ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

BIOINFORMATICS: COMPUTING ROLE & RELEVANCE IN BIOTECHNOLOGY (WITH SPECIAL REFERENCE TO ARTIFICIAL NEURAL NETWORKS & GENETIC ALGORITHMS)

Dr Govind Singh[*1] & Dr Dinesh Singh[2]
[*1]Faculty of Engineering and Technology, Mahatma Gandhi Chitrakoot Gramodaya University Chitrakoot, Satna [M.P.]-485870
[2]Department of Computer Science & Application, TERI, PG College Gajipur, Uttar Pradesh

ABSTRACT

Recently the use of soft computing tools for solving bioinformatics problems have been gaining the attention of researchers because of their ability to handle imprecision, uncertainty in large and complex search spaces. The acquisition of biological data, ranging from molecular characterization and simulations (e.g. protein folding dynamics), to systems biology endeavors (e.g. whole organ simulations) all the way up to ecological observations (e.g. as to ascertain climate change's impact on the biota) is growing at unprecedented speed. The use of computational and networking resources is thus unavoidable. Bioinformatics is a promising and innovative research field in 21st century. Despite of a high number of techniques specifically dedicated to bioinformatics problems as well as many successful applicatins, we are in the beginning of a process to massively integrate the aspects and experiences in the different core subjects such as biology, medicine, computer science, engineering, chemistry, physics, and mathematics..

I. INTRODUCTION

There is an increasing technological convergence occurring between biology and computer science. Today's computer scientists are tried to apply biological concept to solve social problems using genetic algorithms, DNA chips, neural networks, etc. These words are jargon from separate disciplines merged into hybrid, compound terms. Much like with businesses, the biological research world is becoming truly inseparable from the information systems needed to support scientific research and technological development. While these developments are exciting and promising, they also provide the scientific community with serious challenges. The intersection between computer science and biology has recently been described loosely as bioinformatics **(Baxevanis and Oullette, 1998).**

Bioinformatics is the application of computer technology to the management of biological information **(Krishnan et. al., 2003)**. In other word Bioinformatics is the application of computational tools and techniques to the management and analysis of biological data. Modern biotechnology is a set of techniques that involve manipulation or change of the genetic patrimony of living organisms **(Ramani, 2002)**. The term Bioinformatics was first coined by **Paulien Hogeweg and Ben Hesper (1978)** at Utrecht University in 1978 to refer to 'the study of information processes in biotic systems' **(Hogeweg and Hesper, 1983 & 1985).**

Bioinformatics has emerged out of the inputs of specialists from several different areas such as biology, biochemistry, biophysics, molecular biology, biostatistics, and computer science. Specially designed algorithms and organized computer databases are at the core of all bioinformatics operations **(Krane & Raymer, 2003; Gibas & Jambeck, 2003)**. Algorithms, that are necessarily complex, make voluminous data easy to handle for defined purposes, in an amazingly short time, a process that is humanly impossible. The requirements of such an activity make heavy and high level demands on both the hardware and the software capabilities of computers **(Miller, 1998)**.

Instruments for biology and medical research, such as next-generation sequencers, are generating vast amounts of data so much that traditional high performance computing (HPC) servers can't keep up. As researchers and scientists demand more computational performance, computing centers are running out of the resources needed to obtain that

performance: space, electrical power and cooling. The bioinformatics industry needs a more elegant solution, a new computing technology that significantly improves high throughput processing **(Lee, 2015)**.

II. BIOINFORMATICS & COMPUTATIONAL TECHNIQUES

It was well understood that computing would play a vital role in the future progress of biological information. Access to elaborate algorithms on computers increased the awareness of more recent methodological developments in biological information. "Computational statistics is related to the advances of statistical theory and methods through the use of computational methods **(Arhipova, 2006)**. This includes both the use of computation to explore the impact of theories and methods, and development of algorithms to make these ideas available to users." **(Lauro, 1996).** Computation in statistics is based on algorithms which originate in numerical mathematics or in computer science. The core topics of numerical mathematics are numerical linear algebra and optimization techniques but practically all areas of modern numerical analysis may be useful **(Cucker, 2003)**. The group of algorithms highly relevant for computational statistics from computer science is machine learning, artificial intelligence and knowledge discovery in data bases or data mining. These developments have given rise to a new research area on the borderline between biological and computer science **(Cucker, 2003; Arhipova, 2006)**.

A variety of techniques have been developed over the years to explore for and extract biological information from large biological data sets. At the end of the 1980s a new discipline, named data mining, emerged **(Sumathi & Sivanandam, 2005)**. Data Mining is the process of extracting knowledge hidden in large volumes of raw data. Data mining automates the process of finding relationships and patterns in raw data and delivers results that can be either utilized in an automated decision support system or assessed by a human analyst **(George, 2004)**. Internet facilitates linking to current programs and initiatives utilizing the Internet to form clearing-houses and distributed networks of biological information. Some are integrated system for agricultural genome analysis, including databases, conferences, publications, courses and a particularly good plant genome online database tutorial **(Kanaujia, 2004)**. Links to many tools and programs are available from the National level Institutes for sequence analysis and molecular biology, including databases, protocols and tutorials. Many pages include links to a number of model organism databases and to a number of genetic databases **(NCBI, 2015).**

III. AIMS OF COMPUTATIONAL TECHNIQUES & BIOINFORMATICS

Bioinformatics helps in providing practical tools to explore proteins and DNA in number of ways. Bio-computing is useful in recognition techniques to detect similarity between sequences and hence to interrelate structures and functions **(Jena et. al., 2009)**. The aim of Bioinformatics is to organize data in a way that allows researchers to access existing information and to submit new entries as they are produced **(Sharma et. al., 2012)**. Computer techniques are used to develop tools and resources that aid in the analysis and management of data. This data is applied to analyze and interpret the results in a biologically meaningful manner. Its computational aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness, low solution cost, and close resemblance with human like decision-making **(Bele, 2019)**. The constituents of soft computing are fuzzy logic, artificial neural networks, evolutionary algorithms including genetic algorithms, genetic programming, evolutionary strategies, support vector machines, wavelets, rough sets, simulated annealing, swarm particle optimization, memetic algorithms, ant colony optimization etc **(Bandyopadhyay et. al., 2006).**

IV. ARTIFICIAL NEURAL NETWORK

An Artificial Neural Network (ANN) 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 **(Turkson et. al., 2016)**. Its novel structure is represented as multiple layers of simple processing elements, operating in parallel to solve specific problems. An artificial neural network learns and classifies a problem through repeated adjustments of the connecting weights between the elements **(Nahua, 2017)**. The most delicate part of neural network modeling is generalization, the development of a model that is reliable in predicting future accidents **(Jena, 2009)**. Neural networks are suitable

ISSN 2348 – 8034
Impact Factor- 5.070

for the analysis of gene expression patterns, prediction of protein structure and other related processes in bioinformatics.

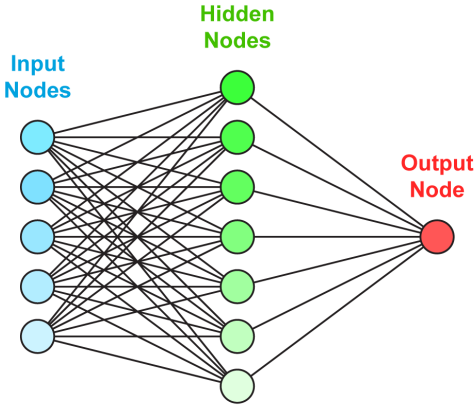

Figure 1: Neural Network Layers

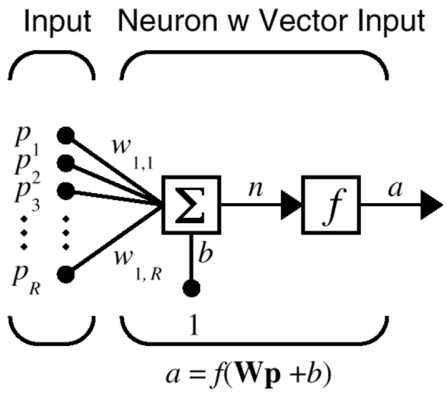

Figure 2: Mathematical Model of a Neural Network

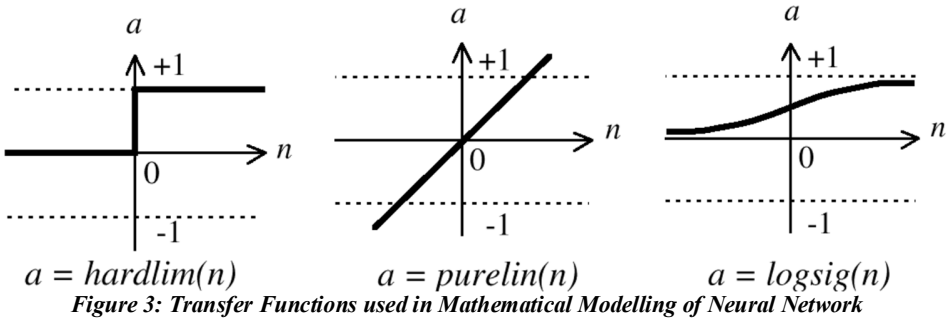

Figure 3: Transfer Functions used in Mathematical Modelling of Neural Network

Neural networks have been widely used in biology **(Baldi and Brunak, 1998)** since the early 1980s. Neural networks can be used to predict the translation initiation sites in DNA sequences **(Hatzigeorgiou and Reckzo, 2004)**. It predicts immunologically interesting peptides by combining an evolutionary algorithm **(Brusic et al., 1998)**. Neural networks carry out pattern classification and signal processing successfully in bioinformatics; in fact, a large number of applications of neural network can be found in this area. Neural networks perform protein sequence classification. Neural networks are applied to protein sequence classification by extracting features from protein data and using them in combination with the Bayesian neural network **(Wu and Mclarty, 2000)**. It analyze the gene expression patterns as an alternative to hierarchical clusters **(Toronen et al., 1999; Ma et al., 2000; Bicciato et al., 2001; Torkkola et al., 2001)**. Gene expression can even be analyzed using a single layer neural network **(Narayanan et al., 2003)**

ISSN 2348 – 8034
Impact Factor- 5.070

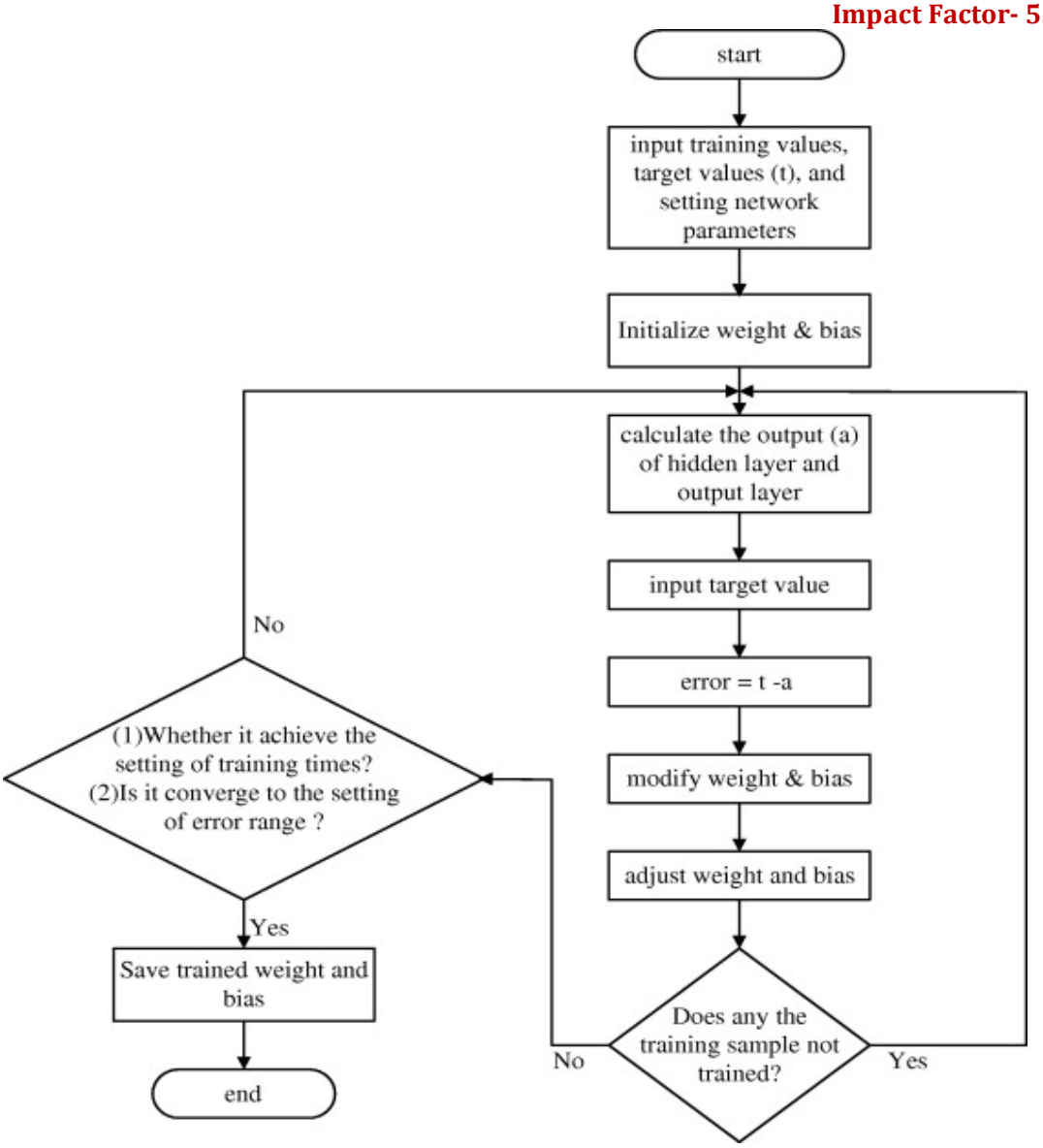

Figure 4: Flowchart of Neural Network for Biological Information Processing

Although more than a dozen NN architectures have been developed and adopted, one of the first and simplest architectures, the feedforward neural network (FNN), is the most frequently applied in protein bioinformatics **(Chen, 2012)**. Besides FNN, the recurrent neural network (RNN) and the radial basis function neural network (RBF) architectures also found several applications in the prediction of bioinformatics data **(Vanessa et. al., 2017)**. A common feature of all prediction applications in protein bioinformatics is the necessity to convert the input (biological) data into the data that can be processed by the NN. This usually involves encoding of the biological data into a fixed-size feature vector **(Seonwoo, 2017)**. NNs are used in a variety of protein bioinformatics applications. They can be categorized into:

- Prediction of protein structure including secondary structure and secondary structure content, contact maps, structural contacts, boundaries of structural domains, specific types of local structures like beta-turns, etc.
- Prediction of binding sites and ligands, which includes prediction of binding residues and prediction of various properties of the binding ligands.

ISSN 2348 – 8034
Impact Factor- 5.070

- Prediction of protein properties such as physicochemical proteins, localization in the host organism, etc.

The most popular architecture used in these methods is a simple three-layer feedforward NN, although other architectures such as RBF and recurrent NNs are also applied **(Hill, 2018)**. Some of the protein bioinformatics applications use multilayered designs in which two (or more) NNs are used in tandem. We show that the popularity of the NN-based designs has been growing over the last decade **(Hang, 2018)**. Three applications that enjoy the most widespread use are discussed in greater detail. They include protein secondary structure prediction, prediction of binding sites, and prediction of relative solvent accessibility **(Zang, 2018)**. The last two decades observed the development of several methods based on NN for the prediction of protein secondary structure. The performance of the methods mainly depends on the representation of the protein sequence and the size of the training dataset **(Iqbal, 2014)**. Since the beta-sheets (strands adjacent in the tertiary structure) are established that are far away in the sequence, the window-based methods (including all present methods for the prediction of protein secondary structure) are inherently incapable of grasping the long-range interactions, which results in a relatively poor result for strands **(Zang, 2018)**.

V. GENETIC ALGORITHMS

Genetic algorithms **(Goldberg, 1989; Bhandari et al., 1996; Booker et al., 1989; Mitchell et al., 1992)**, a biologically inspired technology, are randomized search and optimization techniques guided by the principles of evolution and natural genetics. They are efficient, adaptive, and robust search processes, producing near optimal solutions, and have a large degree of implicit parallelism **(Kumar. 2014)**. Therefore, the application of Genetic Algorithms for solving certain problems of bioinformatics, which need optimization of computation requirements, and robust, fast and close approximate solutions, appears to be appropriate and natural **(Prithivi, 2012)**. The errors generated in experiments with bioinformatics data can be handled with the robust characteristics of Genetic Algorithms.

Genetic Algorithms are executed iteratively on a set of coded solutions, called population, with three basic operators: selection/reproduction, crossover, and mutation **(Chenand, 2007)**. They use only the payoff (objective function) information and probabilistic transition rules for moving to the next iteration. Genetic Algorithms seem particularly suited to implementation using DNA, protein, and other bioinformatics tasks **(Needleman and Wunsch, 1970)**. Genetic Algorithms can process, in parallel, populations billions times larger than is usual for conventional computation. The usual expectation is that larger populations can sustain larger ranges of genetic variation, and thus can generate high-fitness individuals in fewer generations. Problems of bioinformatics seldom need the exact optimum solution; rather, they require robust, fast, and close approximate solutions, which genetic algorithms are known to provide efficiently **(Ye et. al., 2008)**. Problems of bioinformatics seldom need the exact optimum solution; rather, they require robust, fast, and close approximate solutions, which Genetic Algorithms are known to provide efficiently.

ISSN 2348 – 8034
Impact Factor- 5.070

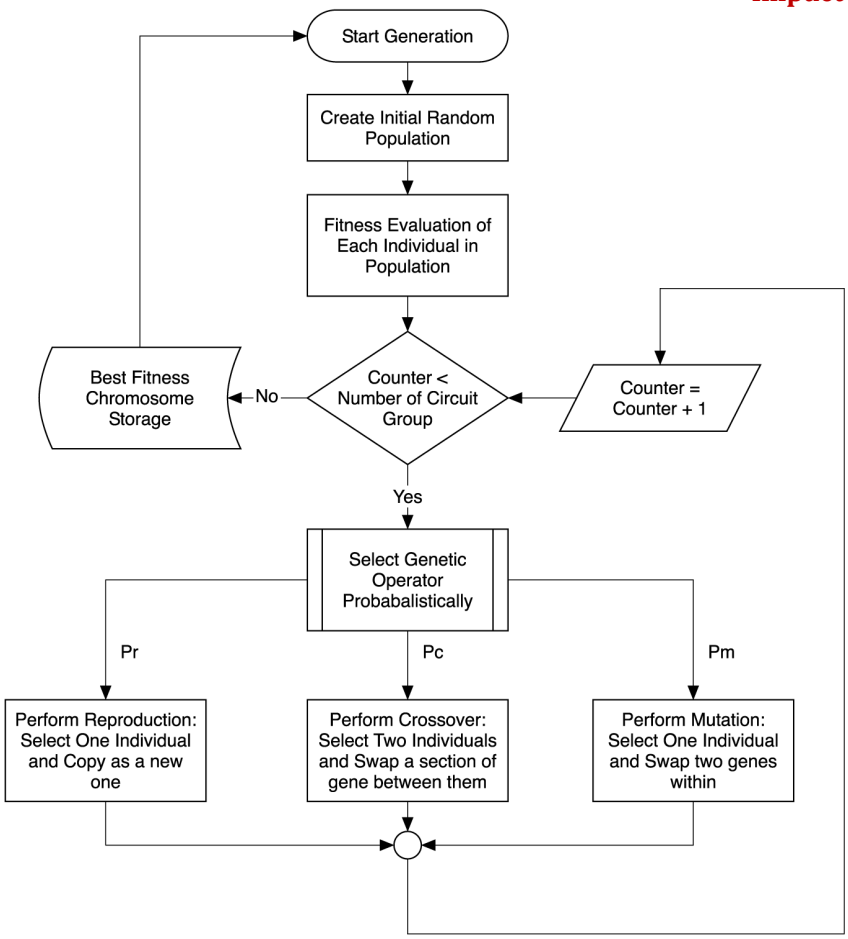

Figure 5: Flowchart of Genetic Algorithms for Biological Information Processing

The field of Bioinformatics has advanced rapidly in recent years. Breakthroughs in sequencing technology have driven an explosion in the production of genetic data. As a result, the development of methods to interpret this data accurately and in a timely manner has become a major bottleneck in bioinformatics research **(Yang, 2016)**. This task is further complicated by imperfect knowledge of genetic features: performance-altering assumptions can easily be introduced during algorithm development, and methodology may quickly become obsolete. Genetic algorithms are an evolution-inspired class of machine learning algorithms that show great promise to resolve these problems **(Manzoni, 2018)**. These algorithms gradually refine solutions through natural selection, evolving a solution to a problem in bioinformatics rather than manually designing a search strategy. Due to this learning process determining how features are identified, genetic algorithms do not rely on human knowledge of the problem **(Jiawei. 2012; Graham, 1994)**. Consequently, biases are largely limited to the data used to determine how fitness is evaluated. Genetic algorithms also make better use of computational resources by reducing search space and utilizing parallel computation **(Mehboob, 2016)**. In order to examine this potential, this work explores many different implementations of genetic algorithms in bioinformatics and their results **(Zhu, 2013)**.

Protein structure prediction is the process of prediction of the three dimensional structure of a protein from its amino acid sequence. Proteins are large biological molecules which contain large amount of amino acid sequence **(Deng, 2018)**. The Bioinformatics industry is in the fledgling condition and gaining more attention of researchers. For the development of a new drug there need a high level of research with large amount of data from different cluster and locations **(Rout et. al., 2013)**. In the field of biology protein structure prediction plays a vital role for the development

a new drug. In recent years many techniques are being used for the protein structure prediction **(Sahu, 2015)**. There are many Soft Computing methods like Fuzzy Logic, Artificial neural network, Genetic algorithms, Swarm optimization, etc are used for this purpose to distinguish, compare or process the various structure of protein **(Timmy, 2013).** It is always a big task for researchers to develop new tools and methods for the purpose of processing of data as well as development of drugs. Protein Structure Prediction is the process of prediction of the secondary, territory & quaternary structure of the protein from its amino acid sequence.

Genetic Algorithm is a computational technique which creates the mimic of the process. In this paper we have proposed a technique, which based on Genetic Algorithm technique for the prediction of protein structure **(Álvarez et. al., 2012)**. This technique will be helpful to work with huge amount of data and for the prediction of protein structure in a large scale. This technique can be adopted by the medicine researchers to develop various drugs after study and analyzing the changes of protein structure **(Tian et. al., 2012)**. The field of Bioinformatics has advanced rapidly in recent years. Breakthroughs in sequencing technology have driven an explosion in the production of genetic data **(Rodger & Rayston, 2005**). As a result, the development of methods to interpret this data accurately and in a timely manner has become a major bottleneck in bioinformatics research. This task is further complicated by imperfect knowledge of genetic features: performance-altering assumptions can easily be introduced during algorithm development, and methodology may quickly become obsolete **(Shan, 2014)**. Genetic algorithms are an evolution-inspired class of machine learning algorithms that show great promise to resolve these problems. These algorithms gradually refine solutions through natural selection, evolving a solution to a problem in bioinformatics rather than manually designing a search strategy **(Goldberg & Holland, 1988)**.

Due to this learning process determining how features are identified, genetic algorithms do not rely on human knowledge of the problem. Consequently, biases are largely limited to the data used to determine how fitness is evaluated **(Liu et. al., 2012)**. Genetic algorithms also make better use of computational resources by reducing search space and utilizing parallel computation **(Mahdevar, 2010)**. In order to examine this potential, this work explores many different implementations of genetic algorithms in bioinformatics and their results.

VI. CONCLUSION

Bioinformatics is an essential component of modern biotechnology. Bioinformatics is a field that combines cutting edge discoveries and innovation in the field of molecular biology and computer science. The level of innovations that happen in this hybrid area is very important for advanced development in biological sciences. Hybrid research aims at computational modeling of biological phenomena and applies techniques from areas such as artificial intelligence, databases, software engineering, theoretical computer science, discrete mathematics, optimization theory, control theory and modeling. The idea of jointing computer science and biological research should be implemented in the field of information technologies in bio systems. Soft computing, in the field of neural networks and genetic algorithms, is a consortium of methodologies that work synergistically and provides, in one form or another, flexible information processing capabilities for handling real life ambiguous situations. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness, low solution cost and close resemblance with human like decision-making.

REFERNCES

1. Álvarez D, Hornero R, Marcos JV, del Campo F.,(2012). Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis. Med Eng Phys 2012; Vol.34, pp.1049 - 1057
2. Arhipova, Irina (2006). The Role of Statistical Methods in Computer Science & Bioinformatics. ICOTS-7. Arhipova, 2006. URL - http://iase-web.org
3. Baldi P. and Brunak S. (1998), "Bioinformatics: the Machine Learning Approach", MIT Press.
4. Bandyopadhyay Sanghamitra, Maulik Ujjwal & Wang Jason T L (2006).Analysis of biological data a soft computing approach World Scientific Publishing Co. Pte. Ltd. 5 Toh Tuck Link, Singapore 596224
5. Baxevanis A. D. & Ouellette B. F. F. (1998). Bioinformatics: a practical guide to the analysis of genes and proteins. John Wiley & Sons, New York, NY

6. Bele, S.B. (2019). The Concept Of Soft Computing, Asian Journal of Science and Technology, Vol. 10, Issue, 12, pp.10556-10558, December, 2019
7. Bhandari D., Murthy C. A., and Pal S. K. (1996). "Genetic algorithm with elitist model and its convergence", Int. J. Pattern Rcognit. Artif. Intell., Vol. 10, No. 6, pp. 731–747.
8. Bicciato S., Pandin M., Didone G., and Bello, C. (2001), "Analysis of an Associative Memory neural Network for Pattern Identification in Gene expression data", BIOKDD01, Workshop on Data mining in Bioinformatics. Soft Computing Methodologies in Bioinformatics 2001
9. Booker L. B., Goldberg D. E.and Holland J. H., (1989). "Classifier systems and genetic algorithms", Artif. Intell., Vol. 40, No. 1–3, pp. 235–282.
10. Brusic V., van Endert P., Zeleznikow J., Daniel S., Hammer J., and Petrovsky N. 1. (1999), "A neural network model approach to the study of human TAP transporter", In Silico Biol. 1: pp. 9–21
11. Chen Ke & Lukasz Kurgan (2012). Neural Networks in Bioinformatics. Handbook of Natural Computing (pp.565-583)
12. Chen Ke, Kurgan L.A., Alberta, Edmonton (2012). Neural Networks in Bioinformatics. Handbook of Natural Computing. Springer-Verlag Berlin Heidelberg, pp. 02-22
13. Chenand K., Kurgan L. (2007), "PFRES: protein fold classification by using evolutionary information and predicted secondary structure", Bioinformatics, 23(21): 2843 – 2850
14. Cucker, F. (2003). Foundations of Computational Mathematics: Special Volume. Handbook of Numerical Analysis. North-Holland Publishing. ISBN 978-0-444-51247-5.
15. Deng Haiyou, Ya Jia and Yang Zhang (2018). Protein structure prediction. Int J Mod Phys B. 2018 Jul 20; 32(18): 1840009.
16. George Fernandez (2004). Conference on Applied Statistics in Agriculture. https://doi.org/10.4148/2475-7772.1148
17. Gibas & Jambeck (2003). Developing Bioinformatics Computer Skills, O'Reilly, 2003
18. Goldberg D., (1989). "Genetic Algorithms in Optimization, Search, and Machine Learning", Reading, MA: Addison-Wesley
19. Goldberg D.E. and Holland J.H.(1988). "Genetic algorithms and machine learning," Machine Learning, vol. 3, no. 2, pp. 95–99, 1988.
20. Graham F. Spencer (1994). Automatic generation of programs for crawling and walking. In Kenneth E. Kinnear, Jr., editor, Advances in Genetic Programming, chapter 15, pages 335–353. MIT Press
21. Hang Li, Xiu-Jun Gong, Hua Yu and Chang Zhou (2018). Deep Neural Network Based Predictions of Protein Interactions Using Primary Sequences. Molecules. 23(8): 1923.
22. Hatzigeorgiou, A. G. Reckzo, M. (2004), "Signal peptide prediction on DNA sequences with artificial neural networks", Biomedical Circuits and Systems
23. Hatzigeorgiou, A. G. Reckzo, M. (2004), "Signal peptide prediction on DNA sequences with artificial neural networks", Biomedical Circuits and Systems
24. Hogeweg, P. and Hesper B., (1983). The ontogeny of the interaction structure in BumbleBee colonies: a MIRROR model. Behav Ecol Sociobiol 12: 271-283.
25. Hogeweg, P. and Hesper B., (1985). Socio-informatic processes, a MIRROR modelling methodology. J Theor Biol 113: 311-330;
26. Iqbal Muhammad Javed, Faye Ibrahima, Samir B.B. and Abas Md Said (2014). Efficient Feature Selection and Classification of Protein Sequence Data in Bioinformatics. The Scientific World Journal. https://doi.org/10.1155/2014/173869
27. Jiawei Han, Micheline Kamber, Jian Pei (2012). Data Mining: Concepts and Techniques. Solution Manual, The University of Illinois at Urbana-Champaign Simon Fraser University
28. Kang Nahua (2017). Multi-Layer Neural Networks with Sigmoid Function— Deep Learning for Rookies. Towards Data Science. https://towardsdatascience.com/multi-layer-neural-networks-with-sigmoid-function-deep-learning-for-rookies-2-bf464f09eb7f
29. Krane S.E & Raymer M. (2003). Fundamental Concepts of Bioinformatics, Pearson, 2003
30. Krishnan R.T., Gupta A., Matta V., (2003). Biotechnology & Bioinformatics: Can India Emulate the Software Success Story? Paper to be presented at NSF-sponsored Workshop on The Indian Development

Experience organized by School of Public Policy, George Mason University and Department of Management Studies, Indian Institute of Science At Bangalore On March 3-5, 2003
31. Kumar Sarvesh, Singh R.P., Mishra A. & Kumar Hemant (2014). Computation of Neural Network using C# with Respect to Bioinformatics. International Journal of Scientific and Research Publications, Volume 3, Issue 9, pp 01-14
32. Lauro, C. (1996). Computational statistics or statistical computing, is that the question? Computational Statistics and Data Analysis, 23, 191–193
33. Lee, Frank (2015). IBM Reference Architecture for Genomics Speed, Scale, Smarts (RedBooks), Dept. HYTD Mail Station P099, 2455 South Road, Poughkeepsie, NY 12601-5400 U.S.A. URL : http://www.redbooks.ibm.com/redpapers/pdfs/redp5210.pdf
34. Liu K.Q., Liu Z.P., J. K. Hao et al (2012). "Identifying dysregulated pathways in cancers from pathway interaction networks," BMC Bioinformatics, vol. 13, article 126, 2012.
35. Mahdevar G., Zahiri J., Sadeghi M., Nowzari-Dalini A., and Ahrabian H., (2010). "Tag SNP selection via a genetic algorithm," Journal of Biomedical Informatics, vol. 43, no. 5, pp. 800–804, 2010.
36. Manzoni C., Kia D. A., Vandrovcova J., Hardy J., Wood N., Lewis P., and Ferrari R. (2018). Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences, Briefings in Bioinformatics, 19-2:286-302.
37. Mehboob Usama, Qadir Junaid, Salman Ali & Athanasios Vasilakos (2016). Genetic algorithms in wireless networking: techniques, applications, and issues. Soft Computing volume 20, pages2467–2501. https://link.springer.com/article/10.1007/s00500-016-2070-9
38. Miller Riel, Wolfgang Michalski and Barrie Stevens (1998). 21st Century Technologies, Promises and Perils of a Dynamic Future. OECD Publications, 2, rue Andr´e-Pascal, 75775 Paris Cedex 16
39. Mitchell M., Forrest S., and Holland J. H., (1992). "The royal road for genetic algorithms: Fitness landscapes and GA performance". In Proc. 1st Eur. Conf. Artificial Life Setubal J. and Meidanis J., (1999). "Introduction to Computational Molecular biology", Boston, MA: Thomson
40. Narayanan A., Keedwell E., and Olsson B., (2003). "Artificial Intelligence Techniques for Bioinformatics", Applied Bioinformatics, Vol.1, No. 4, pp. 191-222
41. NCBI (2015). Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, Volume 44, Issue D1, 4 January 2016, Pages D7–D19, https://doi.org/10.1093/nar/gkv1290
42. Needleman S. B. and Wunsch C. D., (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins", J. Mol. Biol., Vol. 48, pp. 443–453
43. Prithivi P., (2012). Soft Computing Methodologies in Bioinformatics. National Conference on Future Computing Volume 1, March 2012
44. Rabindra Ku. Jena, Musbah M. Aqel, Pankaj Srivastava & Prabhat K. Mahanti (2009). Soft Computing Methodologies in Bioinformatics, European Journal of Scientific Research, Vol.26 No.2 (2009), pp.189-203
45. Ramani Shyama V. (2002). "Who is interested in biotech? R&D strategies, knowledge base and market sales of Indian biopharmaceutical firms," Research Policy, Vol. 31, No. 3, March, pp. 381-398.
46. Roger M. Jarvis, Royston Goodacre (2005). Genetic algorithm optimization for pre-processing and variable selection of spectroscopic data. Bioinformatics, Volume 21, Issue 7, Pages 860–868
47. Rout S. Bhusan, Dehury Satchidananda, Mishra Bhabani Sankar Prasad (2013). Protein Structure Prediction using Genetic Algorithm. IJCSMC, Vol. 2, Issue. 6, June 2013, pg.187 – 192
48. Sahu D.P., Sajid M. and Prakash Shiv (2015). A Note on Bioinformatics using Genetic Algorithms. International Journal of Computer Applications (0975 – 8887), Volume 117 – No. 1, May 2015. https://research.ijcaonline.org/volume117/number1/pxc3902315.pdf
49. Seonwoo Min, Byunghan Lee and Sungroh Yoon (2017). Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, Korea. https://arxiv.org/ftp/arxiv/papers/1603/1603.06430.pdf
50. Shan Li, Kang Liying and Zhao Xing-Ming (2014). A Survey on Evolutionary Algorithm Based Hybrid Intelligence in Bioinformatics. Data Mining in Translational Bioinformatics, Vol. 2014, https://www.hindawi.com/journals/bmri/2014/362738/
51. Sharma M.K, Dhar M. K., Kaul Sanjana (2012). Bioinformatics: An introduction and Overview. International Journal of Engineering Research and Development, Vol 3, No 8, PP. 88-99.

52. Shiva Kanaujia (2004). Bioinformatics & Internet: New Paradigm To Disciplines And Information Technology. Srels Journal of Information Management, Vol. 41, No. 1, March 2004, Paper E. p43-55. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.453.1129&rep=rep1&type=pdf
53. Steven T Hill, Rachael Kuintzle, Amy Teegarden, Erich Merrill, III, Padideh Danaee and David A Hendrix (2018). A deep recurrent neural network discovers complex biological rules to decipher RNA protein-coding potential.Nucleic Acids Res. 46(16): 8105–8113.
54. Steven T Hill, Rachael Kuintzle, Amy Teegarden, Erich Merrill, Padideh Danaee and David A Hendrix (2018). A deep recurrent neural network discovers complex biological rules to decipher RNA protein-coding potential. Nucleic Acids Res. 46(16): 8105–8113.
55. SumathiS,S. & Sivanandam, N.(2005). Introduction to Data Mining and its Applications. Springer Nature Switzerland AG
56. Tian H, Liu C, Gao XD, Yao WB (2012). Optimization of auto-induction medium for G-CSF production by Escherichia coli using artificial neural networks coupled with genetic algorithm. World J Microbiol Biotechnol 2012; pp. 01-09
57. Timmy Manning,Roy D Sleator & Paul Walsh (2013). The use of genetic algorithms in bioinformatics. Naturally selecting solutions. Journal Bioengineered Volume 4, 2013 - Issue 5, Pages 266-278
58. Torkkola Kari, Robert Mike Gardner, Tamma Kaysser-Kranich and Calvin Ma (2001). "Selforganizing maps in mining gene expression data", Information Sciences, 139 (1-2): 79–96
59. Turkson R.F., Yana Fuwu, Ali M K A & Jie Hu (2016). Artificial neural network applications in the calibration of spark-ignition engines: An overview. Engineering Science and Technology, an International Journal, Volume 19, Issue 3, Pages 1346-1359
60. Vanessa Isabell Jurtz, Alexander Rosenberg Johansen, Morten Nielsen, Jose Juan Almagro Armenteros, Henrik Nielsen, Casper Kaae Sønderby, Ole Winther and Søren Kaae Sønderby (2017). An introduction to deep learning on biological sequence data: examples and solutions. Bioinformatics. 33(22): 3685–3690.
61. Wu C. H. and McLarty J., (2000). "Neural Networks and Genome Informatics", Elsevier Science.
62. Yang C., Moi S., Lin Y., and Chuang L. (2016). Genetic algorithm combined with a local search method for identifying susceptibility genes. JAISCR6:203-212
63. Ye L, Liu T, Wu Z, Zhou R (2008) Sequence-based protein domain boundary prediction using BP neural network with various property profiles. Proteins 71:300–307
64. Zhang Buzhong, Jinyan Li & Qiang Lü (2018). Prediction of 8-state protein secondary structures by a novel deep learning architecture. BMC Bioinformatics volume 19, Article number: 293 (2018)
65. Zhang Buzhong, Linqing Li and Qiang Lü (2018). Protein Solvent-Accessibility Prediction by a Stacked Deep Bidirectional Recurrent Neural Network. Biomolecules. 8(2): 33.
66. Zhang C., (1994). "A genetic algorithm for molecular sequence comparison", In Proc. IEEE Int. Conf. Systems, Man, and Cybernetics, Vol. 2, pp. 1926–1931
67. Zhu N, O'Connor I (2013) iMASKO: a genetic algorithm based optimization framework for wireless sensor networks. J Sens Actuator Netw 2(4):675–699.