*Corresponding Author

Hasnat Noor
E-mail
hasnatmalik7127@gmail.com

¥Both Authors contributed equally

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

Open Access

Artificial intelligence in bioinformatics

Waqar Hanif [1]¥, Muhammad Amin Afzal [1]¥, Saba Ansar[1], Muntaha Saleem[1], Ayesha Ikram[1], Shumaila Afzal[1], Seemab Amjad Fateh Khan[2], Sajjad Ahmad Larra[1], Hasnat Noor[1]*

[1]Department of Bioinformatics and Biotechnology, Government College University Faisalabad, Pakistan
[2]Department of Pharmacy, Novomed Hospital, Dubai, United Arab Emirates

Abstract

There are considerable improvements over the years in both Bioinformatics and Computer Sciences leading to multiple breakthroughs in the respective fields. As bioinformatics is about finding new ways to analyze the data (huge data) for logical conclusions. Artificial Intelligence can be used to analyze process and categorize the gigantic amount of biological data in less time. Numerous AI algorithms have been developed and used in bioinformatics analyses. Such algorithms are used to generate knowledge in an efficient way. The algorithms based on the applications of AI helps in molecular dynamic simulations, discovering vaccines for diseases, molecular docking analyses, identifications of novel compounds, ADMET properties, machine and deep learning of tools and *in silico* structure prediction. This review article summarizes the applications of AI that employed in bioinformatics; specifically, genetic algorithm in bioinformatics tools.

Introduction

From last decade, there has been much progress in bioinformatics and Artificial Intelligence (AI), and more opportunities are available to understand the biological data and problems. Bioinformatics is an interdisciplinary field of science utilized to solve numerous biological problems by employing mathematical models, statistical approaches, and algorithms along with computational powers [1]. AI is the ability of a computational system to perform different tasks linked with intelligent beings and as a simulation of human intelligence processes by the computer systems [2]. AI is the ability of a machine to abstract, be creative and perform the tasks based on its training [3]. Bioinformatics approaches have been utilized to solve numerous biological problems [4]. The structural bioinformatics tools have the applications of AI and have effective methodologies to design active novel compounds against neurological disorders [5-11] and cancer [12-17] through *in silico* means by using the tools having AI applications. Bioinformatics approaches are used to analyse the biological data for logical conclusions. The projects of whole genome sequencing are generating an enormous amount of biological data related and bioinformatics approaches help to solve and annotate the data in meaningful ways. Numerous biological problems have been solved by employing the approaches of AI along with bioinformatics approaches and help to design effective algorithms for gene prediction, computational drug designing, protein-protein interactional studies, genome wide association studies, next generation sequences and software development. AI has wide range of application in bioinformatics [18, 19]. The efficiency of AI improves by modifying its code according to the input data. AI is classified into two types as generalized AI and Applied AI. The generalized AI is the development of AI in which the machines perform the tasks as humans do while applied AI is the development that can stimulate as human-like thoughts and expressions [20-22]. AI and heuristic methods showed significance for the present and future developments of bioinformatics and having the potential for a revolutionary impact on biotechnology, pharmacology and medicine [23, 24]. Now a day, applied AI helps in the prediction of personalized medicines by analyzing the genomic data of the patients along with the diagnoses of the patients. The algorithms base on AI has the ability to analyze the data efficiently as compared to the algorithms having no AI training. The algorithms of

AI including linear regression, logistic regression and K-Nearest neighbors' algorithm in bioinformatics improve the efficiency of the computational biology to solve the biological problems [25, 26].

Application of AI in bioinformatics

There is a gigantic amount of data in biological databases and the annotation of this data is still a question for the researchers. The applications of AI in bioinformatics have the ability to annotate the data towards logical conclusions [27]. The simulations of different models, annotations of biological sequences, computational drug designing, virtual screening and gene prediction can efficiently predict through the association of AI and bioinformatics [28]. The major contribution of AI in bioinformatics analyses depends on pattern matching and knowledge-based learning systems to solve the biological problems.

The advancement in AI and bioinformatics has significant contribution in health sciences including immunoinformatics [29] and vaccinology and leads to the clinical bioinformatics, high-throughput screening, disease prevention and epidemiology. The development of new vaccines is becoming more difficult due to the increased rate of mutations in microbes and viruses. The advancement helps to improve the power and algorithms of computational simulation. Now a day, computational systems have the capacity to screen the vaccine targets from >20,000 flavivirus proteins to >100,000 influenza proteins. The generated data can be interpreted by using different methods results to logical conclusions. Researchers designed *in silico* methodologies to discover vaccines (**Figure 1**).

The success of AI in bioinformatics has widely used algorithms and methodologies including neural networks, probabilistic approach, decision trees, cellular automata, hybrid methods and genetic algorithms to solve numerous biological problems [30, 31].

Genetic Algorithm

A genetic algorithm is a heuristic-based search algorithm inspired by the natural evolution theory of Charles Darwin. The genetic algorithm follows the idea of natural selection to select the fittest survival and the significant output [32]. The genetic algorithm is designed to perform the tasks in five major phases
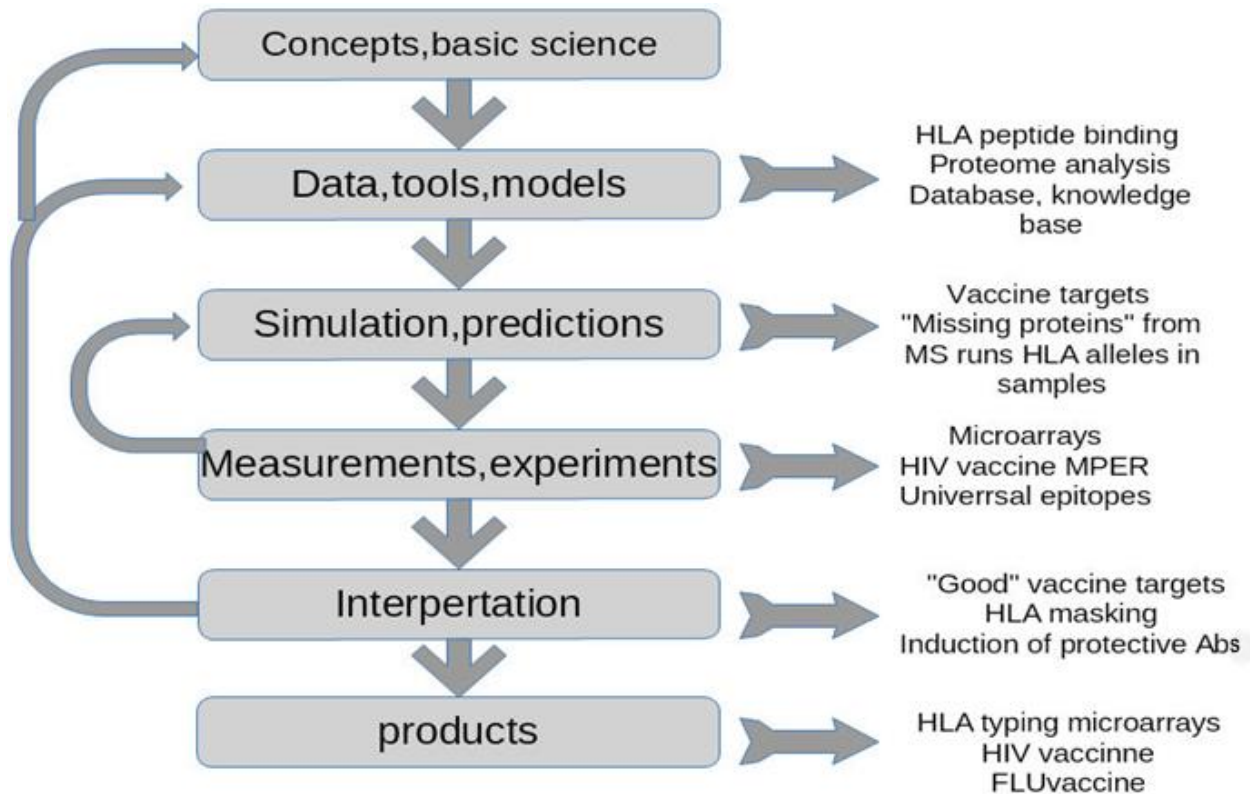
Fig. 1: Workflow of vaccine discovery process including bioinformatics tools and reasoning.

(**Figure 2**) and select the fittest iteration from the generation [33].

Initially, the first population generates a set of individuals known as population and each individual have genes (defined by a set of parameters). The combinations of defined set of genes form a chromosome. The fitness function determines the efficiency of an individual and calculates a score for each individual. The calculated score determines the selection of an individual leads to the selection of the fittest individual based on the fitness score for the progeny. The parents are selected for the reproduction through this process. The crossover is considered as the significant step for each pair of parents for the random selection of the genes results to the generation of offspring. There is a probability that the strings of the genes can flip for the generation of new offspring at mutation phase. The algorithm terminates by the repeated formation of the same generations, considered as the final product [34].
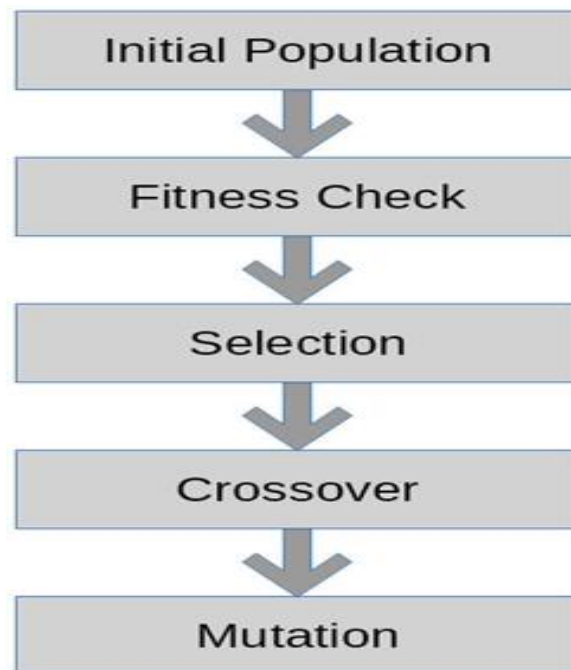

Fig. 2: Flowchart of the major followed phases of the Genetic Algorithm.
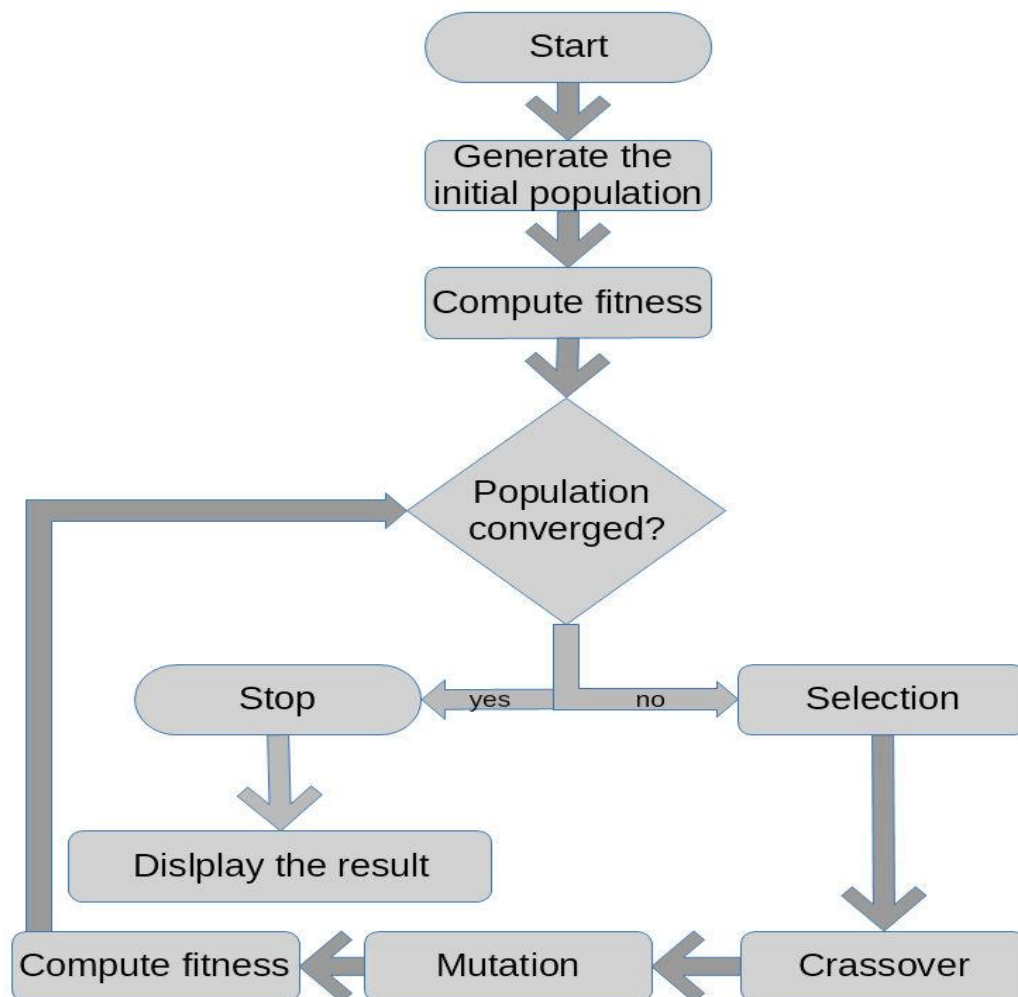
Fig. 3: The flowchart of the Genetic Algorithm.

Genetic Algorithm is utilized to effectively enhance the multiple sequence alignment [35, 36]. This approach uses a population of alignments to generate a fitness score based on the matching and mismatching of the columns [37-39]. Half of the suitable alignments are copied to the next generation leads to the crossover points to choose a cut for the random selection point in the first alignment sequence and another cut is made for the second alignment to adjust the first sequence. One parent is spliced to add the gaps leading to the splicing of another parent to add the gaps to ensure the alignment consistency [40-42]. Genetic algorithm predicts efficient alignments results as compared to other alignment algorithms [43-45]. The applications of AI in bioinformatics play an important role to solve the biological problems computationally [46-50].

Conclusion

In future, Bioinformatics will play a significant role to analyze the huge data sets by using AI leads to save the time and resources. It will also help to accelerate the biological discoveries, especially in medicine, biomedical fields and robotic surgery.

Acknowledgments

Conflict of interest

The authors declare no conflict of interest.

References

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

[2] 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.

[3] 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.

[4] 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.

[5] Sehgal SA, Hassan M, Rashid S. Pharmacoinformatics elucidation of potential drug targets against migraine to target ion channel protein KCNK18. Drug design, development and therapy, 2014; 8, 571.

[6] Sehgal SA. Pharmacoinformatics and molecular docking studies reveal potential novel Proline Dehydrogenase (PRODH) compounds for Schizophrenia inhibition. Medicinal Chemistry Research 2017; 26(2):314-326.

[7] 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.

[8] Tahir, Rana A, Sehgal SA. Pharmacoinformatics and molecular docking studies reveal potential novel compounds against schizophrenia by target SYN II. Bentham Science Publishers. 2018; 21(3): 5-8.

[9] Jamil F, Ali A, Sehgal SA. Comparative modeling, molecular docking, and revealing of potential binding pockets of RASSF2; a candidate cancer gene. Interdisciplinary Sciences: Computational Life Sciences 2017; 9(2): 214-223.

[10] 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.

[11] Sehgal SA, Kanwal S, Tahir RA, Khalid Z, Hammad MA. In silico elucidation of potential drug target sites of the Thumb index Fold Protein, Wnt-8b. Tropical Journal of Pharmaceutical Research. 2018; 17(3):491-7.

[12] Tahir RA, Sehgal SA, Khattak NA, Khattak JZK, Mir A. Tumor necrosis factor receptor superfamily 10B (TNFRSF10B): an insight from structure modeling to virtual screening for designing drug against head and neck cancer. Theoretical Biology and Medical Modelling 2013; 10(1): 38.

[13] Dali Y, Abbasi SM, Khan SAF, Larra SA, Rasool R, Ain QT, Jafar TH. Computational drug design and exploration of potent phytochemicals against cancer through in silico approaches. Biomedical Letters 2019; 5(1):21- 26.

[14] 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.

[15] 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.

[16] 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.

[17] 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

[18] Cannata N, Schröder M, Marangoni R, Romano P. A semantic Web for bioinformatics: goals, tools, systems, applications. BMC Bioinformatics. 2008; 6(7):4-8.

[19] Yu, K. H., Beam, A. L. Kohane IS. Artificial intelligence in healthcare. Nature Biomedicine. 2018; 2(12): 719–731.

[20] Keedwell, E. Narayanan. Intelligent bioinformatics: The application of artificial intelligence techniques to bioinformatics problems. John Wiley & Sons. 2005; 5(7):3-7.

[21] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015; 521: 452–459.

[22] 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.

[23] Saeys Y, Inza I, Larran˜aga P. A review of feature selection techniques in bioinformatics. BMC Bioinformatics. 2007; 23(19):2507–2517.

[24] 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.

[25] 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.

[26] Watson JD, Laskowski RA, Thornton JM. Predicting protein function from sequence and structural data. Current Opinion in Structural Biology 2005; 15 (3):19-23.

[27] Sajda, Paul. Machine learning for detection and diagnosis of disease. Annual Revision Biomedical Engineering. 2006; 8: 537-565.

[28] Tahir RA, Wu H, Rizwan MA, Jafar TH, Saleem S, Sehgal SA. Immunoinformatics and molecular docking studies reveal potential epitope-based peptide vaccine against DENV-NS3 protein. Journal of theoretical biology. 2018; 459: 162-170.

[29] Lee JW, Lee JB, Park M, Song SH. An extensive comparison of recent classification tools applied to microarray data. Computational Statistics & Data Analysis. 2005; 48 (4):9-15.

[30] Altman R. Challenges for intelligent systems in biology. IEEE Intelligent Systems November/December. 2001; 2(32): 14-18.

[31] Narayanan A, Wu X, Yang ZR. Mining viral protease data to extract cleavage knowledge. BMC Bioinformatics. 2002; 1: 5-13.

[32] Agrawal R, Srikant R. Fast algorithms for mining association rules. BMC Bioinformatics. 1994; 3(35): 12-16.

[33] Bhandari D, Murthy CA, Pal SK. Variance as a stopping criterion for genetic algorithms with elitist model. Fundamata Informaticae. 2012; 120: 145-164.

[34] Burge C. Karlin S. Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology. 1997; 268: 78-94.

[35] 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

[36] Fogel GB. Computational intelligence approaches for pattern discovery in biological systems. Briefings in Bioinformatics. 2008; 9(4):307–316.

[37] Goldberg DE. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley. 1989; 4 (54): 20-26.

[38] Corpet F. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Research. 1988; 16 (22): 10881-10890

[39] Bard JB, Rhee SY. Ontologies in biology: design, applications and future challenges. Nature Reviews Genetics. 2004; 5(3):213.

[40] Cannata N, Schröder M, Marangoni R, Romano PA. Semantic Web for bioinformatics: goals, tools, systems, applications. BMC Bioinformatics. 2008; 9(4): 1.

[41] Feng DF, Doolittle RF. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. Journal of Molecular Evolution. 1987; 25(4):351-360.

[42] Deagustini CAD, Martinez MV, Falappa MA, Simari GR. Datalog ontology consolidation. Journal of Artificial Intelligence Research. 2016; 56: 613–656.

[43] Hassanien AE, Milanova MG, Smolinski TG, Abraham. Computational Intelligence in Solving Bioinformatics Problems: Reviews, Perspectives, and Challenges. Springer Berlin Heidelberg, Berlin, Heidelberg. 2008; 2(34):3–47.

[44] Lipman DJ, Altschul SF, Kececioglu JD. A tool for multiple sequence alignment, Proceedings of the National. Academy of Science. 1989; 86(12): 4412-4415.

[45] Michalewicz Z. Genetic Algorithms + Data Structures = Evolution Programs. Springer-Verlag,Berlin. 1996; 1(32):3.

[46] Notredame C, Higgins DG. SAGA: Sequence alignment by genetic algorithm, Nucleic Acids Research. 1996; 24 (8): 1515-1524.

[47] Rothlauf F. Representations for Genetic and Evolutionary Algorithms. Springer. 2002; 5(2):4.

[48] Holland JH. Genetic algorithms and the optimal allocation of trials. SIAM Journal on Computing, 1973; 2:88-105.

[49] Fogel GB, Corne DW. Evolutionary Computation in Bioinformatics. Proceedings of the National Academy of Science. 2002; 2(3):2.

[50] Chung YC, Randy LH. Amplitude and phase adaptive nulling with a genetic algorithm. Journal of Electromagnetic Waves and Applications 14.5 2000: 631-649.