
- (e) Carry out pattern classification and signal processing successfully in bioinformatics [32].
- (f) Perform protein sequence classification [6,33,34].
- (g) Predict protein secondary structure prediction [7].

- (h) Analyze the gene expression patterns as an alternative to hierarchical clusters [5,35].

In summary, a neural network is presented with a pattern on its input nodes, and the network produces an output pattern based on its learning algorithm during the training phase [36]. Once trained, the neural network can be applied to classify new input patterns. This makes neural networks suitable for the analysis of gene expression patterns, prediction of protein structure, and other related processes in bioinformatics.

### 2.3. Relevance of Fuzzy Logic in Bioinformatics

Fuzzy logic [28,29] can be easily used to implement systems ranging from simple, small or even embedded up to large networked ones. Fuzzy logic is that it accepts the uncertainties that are inherited in the realistic inputs and it deals with these uncertainties in their affect is negligible and thus resulting in a precise outputs. Fuzzy Logic reduces the design steps and simplifies complexity that might arise since the first step is to understand and characterize the system behavior by using knowledge and experience.

The concept of Fuzzy Logic (FL) was conceived by Lotfi Zadeh, FL provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information. It mimics human control logic [37].

Fuzzy systems have been successfully applied to several areas in practice like for building knowledge-based systems, fuzzy logic-based and fuzzy rule-based models. They can control and analyze processes and diagnose and make decisions in biomedical sciences. There are many application areas in biomedical science and bioinformatics, where fuzzy logic techniques [9,38] can be applied successfully. Some of the important uses of fuzzy logic are listed below:

- (a) Increasing flexibility of protein motifs [37].
- (b) Studying differences between various polynucleotides [26].
- (c) Analyzing experimental expression data [6] using fuzzy adaptive resonance theory [37].
- (d) Studying aligning sequences based on a fuzzy dynamic programming algorithm [7,4].
- (e) Mathematical modeling of complex traits influenced by genes with fuzzy-valued in pedigreed populations.
- (f) Finding cluster membership values to genes applying a fuzzy partitioning method using fuzzy C-Means and fuzzy c-hard mean algorithms [37].
- (g) Generating DNA sequencing using genetic fuzzy and neuro-fuzzy systems by anticipating disturbances due to intangible parameters [19].
- (h) Identifying the cluster genes from micro-array data [8].
- (i) Predicting protein's sub-cellular locations fuzzy k- nearest neighbors algorithm.
- (j) Mapping specific sequence patterns to putative functional classes since evolutionary comparison leads to functional characterization of hypothetical proteins.
- (k) Developing gene expression data [21].

- (l) Establishing and classification relationships between proteins [39] via fuzzy alignment methods or using a generalized radial basis function neural network architecture that generates fuzzy classification rules [21].
- (m) Analyzing the relationships between genes and decipher a genetic network.
- (n) Processing complementary deoxyribonucleic acid (cDNA) micro-array images. The procedure should be automated due to the large number of spots and it is achieved using a fuzzy vector filtering framework [11].
- (o) Classifying amino acid sequences into different super families [22].
- (p) Predicting occurrence of Diseases like cardi-vascular, kidney and cancer diseases [9,38,40-44].

#### **2.4. Genetic Algorithms in Bioinformatics**

Genetic algorithms [1,16] (GA), are randomized search and optimization techniques guided by the principles of evolution and natural genetics [45]. The applications of GAs are for solving certain multi objective problems of bioinformatics, which yields optimization of computation requirements, and robust, fast and close approximate solutions. Moreover, the errors generated in experiments with bioinformatics data can be handled with the robust characteristics of GAs. To some extent, such errors may be regarded as contributing to genetic diversity, a desirable property. The problem of integrating GAs and bioinformatics constitutes a new research area.

GAs [30] are executed iteratively on coded solutions (population) biological basic operators: selection/reproduction, crossover, and mutation. They use objective function information and probabilistic transition rules for moving to the next iteration. Of all the evolutionarily inspired approaches, Gas seem particularly suited to implementation using DNA, protein [46,47] and other bioinformatics tasks. This is because GAs are generally based on manipulating populations of bit-strings using both crossover and point-wise mutation.

The most suitable applications of GAs in bioinformatics are:

- (a) Alignment and comparison of DNA, RNA, and protein sequences [4,26,10].
- (b) Gene mappings in chromosomes [15].
- (c) RNA structure prediction [30].
- (d) Protein structure prediction and clustering [48].
- (e) Molecular design and molecular docking [14].
- (f) Gene finding and promoter identification from DNA sequences [45].
- (g) Interpretation of gene expression and micro array data [15].
- (h) Gene regulatory network identification [15].
- (i) Construction of phylogenetic tree for studying evolutionary relationship [13].
- (j) DNA structure prediction [13].

#### **2.5. Swarm Intelligence in Bioinformatics**

Historically, the phrase Swarm Intelligence (SI) was coined by Beny & Wang in late 1980's [48] in the context of cellular robotics. SI systems are typically made up of a population of simple agents (an entity capable of performing/executing certain operations) interacting locally with one another and with their environment. Although

there is normally no centralized control structure dictating how individual agents should behave, local interactions between such agents often lead to the emergence of global behavior [27]. Many biological creatures such as fish schools and bird flocks clearly display structural order, with the behavior of the organisms so integrated that even though they may change shape and direction, they appear to move as a single coherent entity. The main properties of the collective behavior can be given below:

Individuals attempt to maintain a minimum distance between themselves and others at all times. This rule is given the highest priority and corresponds to a frequently observed behavior of animals in nature. If individuals are not performing, an avoidance maneuver they tend to be attracted towards other individuals (to avoid being isolated) and to align themselves with neighbors. The structure and function of fish schools identified four collective dynamical behaviors as [49]:

- **Swarm**: an aggregate with cohesion, but a low level of polarization (parallel alignment) among members.
- **Torus**: individuals perpetually rotate around an empty core (milling). The direction of rotation is random.
- **Dynamic parallel group**: the individuals are polarized and move as a coherent group, but individuals can move throughout the group and density and group form can fluctuate.
- **Highly parallel group**: much more static in terms of exchange of spatial positions within the group than the dynamic parallel group and the variation in density and form is minimal.

A swarm can be viewed as a group of agents cooperating to achieve some purposeful behavior and achieve some goal. This collective intelligence seems to emerge from what are often large groups. According to Milonas, five basic principles define the swarm intelligence paradigm, phase transitions, and collective intelligence [1,18]:

- (a) The proximity principle: the swarm should be able to carry out simple space and time computations.
- (b) The quality principle: the swarm should be able to respond to quality factors in the environment.
- (c) The principle of diverse response: the swarm should not commit its activities along excessively narrow channels.
- (d) The principle of stability: the swarm should not change its mode of behavior every time the environment changes [49].
- (e) The principle of adaptability: the swarm must be able to change behavior more when it is worth the computational price.

## 2.6. Application of Swarm Intelligence in Bioinformatics

Swarm Intelligence algorithms [27,48,49] are mainly stochastic search and optimization techniques, guided by the principles of collective behavior and self organization of insect swarms. They are efficient, adaptive and robust search methods producing near optimal solutions and have a large amount of implicit parallelism.

The main properties of the collective behavior of swarms can be given below:

- **Homogeneity:** every bird in flock has the same behavioral model. The flock moves without a leader, even though temporary leaders seem to appear.
- **Locality:** its nearest flock-mates only influence the motion of each bird. Vision is considered to be the most important senses for flock organization.
- **Collision avoidance:** avoid colliding with nearby flock mates.
- **Velocity matching:** attempt to match velocity with nearby flock mates.
- **Flock centering:** attempt to stay close to nearby flock mates.

On the other hand, several tasks in bioinformatics involve optimization of different criteria (like energy, alignment score, overlap strength and so on):

- (a) Most of the ordering problems in bioinformatics, such as the sequence alignment, fragment assembly problem (FAP) and gene mapping (GM) [46,47,50], are quite similar to the TSP (one of the most difficult ordering problems till date) with notable differences. Thus, ACO can be tried on many of these problems and the results can be compared with the classical methods used in these contexts up to now.
- (b) To develop algorithms and mathematical models for probing the relationships among the members of a large biological dataset [51].
- (c) To analyze and interpret the heterogeneous kind of data including nucleotide and amino acid sequences, protein domains and protein structures [52].
- (d) To implement tools that enable efficient storage, retrieval and management of high-volume biological databases [52].
- (e) Hierarchical classification of protein function with ensembles of rules and particle swarm optimization [51].
- (f) Protein Structure Comparison [51].
- (g) Metabolic Pathway Analysis [27].
- (h) Patterns and Databases [49,50].

### 3. CONCLUSION

With an explosive growth of the annotated genomic sequences in available form, bioinformatics has emerged as a challenging and fascinating field of science. It presents the perfect harmony of statistics, biology and computational intelligence methods for analyzing and processing biological information in the form of gene, DNA, RNA and proteins. Soft computing algorithms on the other hand, have recently gained wide popularity among the researchers, for their amazing ability in finding near optimal solutions to a number of NP hard, real world search problems. A survey of the bioinformatics literature reveals that the field has a plethora of problems that need fast and robust search mechanisms. Problems belonging to this category include (but are not limited to) the multiple sequence alignment (MSA), protein secondary and tertiary structure prediction, protein ligand docking, promoter identification and the reconstruction of evolutionary trees. Classical deterministic search algorithms and the derivative based optimization techniques are of no use for them as the search space may be enormously large and discontinuous at several points.

### RESOURCES

European Bioinformatics Institute: [www.ebi.ac.uk](http://www.ebi.ac.uk)



National Center for Biotechnology Information: [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)  
Protein Data Bank: [www.rcsb.org/pdb](http://www.rcsb.org/pdb)  
Swiss-Prot Database: [www.expasy.ch/sprot/sprot-top.html](http://www.expasy.ch/sprot/sprot-top.html)  
CATH Database of Folds: [www.biochem.ucl.ac.uk/bsm/cath](http://www.biochem.ucl.ac.uk/bsm/cath)  
SCOP Database: [scop.mrc-lmb.cam.ac.uk/scop](http://scop.mrc-lmb.cam.ac.uk/scop)  
DALI: [www2.ebi.ac.uk/dali](http://www2.ebi.ac.uk/dali)

## REFERENCES

- [1] Sushmita Mitra and Yoichi Hayashi; "Bioinformatics with Soft Computing", IEEE transactions on systems, man, and cybernetics—part c: Applications and Reviews, Vol. 36(5), September 2006.
- [2] S. Mitra and T. Acharya; "Data Mining: Multimedia, Soft Computing, and Bioinformatics", New York: Wiley, 2003.
- [3] J.D. Szustakowski and Z. Weng; "Protein structure alignment using a genetic algorithm", Proteins, Vol. 38(4), pp. 428-440, 2000.
- [4] N. Qian and T.J. Sejnowski; "Predicting the secondary structure of globular proteins using neural network models", J. Mol. Biol., Vol. 202(4), pp. 865-884, 1988.
- [5] D. Wang and G. B. Huang; "Protein sequence classification using extreme learning machine", Proc. Int. Joint Conf. Neural Networks (IJCNN'05), Montreal, QC, Canada, pp. 1406-1411, August 2005.
- [6] H. Saigo, J.P. Vert, N. Ueda, and T. Akutsu; "Protein homology detection using string alignment kernels," Bioinformatics, Vol. 20, pp. 1682-1689, 2004.
- [7] M. Xiong, J. Li, and X. Fang; "Identification of genetic networks," Genetics, Vol. 166, pp. 1037-1052, 2004.
- [8] C.I. Branden and J. Tooze; "Introduction to Protein Structure", Garland Publishing, New York, 2nd edition, 1999.
- [9] L. Liu, Y. Fang, M. Li and C. Wang; "Prediction of beta-turn in protein using E-SSpred and support vector machine", Protein J., Vol. 28, pp. 175-181, 2009.
- [10] Zhong Wei, Altun Gulsah, Tian Xinmin, Harrison Robert, Tai Phang and Pan Yi; "Parallel protein secondary structure prediction schemes using Pthread and Open MP over hyper-threading technology", The Journal of Supercomputing, Vol. 41(1), pp. 1-16, 2007.
- [11] A.G. Hatzigeorgiou and M. Reckzo; "Signal peptide prediction on DNA sequences with artificial neural networks", Biomedical Circuits and Systems, 2004.
- [12] S. Brunak, J. Engelbrecht, and S. Knudsen; "Prediction of human mRNA donor and acceptor sites from the DNAsequence", J. Mol. Biol., Vol. 220, pp. 49-65, 1991.
- [13] B. Qian, S. Raman, R. Das, P. Bradley, A.J. McCoy, R.J. Read and D. Baker; "High-resolution structure prediction and the crystallographic phase problem", Nature, Vol. 450, pp. 259-264, 2007.
- [14] "Special Issue on Bioinformatics Part I: Advances and Challenges", Proc. IEEE, Vol. 90(11), November 2002.
- [15] G. Fogel and D. Corne (eds.); "Evolutionary Computation in Bioinformatics", San Francisco, CA: Morgan Kaufmann, 2002.

- 
- [16] Aboul Ella Hassanien, Mariofanna G. Milanova and Tomasz G. Smolinski; "Computational Intelligence in Solving Bioinformatics Problems: Reviews, Perspectives, and Challenges", *Comp. Intel. in Biomed. & Bioinform.*, SCI 151, pp. 3-47, Springerlink 2008.
- [17] P. Baldi and S. Brunak; "Bioinformatics: The Machine Learning Approach", Cambridge, MA: MIT Press, 1998.
- [18] Rabindra Ku. Jena, Musbah M. Aqel, Pankaj Srivastava and Prabhat K. Mahanti; "Soft Computing Methodologies in Bioinformatics", *European Journal of Scientific Research*, Vol.26(2), pp. 189-203, 2009.
- [19] J. Cheng, P. Baldi; "Improved residue contact prediction using support vector machines and a large feature set", *BMC Bioinformatics*, Vol. 8, pp. 113, 2007.
- [20] Juan R. González, David A. Pelta, and José L. Verdegay; "Solving Bioinformatics Problems by Soft Computing Techniques: Protein Structure Comparison as Example", *Intel. Sys. and Tech.*, SCI 217, pp. 123-136, Springer-Verlag Berlin Heidelberg 2009.
- [21] A. Dal Palu, A. Dovier, and F. Fogolari; "Constraint logic programming approach to protein structure prediction", *BMC Bioinformatics*, Vol. 5, pp. 186, 30 November 2004.
- [22] Special Issue on Bioinformatics, *IEEE Computer*, Vol. 35, July 2002.
- [23] D. Dasgupta (Editor); "Artificial Immune Systems and Their Applications", Springer-Verlag, Inc. Berlin, January 1999.
- [24] J. Setubal and J. Meidanis; "Introduction to Computational Molecular Biology", Boston, MA: Thomson, 1999.
- [25] A. Narayanan, E. Keedwell, and B. Olsson; "Artificial Intelligence Techniques for Bioinformatics", *Applied Bioinformatics*, Vol. 1(4), pp. 191-222, 2003.
- [26] Y. Huang and Y. Li; "Prediction of protein subcellular locations using fuzzy k-NN method", *Bioinformatics*, Vol. 20(1), pp. 21-28, 2004.
- [27] Swagatam Das, Ajith Abraham, and Amit Konar; "Swarm Intelligence Algorithms in Bioinformatics", (SCI) 94, pp. 113-147, Springerlink., 2008.
- [28] V. Kecman; "Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models. Complex Adaptive Systems", Cambridge, MA: MIT Press, 2001.
- [29] S. K. Pal and S. Mitra; "Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing", New York: Wiley, 1999.
- [30] Zou Xiu-fen, Pan Zi-shu, Kang Le-shan and Zhang Chu-yu; "Evolutionary computation techniques for Protein structure prediction: A Survey", *Wuhan University Journal of Natural Sciences*, Vol. 8(1B), 2003.
- [31] E.E. Snyder and G.D. Stormo; "Identification of protein coding regions in genomic DNA", *J. Mol. Biol.*, Vol. 248, pp. 1-18, 1995.
- [32] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman; "Gapped BLAST and PSI-BLAST: A new generation of protein database search programs", *Nucleic. Acids Res.*, Vol. 25, pp. 3389-3402, 1997.
- [33] J. Moult, K. Fidelis, A. Kryshtafovych, B. Rost, T. Hubbard and A. Tramontano; "Critical assessment of methods of protein structure prediction - Round VII", *Proteins*, Vol. 69(8), pp. 3-9, 2007.
- [34] The UniProt Consortium; "The Universal Protein Resource (UniProt)", *Nucleic. Acids Res.*, Vol. 35(D), pp. 193-197, 2007.

- [35] C.H. Wu and J.W. McLarty; "Neural Networks and Genome Informatics, in *Methods in Computational Biology and Biochemistry*", Amsterdam: Elsevier, 2000.
- [36] R. Farber, A. Lapedes, and K. Sirotkin; "Determination of eukaryotic protein coding regions using neural networks and information theory", *J. Mol. Biol.*, Vol. 226, pp. 471-479, 1992.
- [37] L. A. Zadeh; "Fuzzy logic, neural networks, and soft computing", *Commun. ACM*, Vol. 37, pp. 77-84, 1994.
- [38] G. Pollastri, A.J. Martin, C. Mooney and A. Vullo; "Accurate prediction of protein secondary structure and solvent accessibility by consensus combiners of sequence and structure information", *BMC Bioinform.*, Vol. 8, pp. 201, 2007.
- [39] N. Krasnogor, W.E. Hart, J. Smith and D.A. Pelta; "Protein structure prediction with evolutionary algorithms", *Proceedings of the Genetic & Evolutionary Computing Conf.*, 1999.
- [40] V. Jayde, K. White and P. Blomfield; "Symptoms and diagnostic delay in ovarian cancer: A summary of the literature", *Contemp. Nurse.*, Vol. 34, pp. 55-65, 2009.
- [41] P.M. Das and R.C. Bast; "Jr Early detection of ovarian cancer", *Biomark Med.*, Vol. 2, pp. 291-303, 2008. doi: 10.2217/17520363.2.3.291
- [42] M. Markman; "Pharmaceutical management of ovarian cancer: current status", *Drugs*, Vol. 68, pp. 771-789, 2008. doi: 10.2165/00003495-200868060-00004
- [43] A. Jemal, R. Siegel, J. Xu and E. Ward; "Cancer statistics", *CA Cancer J. Clin.*, Vol. 60, pp. 277-300, 2010. doi: 10.3322/caac.20073
- [44] Alan P. Boyle, Sean Davis, Hennady P. Shulha, Paul Meltzer, Elliott H. Margulies, Zhiping Weng, Terrence S. Furey and Gregory E. Crawford; "High-resolution mapping and characterization of open chromatin across the genome", *Cell*, Vol. 132(2), pp. 311-322, 2008.
- [45] D.E. Goldberg; "Genetic Algorithms in Search, Optimization and Machine Learning", Reading, MA: Addison-Wesley, 1989.
- [46] N. McCarthy; "Tumor profiling: Networking, protein style. *Nature Reviews*", *Cancer*, Vol. 7, pp. 892-893, 2007.
- [47] Y. Zhang, J.S. Liu; "Bayesian inference of epistatic interactions in case-control studies", *Nat. Genet.*, Vol. 39, pp. 1167-1173, 2007.
- [48] A.P. Engelbrecht; "Fundamentals of Computational Swarm Intelligence", Wiley, 2005.
- [49] M. Carolina Hinestrosa, Kay Dickersin, Pamela Klein, Musa Mayer, Karin Noss, Dennis Slamon, George Sledge and Frances M. Visco; "Shaping the future of biomarker research in breast cancer to ensure clinical relevance", *Nature Reviews Cancer*, Vol. 7, pp. 309-315, 2007.
- [50] C. Yang, Z. He, X. Wan, Q. Yang, H. Xue and W. Yu; "SNP Harvester: a filtering-based approach for detecting epistatic interactions in genome-wide association studies", *Bioinformatics*, Vol. 25, pp. 504-511, 2009.
- [51] M.H. Aghdam, N. Ghasem-Aghaee and M.E. Basiri; "Application of Ant Colony Optimization for Feature Selection in Text Categorization", *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 2872-2878, 2008.
- [52] L. P. Wang and X. Fu; "Data Mining with Computational Intelligence", Berlin, Germany: Springer, 2005.