Biomedical Letters ISSN 2410-955X– An International Biannually Journal

Review Article
2020 | Volume 6 | issue 1 | Pages 23-31

ARTICLE INFO

Received
December 24, 2019
Revised
February 13, 2020
Accepted
March 21, 2020

***Corresponding Author**
Ayesha Ikram

E-mail
ayshik56@gmail.com

Keywords
Machine learning
Artificial Intelligence
Bioinformatics
Decision Tree
Neural Networks
Genetic Algorithm

How to Cite
Noor H, Ikram A, Khan AH, Haider A, Razzaq MS. Applications of the Artificial Intelligence and Machine learning In Computational Biology. Biomedical Letters 2020; 6(1):23-31.

[¥]Both authors contributed equally.

Open Access

Applications of the artificial intelligence and machine learning in computational biology

Hasnat Noor[1, ¥], Ayesha Ikram[1*, ¥], Ali Hassaan Khan[2], Ali Haider[1], Muhammad Sufyan Razzaq[1]

[1]Department of Bioinformatics and Biotechnology, Government College University Faisalabad, Pakistan
[2]Institute of Horticulture and Sciences, University of Agriculture, Faisalabad, Pakistan

Abstract

Artificial Intelligence (AI) is a wide-ranging branch of computer science that deals with the construction of smart machines that typically require human intelligence. Machine learning is a method of data analysis that automates the building of analytics models. It is based on the theory that the system can learn from statistics, identify patterns and make decisions with minimal human intervention. Artificial intelligence and machine learning are interconnected fields. Bioinformatics approach has been used to address numerous biological problems. Machine learning and AI are revolutionizing computational biology and bioinformatics. There has been progressive advancements in computer sciences and bioinformatics. The use of AI and machine learning in bioinformatics are helping to make new ways for biological data management and to perform different analyses for logical conclusions. This article describes the use of some key tools of applied AI and machine learning in bioinformatics.

Scan QR code to see this publication on your mobile device.

Introduction

Bioinformatics is an interdisciplinary field, promising the preeminent analysis of huge biological data with powerful computing and mathematical techniques [1]. The advancement in bioinformatics solves many complex biological problems. Bioinformatics approach has been used to address numerous biological problems. Artificial intelligence and machine learning are interconnected fields. Both of these technologies are highly trending technologies used to create intelligent systems. Artificial intelligence systems do not need to be reprogrammed, but instead, they use algorithms that can work with their own intelligence. In machine learning machines are trained by providing data to make future prediction by learning the pattern of data provided as input. Use of AI in structural bioinformatics tools is an effective way to design novel compounds against neurological disorders. Today, most of the biological work is based on the prediction of biological systems [2-5]. To solve the biological functions through bioinformatics approaches helps to treat the diseases and other biological processes [6]. The mutational analyses, protein structures and protein functions are considered as an effective area to solve through bioinformatics approaches. However, the complexity of the biological data and the amount of it can cause the management and analysis issues [7]. As the data from whole genome sequencing projects increasing and coming along with complexities, it is enhancing the need of more powerful computational techniques to be used in bioinformatics [8-10]. Artificial intelligence and machine learning techniques transforming the computational powers, as they were developed to solve problems intelligently, following the concept of machine having human intelligence. Simulation of different models, biological sequence annotations, computational drug designing, virtual screening, binding site prediction and gene prediction can be effectively predicted through the use of AI and machine learning algorithms in bioinformatics. The probabilistic powers of aforementioned techniques is of great essence in bioinformatics to solve the complex biological problems [11-17]

Decision Trees for Classification: A Machine Learning Algorithm

In bioinformatics, researchers are designing efficient algorithms to analyze gigantic amount of data [18,19]. The microarray data contains information of thousands of genes on their expression profiles in different conditions [20-24]. The expression profiles can be used to detect and track the ailments and response of patients to medication. It needs to develop a technique that can detect similar expressions and make a cluster of sick and healthy people (**Fig. 1**) [25-33]. The decisions trees are helpful in these conditions to provide suitable classification results. It is an algorithm that gets an input object and in return shows an output based on decision. It has nodes on it and each node is responsible to test the attribute of an input. The branches on the nodes show correspondence to the values which are possible for the attributes. The leaf nodes repeat the values. It is a tool that supports decision by using a tree-like model. Sorting rules are characterized in tracks from root to leaf. It has three types of nodes as, decision nodes-are usually denoted by squares, chance nodes are in circles and end nodes-by triangles [34-38].

Decision tree is basically a predictive modeling approach. This approach is used to construct a decision tree that finds different ways of splitting a data set grounded on various conditions. Decision trees where the board variable can take continuous qualities (regularly genuine numbers) are called regression trees, also known as a classification and regression tree (CART). It is one of the immensely used and applied methodologies in machine learning. Decision Trees are a non-parametric regulated learning procedure utilized for collaboration arrangement and regression processes. Decision tree gives data about gene interaction by stepwise parting of the data set – each split uncovers one quality and the progressive structure shows the nature of the interactions [39-44].

The building blocks of the decision tree includes; elements of decision tree: no sink nodes are in it however the decision tree only consists of burst nodes. Consequently, it can grow and turn enormous and that is the reason they are customarily strenuous to draw physically. The decision tree can be integrated into the principles of decision, where the result is the contents of the leaf node, and if the conditions along the path make a result in the clause. Generally, there is a form of rules; If condition 1, condition 2 and condition 3 then outcome. Decision regulations can be generated by building link regulations on the right side with the board variable. The casual or temporal relation can also be

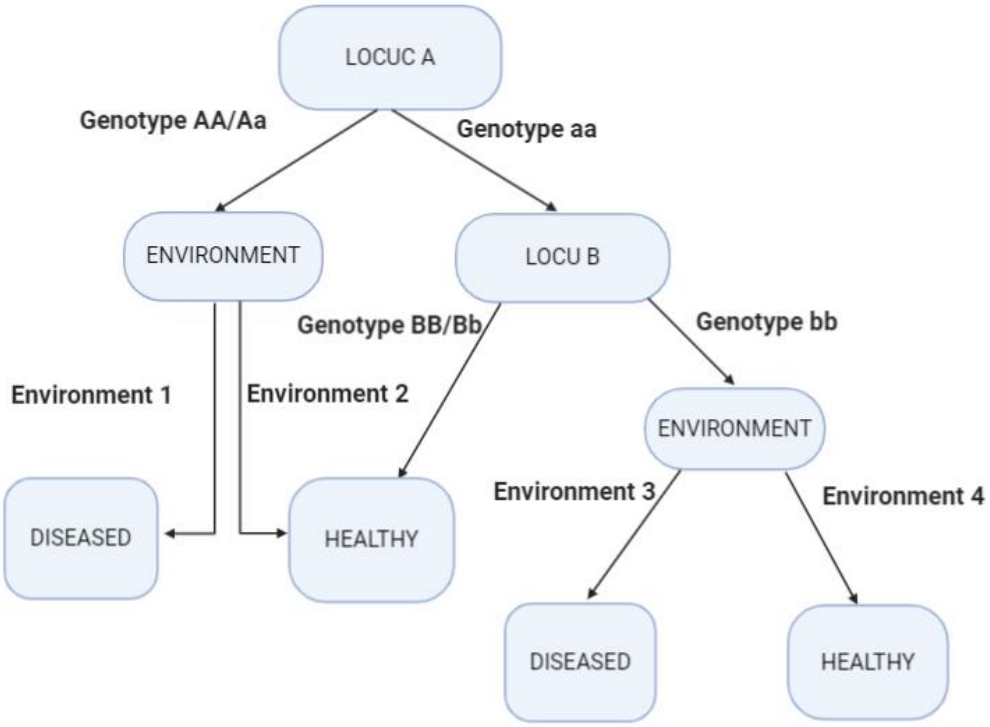

Fig. 1: A simple decision tree. Disease can be predicted from the genotype and environment, *e.g.*, an individual in environment-4 with genotype aabb will be diseased.

represented. The approach for decision tree has different types of queries at each node. The information gains correspondence to the queries to calculate. The gained information is used as an anchor to decide the feature at each step. The small tree splits at each step. The obtained information can be described as a measure generally used for purity. The particular amount of information gained for a class is measured by using the information value. The highest gained information is used to select the first split and the process is continued till the level is reached where children are pure, or the information value turns to 0 [45-49].

Pure denotes the sample chose from a data set fits to the same class while the impure data does not belong to the same class in fact it is the blend of various classes. Gini impurity is a parameter for the measurement. It is used to measure the probability of an unfitting classification of a new case variable. It depends upon the dataset if it is pure the chance of improper classification is zero. The likelihood of improper classification is higher if the sample is a blend of various classes. The basic steps of decision tree include, all the vagueness of dataset is calculated. A list is produced containing all the questions that need to inquire at the nodes. On the basis of questions inquired at each node, rows are divided as true and false. The information obtained from the division of data and gini impurity is calculated. The maximum information gained on the basis of calculations is then updated. The question on the base of information obtained is updated and the partition of the nodes has been done on the basis of questions. These steps are repeated from the beginning in a loop until the pure nodes are obtained. Decision tree is simple to understand even with a transitory insight of its model. For different situations, it aids in governing the best, worst and expected values. The stumbling blocks of decision trees inclined to over fitting. It needs some sort of measurement to keep checking the work. [50-53].

K. nearest neighbor (KNN)

The k. nearest neighbor is a simple and easy to implement supervised machine learning algorithm. It can be used to solve both classification regression analysis problems. This algorithm works to find an entity from knowing its neighbor. It is an instance-based learning. KNN has different example-based

reasoning, lazy learning and instance-based learning. In regression, KNN helps to analyses the average values. It classifies new data points by checking similarity and placing it in a group containing closest neighbors. It has powerful classification algorithm that is used in pattern recognition. This algorithm uses no assumptions, non-parametric and lazy learning algorithm. It is first trained to encounter the problem. It does not decide itself it works by pulling the stored information and using it to solve the issues whether it is classification or regression problem regarding new data point [54]. For classifying a new data point, it uses voting system for its neighbor classes. It is measured by a distant function. The new object is assigned to the group containing closest neighbors (**Fig. 2**).

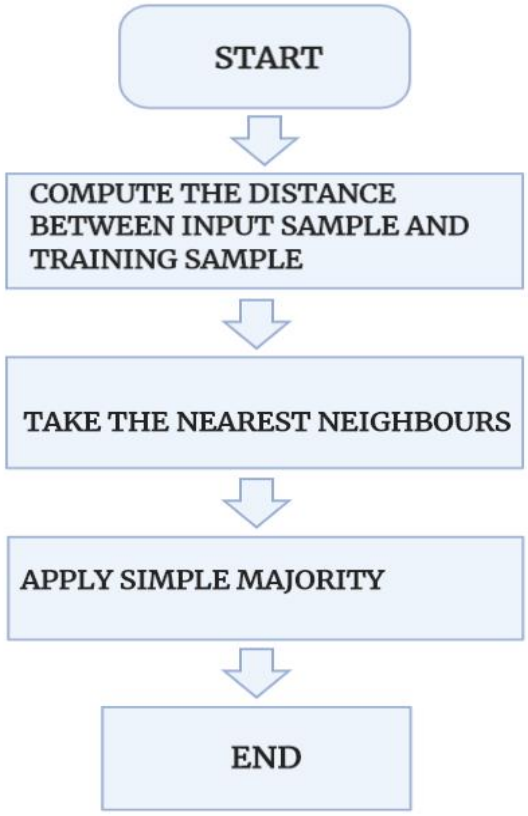

Fig. 2: KNN classifier algorithmic steps

Genetic Algorithm

The genetic algorithm is a heuristic search method used in artificial intelligence. It is used to find better solutions to search problems based on the theory of natural selection and evolutionary biology. Genetic algorithms are mostly recommended for searching through large and intricate data sets. The concept of genetic algorithm relies on the concept of evolution. It generates samples and work based on the phenomena of survival of the fittest [56]. It consistently tries to improve it as well by selecting the fit ones and rejecting the weak ones. The population of the selected ones are increased day by day while the unselected are decreased in numbers. Nature has adapted this method of experimentation for the sake of improvement. The natural method of selection is based on various bio operators including mutations and crossing over which are also used in genetic algorithm. Genetic algorithm mimics the natural process of selection of fit and generation of off springs. It also considers the fact of robustness like nature. Genetic algorithm shuffles the genes and moves from simplicities to complexities for the sake of optimization. The genetic algorithm consists of sequence of events used to find the suitable solution. The events include initialization to look for a population which can serve as initial population from which the fittest can be selected later. The purpose of selecting this parent population is to generate an offspring population. This population contains candidates which serve as variety of genetic materials. The fitness function is the selection of the fittest parent population to generate offspring population. It is done by evaluating parent population by fitness function which is similar to heuristic approach selection. The fittest population is selected to generate the offspring population based on the value of fitness function. The population is selected based on the value assigned by fitness function in order to mimic the natural phenomena of survival of the fittest. Cross overs are applied by swapping the tail of parent population. The bio operator used to swap the bits. The genetic representation in this program has array of bits. This sequence of events works like a loop. They are continued until a constant value of the generation is achieved. Once it starts getting constant values, it results in the termination of algorithm [57].

Artificial Neural Networks (ANNs)

Artificial neural networks (ANN) is the key tool of machine learning. ANNs is a representation of the human brain function in the form of computational models. ANN is a machine learning algorithm for the pattern recognition. It is composed of input layer, hidden layer and output layer. Nodes are referred as the units of information in each layer. Nodes of different layers are interconnected to form a network which resembles the natural biological system [51-

54]. The nodes are mathematical weight constraints that can be trained with known patterns and can be used later for predictions. Once the network has been trained, it is capable of identifying the interconnection between the input and the output. The process of learning in humans has some minute regulations to the synaptic networks among neurons. Contrarily, ANNs learning phase is grounded on the associations among the processing elements that set up the network topology. ANNs have become one of the requisite gears in bioinformatics. It was fueled through the expansion and shoot up of many biological databases, which store the data correlating to RNA and DNA sequences some proteins and other macromolecular structures. Massive amount of data requires the use of computational tools to deal with its complexity to retrieve information of interest [58-61]. The diverse uses of neural networks include predictive modelling, classification and also for the identification of biomarker within datasets of high complexity and desire. The three primary forms of ANNs are radial basis function networks, the multilayer perception and recurrent neural networks [62-64].

Multilayer Perceptron (MLP)

It is multilayer perceptron (feedforward) artificial neural network and ANNs are arranged into multiple layers, corresponding neurons present in each layer or processing elements to make layers. (**Fig. 3**).

ANNs have alike topology consisting of an input layer, one or more hidden layers and an output layer. The complexity of the problem is resolved by the quantity of input neurons as it determine the number of hidden layers. The input layer networks along with outer environs collect the data as a vector of interpreter variables, each regarded as a node. An output is produced which is the sum of all the products processed by a non-linear transfer function.

The MLP helps in secondary structure of protein, relatively solvent accessibility of proteins residues, binding residues, transmembrane regions and quantitative traits from genotype data [65-68].

Recurrent Neural Network (RNN)

Recurrent neural networks are designed for sequential or time-series data. The modification in the MLP structure generates a recurrent neural network. A context layer is present which serves the purpose of holding the figures through observations. Every iteration has a new feature vector is thrust into the input layer. The preceding contents of the hidden layer are reprinted to the context layer and shoved into the hidden layer in the next iteration. The RNN processes include the input value into the input nodes. The calculation of net inputs from input nodes and from the nodes present in the context layer. The computing of hidden nodes and activation from net input Lead to calculate the output node activations. The use of back propagation algorithm to compute the new weight values and the insertion of new hidden weights in the context layer are also analyzed. The weight between the input layer and the hidden layer is taken by using the similar method MLP. To find the error values, the weights between the context and the hidden layer play a vital role as the error values depends upon the hidden nodes, received at the t_{th} iteration (**Fig. 4**).

RNN architecture can be successfully applied in the prediction of β-turns, secondary structure of proteins, number of residue contacts, continuous B-cell epitopes, binding sites of transcription factors and sequential phenotype prediction in genomics [69-71].

Radial Basis Function Neural Networks

Radial basis function (RBF) neural networks comprises of three layers as an input layer, a hidden layer (feature vector having a non-linear RBF activation function) and a linear output layer. Before the problem is determined and the classification is done, non-linear transfer function is applied to the hidden layer. The linear differentiation of the hidden layer grows by increasing the dimension of the hidden layer (**Fig. 5**).

In the hidden layers, the vectors are organized onto each RBF. It is usually applied as a Gaussian function. The values of the center and spread are set up by the aid of training dataset; the utilized method contains K means clustering or a random subset of the training vectors. While considering the regression hurdles, the output layer is a linear arrangement of values generated by the hidden layer, which associates to the mean forecasted output. In prediction hurdles, the output layer is conventionally accomplished by the aid of a sigmoid function of a linear amalgamation of hidden values layer values, denoting a posterior likelihood.

The RBF neural networks can be successfully applied in the prediction of inter-residue contact maps, cleavage sites of proteases and targets for protein-targeting compounds in drug designing [72-76].

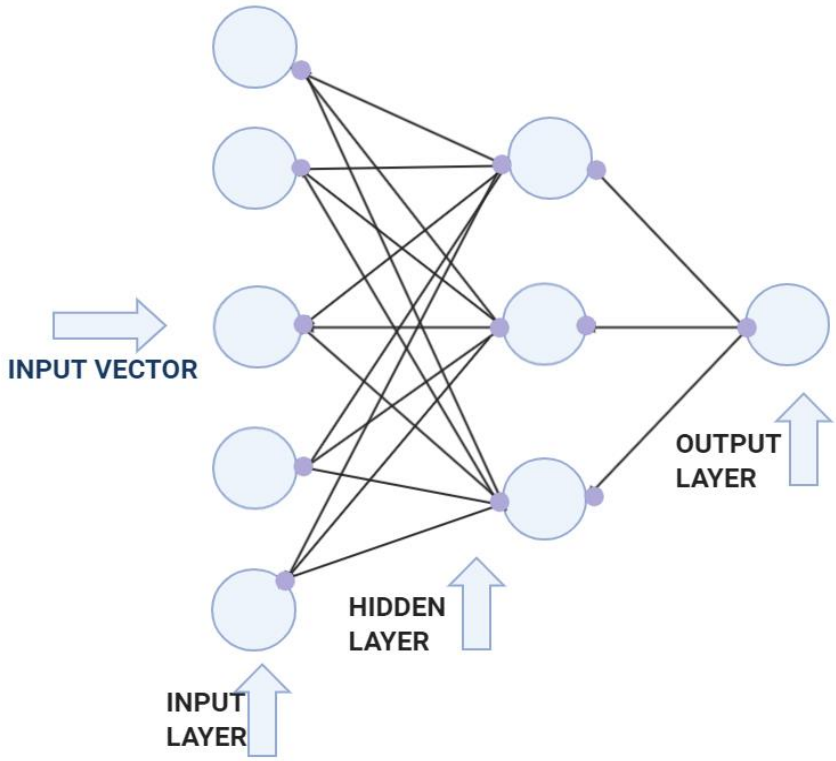

Fig. 3: Architecture of a typical multi-layered perceptron artificial neural network.

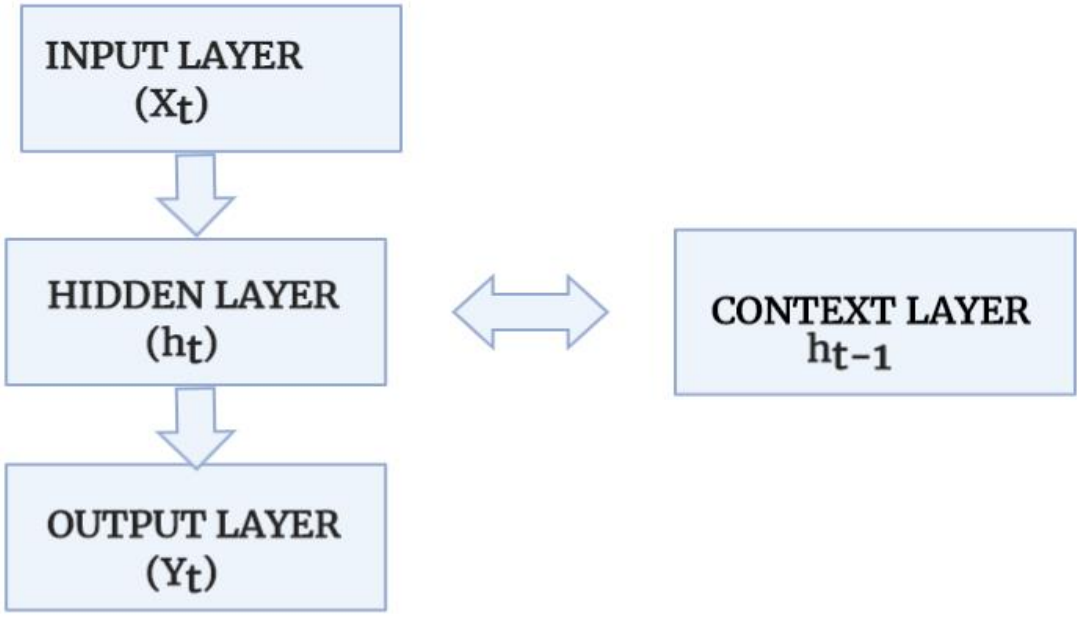

Fig. 4: Architecture of RNN. Like deep neural networks, RNN also contains an input layer, a hidden layer, and an output layer. An additional context layer is connected to the hidden layer.

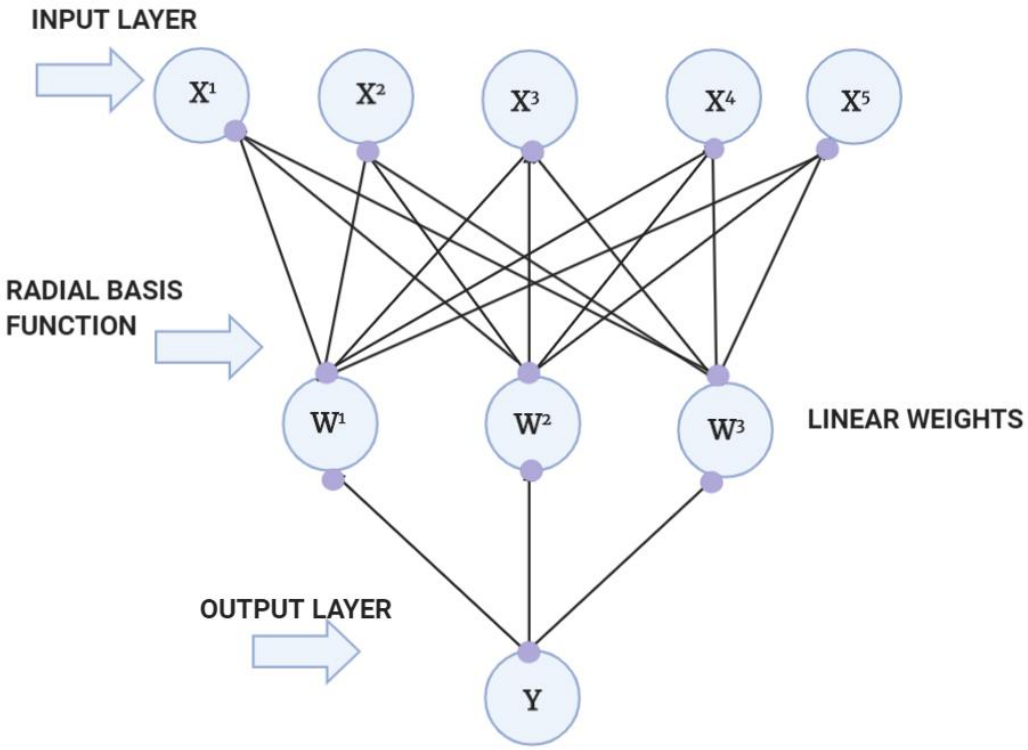

Fig. 5: Architecture of RBF neural network.

Conclusion

Artificial intelligence and machine learning are strengthening the field of computational biology and bioinformatics. As the volume of biological data is increasing, AI and machine learning promise great support in managing and performing analysis. The efficient prediction methods provide a great deal of support for researchers in the early diagnosis of diseases.

Acknowledgments

Authors are thankful to the Department of Bioinformatics and Biotechnology, Government College University Faisalabad to provide the research platform.

References

[1] Xia F, Shukla M, Brettin T, Garcia-Cardona C, Cohn J, Allen JE, et al. Predicting tumor cell line response to drug pairs with deep learning. BMC Bioinformatics. 2018; 2(23):12-15.

[2] Sehgal SA. Pharmacoinformatics, Adaptive Evolution, and Elucidation of Six Novel Compounds f or Schizophrenia Treatment by Targeting DAOA (G72) Isoforms. BioMed Research International. 2017; 2:12-16.

[3] Sehgal SA, Mirza AH, Tahir RA, Mir A. Quick Guideline for Computational Drug Design 2018; Bentham Science Publishers.

[4] Sehgal SA, Khattak NA, Mir A. Structural, phylogenetic and docking studies of D-amino acid oxidase activator (DAOA), a candidate schizophrenia gene. Theoretical Biology and Medical Modelling. 2013; 10(1):3.

[5] Sehgal SA, Mannan S, Kanwal S, Naveed I, Mir A. Adaptive evolution and elucidating the potential inhibitor against schizophrenia to target DAOA (G72) isoforms. Drug design, development and therapy 2015; 9: 3471.

[6] Sehgal SA, Mannan S, Ali S. Pharmacoinformatics and molecular docking studies reveal potential novel antidepressants against neurodegenerative disorders by targeting HSPB8. Drug Design, Development and Therapy 2016; 10:1605.

[7] Sun D, Wang M, Li A. A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. IEEE/ACM transactions on computational biology and bioinformatics. 2018; 3 (12): 3-6.

[8] Marechal, Eric. Chemogenomics: A Discipline at the Crossroad of High Throughput Technologies, Biomarker Research, Combinatorial Chemistry, Genomics, Cheminformatics, Bioinformatics and Artificial Intelligence. Combinatorial Chemistry & High Throughput Screening. 2008; 11(8): 516-532.

[9] Braga-Neto UM, Dougherty ER. Is cross-validation valid for small-sample microarray classification? Bioinformatics. 2004; 20(3): 374-80.
[10] Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, Moreira AL, Razavian N, Tsirigos A. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature Medicine. 2018; 24(10):1559
[11] Cannata N, Schröder M, Marangoni R, Romano P. A semantic Web for bioinformatics: goals, tools, systems, applications. BMC Bioinformatics. 2008; 6(7):4-8.
[12] Yu, K. H., Beam, A. L. Kohane IS. Artificial intelligence in healthcare. Nature Biomedicine. 2018; 2(12): 719–731.
[13] Keedwell, E. Narayanan. Intelligent bioinformatics: The application of artificial intelligence techniques to bioinformatics problems. John Wiley & Sons. 2005; 5(7):3-7.
[14] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015; 521: 452–459.
[15] Heinson AI, Gunawardana Y, Moesker B, Hume CC, Vataga E, Hall Y, Stylianou E, McShane H, Williams A, Niranjan M, Woelk CH. Enhancing the biological relevance of machine learning classifiers for reverse vaccinology. International journal of molecular sciences. 2017; 18(2): 312.
[16] Pavillon N, Hobro AJ, Akira S, Smith NI. Noninvasive detection of macrophage activation with single-cell resolution through machine learning. Proceedings of the National Academy of Sciences. 2018; 115(12): 2676-85.
[17] Sehgal SA, Tahir RA, Shafique S, Hassan M, Rashid S. Molecular modeling and docking analysis of CYP1A1 associated with head and neck cancer to explore its binding regions. Journal of Theoretical & Computational Science. 2014; 1(112):2.
[18] Timmis J, Andrews P, Hart E. On artificial immune systems and swarm intelligence. Swarm Intelligence. 2010; 4(4): 247-73.
[19] Li Y, Niu M, Zou Q. ELM-MHC: an improved MHC identification method with extreme learning machine algorithm. Journal of proteome research. 2019; 18(3): 1392-401.
[20] Eden, Amnon, Søraker J, Moor JH, Steinhart E. Singularity Hypotheses: A Scientific and Philosophical Assessment. The Frontiers Collection. Berlin: Springer.2012; 2(23):23-30.
[21] Saeys Y, Inza I, Larran˜aga P. A review of feature selection techniques in bioinformatics. BMC Bioinformatics. 2007; 23(19):2507–2517.
[22] Larranaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, Lozano JA, Armananzas R, Santafe G, Perez A, Robles V .Machine learning in bioinformatics. Briefings in Bioinformatics. 2006; 7(1):86–112.
[23] Castiglione F, Pappalardo F, Bernaschi M, Motta S. Optimization of HAART with genetic algorithms and agent-based models of HIV infection. Bioinformatics. 2007; 23(24): 3-5.
[24] Watson JD, Laskowski RA, Thornton JM. Predicting protein function from sequence and structural data. Current Opinion in Structural Biology 2005; 15 (3):19-23.
[25] J. Huang, Hong Fang, Xiaohui Fan. Decision forest for classification of gene expression data. Computers in Biology and Medicine 2010; 40(8): 698-704.
[26] E. Zintzaras, A. Kowald. Forest classification trees and forest support vector machines algorithms: Demonstration using microarray data. Computers in Biology and Medicine 2010; 40(5): 519-524.
[27] R. Dıaz-Uriarte, S. Alvarez de Andres. Gene selection and classification of microarray data using random forest. BMC bioinformatics 2006; 7(1):29-34.
[28] Pan W. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. Bioinformatics. 2002; 18(4): 546-54.
[29] J. W. Lee, J. B. Lee, M. Park, S. H. Song. An extensive comparison of recent classification tools applied to microarray data. Computational Statistics & Data Analysis 2005; 48(4): 869-885.
[30] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, S. Levy. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. Bioinformatics 2005; 21(5): 631-643.
[31] M. Z. Man, G. Dyson, K. Johnson, B. Liao. Evaluating Methods for Classifying Expression Data. Journal of Biopharmaceutical Statistics 2004; 14 (4): 1065-1086.
[32] Y. Lu, J. Han. Cancer classification using gene expression data. Information Systems 2003; 28(4): 243-268.
[33] López-Bigas N, Ouzounis CA. Genome-wide identification of genes likely to be involved in human genetic disease. Nucleic acids research. 2004; 32(10): 3108-14.
[34] Tang B, Pan Z, Yin K, Khateeb A. Recent advances of deep learning in bioinformatics and computational biology. Frontiers in genetics. 2019; 10.
[35] Michener, C. D., Sokal R. R. A quantitative approach to a problem in classification. Evolution. 1975; 11: 130-162.
[36] Plis SM, Hjelm DR, Salakhutdinov R, Allen EA, Bockholt HJ, Long JD, Johnson HJ, Paulsen JS, Turner JA, Calhoun VD. Deep learning for neuroimaging: a validation study. Frontiers in neuroscience. 2014; 8: 229.
[37] Leung MK, Xiong HY, Lee LJ, Frey BJ. Deep learning of the tissue-regulated splicing code. Bioinformatics. 2014; 30(12): 121-9.
[38] Hanif W, Afzal MA, Ansar S, Saleem M, Ikram A, Afzal S, Khan SAF, Larra SA, Noor H. Artificial intelligence in bioinformatics. Biomedical Letters 2019; 5(2):114-119.
[39] Mamoshina P, Vieira A, Putin E, Zhavoronkov A. Applications of deep learning in biomedicine. Molecular pharmaceutics. 2016; 13(5): 1445-54.
[40] J. R. Quinlan. Induction of Decision Trees. Machine learning 1986; 1(1): 81-106.
[41] Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nature Reviews Genetics. 2015; 16(6): 321-32.
[42] Chen Y, Li Y, Narayan R, Subramanian A, Xie X. Gene expression inference with deep learning. Bioinformatics. 2016; 32(12): 1832-9.
[43] Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. Nature biotechnology. 2015: 33(8): 831-8.
[44] J. R. Quinlan. Simplifying decision trees. International Journal of Man-Machine Studies 1987; 27(3): 221-248.
[45] Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow PM, Zietz M, Hoffman MM, Xie W. Opportunities and obstacles for deep learning in biology and medicine. Journal of The Royal Society Interface. 2018; 15(141): 20-34.
[46] Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. Molecular systems biology. 2016; 12 (7). 35-40.
[47] Michalewicz Z. Genetic Algorithms + Data Structures = Evolution Programs. Springer-Verlag,Berlin. 1996; 1(32):3.
[48] Notredame C, Higgins DG. SAGA: Sequence alignment by genetic algorithm, Nucleic Acids Research. 1996; 24 (8): 1515-1524.
[49] Rothlauf F. Representations for Genetic and Evolutionary Algorithms. Springer. 2002; 5(2):4.
[50] Moghram BA, Nabil E, Badr A. Ab-initio conformational epitope structure prediction using genetic algorithm and SVM for vaccine design. Computer methods and programs in biomedicine. 2018; 153: 161-70.
[51] Holland JH. Genetic algorithms and the optimal allocation of trials. SIAM Journal on Computing, 1973; 2:88-105.
[52] Fogel GB, Corne DW. Evolutionary Computation in Bioinformatics. Proceedings of the National Academy of Science. 2002; 2(3):2.

[53] Chung YC, Randy LH. Amplitude and phase adaptive nulling with a genetic algorithm. Journal of Electromagnetic Waves and Applications 14.5 2000: 631-649.
[54] Goldberg DE. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley. 1989; 4 (54): 20-26.
[55] Douzono H, Hara S, Noguchi Y. An application of genetic algorithm to DNA sequencing by oligonucleotide hybridization. Proceedings of the IEEE international joint symposia on intelligence and systems Rockville, Maryland, USA. 1998; 5(34): 92–98
[56] Fogel GB. Computational intelligence approaches for pattern discovery in biological systems. Briefings in Bioinformatics. 2008; 9(4):307–316.
[57] Agrawal R, Srikant R. Fast algorithms for mining association rules. BMC Bioinformatics. 1994; 3(35): 12-16.
[58] Adamczak R, Porollo A, Meller J. Combining prediction of secondary structure and solvent accessibility in proteins. Proteins 2005; 59: 467–475.
[59] Ahmad S, Gromiha MM, Sarai A (2003) Real value prediction of solvent accessibility from amino acid sequence. Proteins 2005; 50: 629–635.
[60] Ahmad S, Gromiha MM, Sarai A. Analysis and prediction of DNA-binding proteins and their binding residues based on composition, sequence and structural information. Bioinformatics 2004; 20: 477–486.
[61] Ahmad S, Sarai A. PSSM-based prediction of DNA binding sites in proteins. BMC Bioinformatics 2005; 6: 33
[62] Reinhardt A, Hubbard T .Using neural networks for prediction of the subcellular location of proteins. Nucleic Acids Research 1998; 26:2230–2236
[63] Blom N, Hansen J, Blaas D, Brunak S. Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. Protein Science 1996; 5: 2203–2216.
[64] Byvatov E, Schneider G. Support vector machine applications in bioinformatics. Applied Bioinformatics 2003; 2(2): 67–77.
[65] Cai YD, Zhou GP. Prediction of protein structural classes by neural network. Biochemistry 2000; 82: 783–785
[66] Cai YD, Liu XJ, Chou KC (2002) Artificial neural network model for predicting protein subcellular location. Computational Chemistry 2002; 26: 179–182.
[67] Giorgi JM, Bader GD. Transfer learning for biomedical named entity recognition with neural networks. Bioinformatics. 2018; 34 (23): 4087-94.
[68] Chandonia JM, Karplus M. Neural networks for secondary structure and structural class predictions. Protein Science 1995; 4: 275–285
[69] Ranawana R, Palade V. A neural network based multi-classifier system for gene identification in DNA sequences. Neural Computing & Applications. 2005; 14(2): 122-31.
[70] Fogel GB. Computational intelligence approaches for pattern discovery in biological systems. Brief Bioinformatics 2008; 9(4): 307–316.
[71] Khashei M, Hamadani AZ, Bijari M. A fuzzy intelligent approach to the classification problem in gene expression data analysis. Knowledge-Based Systems. 2012; 27: 465-74.
[72] Wu CH. Artificial neural networks for molecular sequence analysis. Computers & chemistry. 1997; 21(4): 237-56.
[73] Jacoboni I, Martelli PL, Fariselli P, De Pinto V, Casadio R. Prediction of the transmembrane regions of beta-barrel membrane proteins with a neural network-based predictor. Protein Science 2001; 10:779–787.
[74] Qian N, Sejnowski TJ. Predicting the secondary structure of globular proteins using neural network models. Journal of Molecular Biology 1988; 202:865–884.
[75] Saeys Y, Inza I, Larran˜aga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007; 23(19):2507–2517.
[76] Dor O, Zhou Y. Real-SPINE: an integrated system of neural networks for real-value prediction of protein structural properties. Proteins 2007; 68: 76–81.