Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2582-2845.7918

ISSN: 2582 – 2845

Ind. J. Pure App. Biosci. (2019) *7*(6), 317-321

Indian Journal of Pure & Applied Biosciences

Review Article

Role of Artificial Intelligence in Bioinformatics

Vijay Rana*

Assistant Professor, Sant Baba Bhag Singh University, Jalandhar, India

*Corresponding Author E-mail: vijay.rana93@gmail.com

Received: 4.11.2019 | Revised: 17.12.2019 | Accepted: 26.12.2019

ABSTRACT

The bioinformatics science is characterized by a huge amount of information from heterogeneous resources and challengeable to depict the desirable relationships between heterogeneous datasets still remain issue. The Artificial Intelligence (AI) has abilities to dominate bioinformatics data structure for its unrecorded learning abilities to process complex information. The several AI techniques are attained from computer vision, recognition system, or natural language processing tools that signify the realization of precision medicine. This main idea behind this work to summarized the valuable contribution of AI in the areas of bioinformatics and computational molecular biology. This work also describes the artificial techniques that developed by the eminent researchers in order to classify, managed and mine different available biological databases and biological experiments.

Keywords: *DNA, Bioinformatics, Artificial Intelligence, Biology.*

INTRODUCTION

Bioinformatics is an area of evaluation of biological information and basic appliance in this area includes estimation of biological series and molecular constitution where as emerging services like modelling of biological systems. With the rapid growth of bioinformatics applications, system required some intelligent model for development of real time model in biological aspect. The Artificial intelligence (AI) has rapidly obtained concentration in biological research and automatically molecular biology. With the accessibility of dissimilar types of AI implementations, it has become general technique for the researchers to use the off-shelf model to organize and attain their databases. Present time, with several intelligent algorithms obtainable in the literature and scholars are expression problems in selecting the preeminent approach that could be used to a particular information set. In this work, we summarize the review the role of artificial intelligent in the field of bioinformatics and computational molecular biology.

Bioinformatics and artificial intelligent are related with the exploit of computerized measure to recognize biological techniques and to obtain and use biological information, progressively more high-scale data.

Cite this article: Rana, V. (2019). Role of Artificial Intelligence in Bioinformatics, *Ind. J. Pure App. Biosci. 7*(6), 317-321. doi: http://dx.doi.org/10.18782/2582-2845.7918

The techniques from bioinformatics and artificial intelligence are progressively more utilized to enhance or control conventional laboratory and practical based real biology applications (Navjot et al., 2019). These techniques have become significant in biological domain due to latest development in area of informatics and identify to obtain a huge biological information sets, and due to the omnipresent, booming biological imminent, which have come from the operation of earlier information. This development from poor information set to a rich data inset from the work start with DNA sequence information, but is now occurring in several other filed of biology.

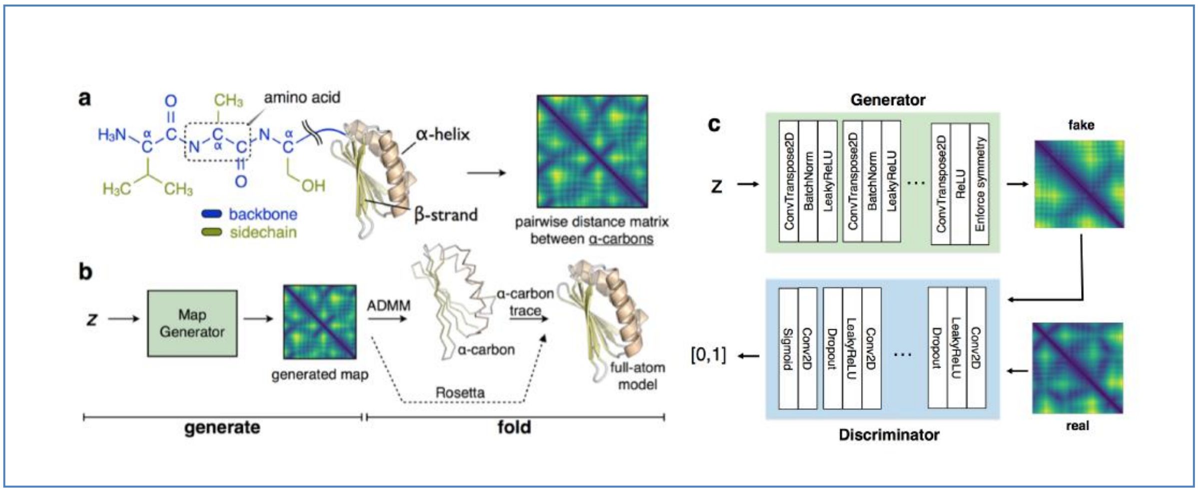

Fig. 1: Artificial and Bioinformatics

This combination obtains a close impact in the expansion of bioinformatics for at least two forms. The first phase defines the working of Artificial Intelligence, when processing make a resource of encouragement for the use of real time computational applications that are stronger to the management of real information such as tolerance of tentative and unfocused information, auto edition and learning skills.

This paper has been formalized into five sections. Section 2 presents the work of eminent researchers. Section 3 defines the proposed work and Section 4 discusses the results. Section 5 finally concludes.

RELATED WORK

The section highlights the work of researchers and describes the unfolded issues which further needed kind of attentions.

Zoheir et al. (2005) summarized the overview of Artificial Applications in bioinformatics and DNA sequencing and review shown that existing tools have lot of data for enhancing the existing biology techniques with machine learning that are already being used to retrieve valuable information and to discover the functional patterns from the enormous amount of raw biological data sets. This work also highlights the several software tools that are specially developed for bioinformatics applications by the research community in order to discover, organize and access dissimilar obtainable biological databases to simulate biological experiments with and without errors ratio.

Janice et al. (2004) described Artificial Intelligence techniques from ontology domain representation, machine learning structure, information accessibility and reasoning are at the forefront in defining the emerging challenges that are occurring in molecular biology. The authors highlights the several issues, which are related with molecular progression, structure and function rely on the capability to control and comprehend these information. It is assumed that machine learning technique is able to addressing the significant issues, which are occurring in molecular biology.

Miguel et al. (2018) proposed a flexible model to work with soilless measure in recirculation greenhouses utilizing reasonably saline water. That totally depended on IoT system with low cost hardware and maintained by an open source software service at both local and cloud environment. The combination of IoT technology and cloud computing provides more opportunities to improve Precision Agriculture (PA) connectivity. The experimental result shown that proposed system entirely instantiated in a real time working features in frames of the Indian drain project and allocating the control of a real hydroponic system with the help of existing software.

Kaitao et al. (2018), highlights the review of supervised and unsupervised machine learning techniques in bioinformatics domain and also recent development in the areas of biology computational. The authors also define the training dataset, validation dataset and test dataset and the training set is used to make a models. The validation dataset is used to analysis the oversimplification issue of the final model for evaluating the parameters. The experimental results shown that machine learning is most appropriate technique for biological applications.

Nathan et al. (2019) proposed a integrated approach for automatically extract patients diagnosis information from the patients digital data. This approach generally consists of enriching the AI database accessible to the clinician for the decision support of diagnosis and recently development described that bioinformatics has contains a novel turn with the arrival of AI and machine learning. The Image evaluation reports defined machine learning has attained the most remarkable results performance, especially with use of convolution neural networks (CNN).

ROLE OF ARTIFICIAL INTELLIGENCE IN BIOINFORMATICS

The artificial intelligence has described probable in evaluating outsized, compound biological information. However, this development is needed in addition to intelligence learning for successful application of bioinformatics. The AI technique has been widely utilized in biological studies for calculation and sighting. With the expending the accessibility of dissimilar types of biological database, the application of artificial intelligence, especially machine learning techniques, has become more recurrent.

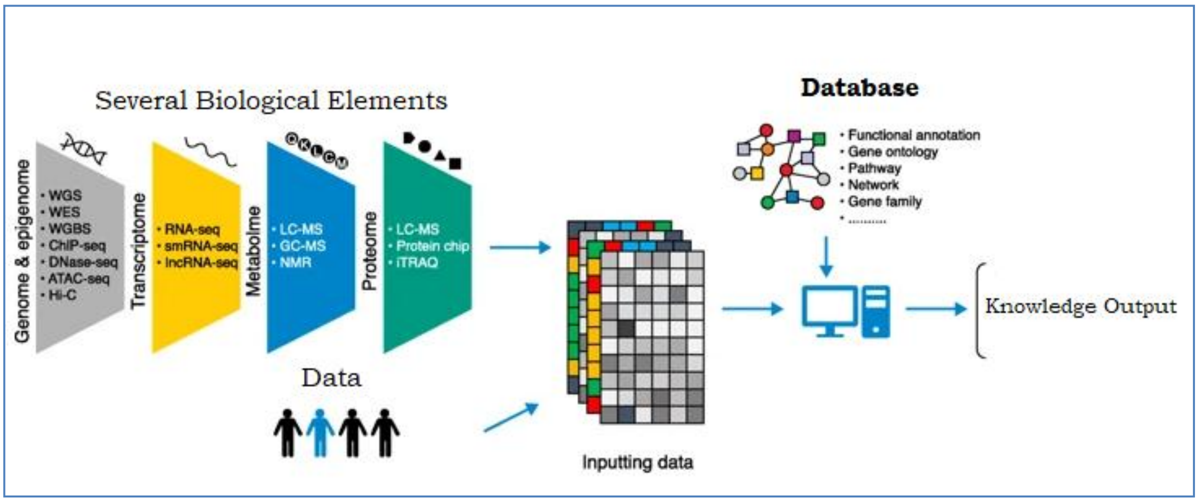

Fig. 2: Elements Model

The major area of implement the artificial intelligence techniques is in the determination of genomic features, especially those that are difficult to measure through present techniques such as regulatory regions.

Protein Classification

There are number of techniques developed to measure the biological aspects, specifically protein classification such techniques are support vector machines, machine learning, decision tree algorithm, logistic regression, KNN algorithm and neural networks algorithm. The logistic regression is a regression technique that works on dependent variable (DV) and contained the feature of robustness and efficiency algorithm. This technique based on a training data set produced from 18 proteins' mixture dataset. The AI provides most cost effective techniques to automated find of a priori unknown extrapolative connections from outsized data sets in bioinformatics database. To measure the proteins function, it formalized into various groups as per the functions they perform.

Biological Sequence Matching

Biological sequence is a measure described a set of terms, where the idea of distance among them is depend on the similarity of their biological aspects (Sunny et al., 2019). It used to calculate the amount of the semantic similarity among biological elements of nature and concepts through a cosine similarity measure according to the evaluation of data relating with their meaning or defining their domain. The cosine similarity measure finds similarity among two vectors (two documents on the Vector Space) and measures the cosine of the angle among them. The cosine similarity measure is to resolve the equation of the dot product for the $\emptyset$:

$$a, b = ||a||b|| \cos.\theta \quad (1)$$

$$cos\emptyset = \frac{\mathrm{a} * \mathrm{b}}{||\mathrm{a}||\mathrm{b}||} \quad (2)$$

Cosine Similarity measure generates a metric that find the related between two documents by defining at the angle instead of magnitude.

Biological Information Exploring

This technique is used to extract the biological information of elements, which contains the Log Files, cookies and session identification (Sunita et al., 2019). A log document is a record that records either occasions that happen in a working framework or other programming runs, or messages between various clients of correspondence programming. Logging is the demonstration of keeping a log.

Biological Information Exploring

```
Step 1: Initialization of all word txt = $Watson()$;
        $wn.synset('txt')$
        $wn.synset.lemma()$
Step 2: Extraction and similarity computation
        $if\ (wn.synste('txt') < 1)$
            $then\ return\ ('txt')$
        $else$
        $for\ synset\ \in wn.synset('txt')$
            $Sense = txt \in W \times POS \rightarrow 2^{SYNSETS}$
$Step$ 3: If (IP address is unique)then Create similar User Cluster
            $Return\ similar\_user\_cluster$;
            Similarity Matrix (sim_matrix)
Step 4: Define the Result
```

For Web seeking, an exchange log is an electronic record of connections that have happened amid a looking scene between a web search tool and clients hunting down data on that web crawler. Later sections are ordinarily attached as far as possible of the document. Data about the demand, including customer IP address, ask for date/time, page asked for, HTTP code, bytes served, client specialist, and referrer are normally included. This information can be joined into a solitary record, or isolated into unmistakable logs, for example, an entrance log, mistake log, or referrer log. Be that as it may, server logs normally don't gather client explicit data.

CONCLUSION

The artificial intelligence will become instrumental in all future bioinformatics research and more researchers in bioinformatics field are interested to deal with huge amounts of information to amalgamate, systematize, and explore this current data structure. The literature highlights the rapid expansion of bioinformatics domain with vital information that is significant for making the sound biomedical research.

REFERENCES

Adhikari, B., Bhattacharya, D., Cao, R., & Cheng, J. (2015). CONFOLD: Residue-residue contact-guided ab initio protein folding, *Proteins*, *83*, 1436–1449.

Ben-Hur, A., & Noble, W.S., (2005). Kernel methods for predicting protein-protein interactions. *Bioinformatics*, *21*, 38-46.

Bhattacharya, D., Cao, R., & Cheng, J. (2016), UniCon3D: de novo protein structure prediction using united-residue conformational search via stepwise, probabilistic sampling, *Bioinformatics*, *18*, 2791–2799.

Ezziane, Z. (2005), Applications of artificial intelligence in bioinformatics: A review, *Expert Systems with Applications 30*, 2–10.

Foulquier, N., Redou, P., Saraux, A. (2019), How Health Information Technologies and Artificial Intelligence May Help Rheumatologists in Routine Practice, Rheumatology and Therapy – *Springer*. *6*, pp 135-138.

Glasgow, J., Jurisica, I., & Rost, B. (2004). AI and Bioinformatics, *AI Magazine*. *25*, 4-6.

Kaur, N., & Rana, V. (2019). Word Sense Disambiguation for a domain-specific lexical sample task, Research Review *International Journal of Multidisciplinary*. *4*(2), 1002-1107.

Li, D., Ju, Y., & Zou, Q. (2016). Protein Folds Prediction with Hierarchical Structured SVM, *Curr. Proteomics*, *13*(2), 79-85.

Si, D., Ji, S., Nasr, K. A., & He, J. (2012). A machine learning approach for the identification of protein secondary structure elements from electron cryo-microscopy density maps, *Biopolymers*. *97*, 698–708

Sharma, S., Sunita, & Rana, V. (2018), An Optimum Approach for Preprocessing of Web User Query, *International Journal of Informatics and Communication Technology*, *Springer*. *7*(1), 8-12.

Sunita, & Rana, V. (2019). An Effective Preprocessing Algorithm for Information Retrieval System, *International Journal of Recent Technology and Engineering (IJRTE)*. *8*(3), 6371-6375.

Twine, K.L.N., Guo, Y. Denis, (2018), Artificial Intelligence and Machine Learning in Bioinformatics, *Encyclopedia of Bioinformatics and Computational Biology-Elsvier*. 1-16.

Zamora-Izquierdo, M.A., Santa, J., & Martınez b, J.A. (2018). Smart farming IoT platform based on edge and cloud computing, bio systems engineering, *Elsevier*. *26*, 1-14.