Commentary

Artificial Intelligence as Roots of Computational Biology

Perry Marshall*

Department of Biomolecular Engineering, University of Montreal, Quebec, Canada

DESCRIPTION

Computer algorithms became particularly useful in processing and handling a vast amount of data in the 1960s, when health informatics was born. Artificial Intelligence algorithms [1], which are available in a variety of forms, have grown popular among academics because they allow them to readily study more information that is annotated with descriptions, precision assessments, and explanations. Precision medicine, according to the National Institute of Health (NIH), is the best technique for preventing and curing diseases, but it can be more understandable with the help of powerful supercomputers and creative algorithms.

Genetics, molecular biology, computer science, mathematics, and statistics are all part of bioinformatics, which is an interdisciplinary field [2]. Modeling biological mechanisms at the molecular level and interpreting outcomes from obtained data are the most basic difficulties. A bioinformatics method typically comprises the following steps:

• Compiling data from biological evidence.
• Creating a computation database.
• Identifying and resolving a computational modeling issue.
• Putting computer algorithms to the test and evaluating them.

Artificial Intelligence (AI) is a term that refers to a computer that mimics human thought and behaviour in order to complete tasks that humans have assigned to it. Researchers have been heavily relying on AI technologies to classify biological sequences, identify biological entities, and determine attributes, which requires acquiring, organising, and analysing vast amounts of data that are inaccessible to humans.

AI COULD BE A TREMENDOUS AID IN VARIOUS SECTORS

Generative modeling for protein structures

Since it is impossible to compare and learn about genuine data distribution, generative models learn the true interpretation of data distribution, resulting in distributed data that looks similar to the original. GANs (Generative Adversarial Networks) are utilised in this scenario, and they consist of a generator (which learns to give output) and a discriminator (learns to compare true data from the output). These two are neural networks in which the generator tries to generate the actual image and the discriminator tries to determine if the image generated is true or false. GANs, for example, are used to generate protein structure and to anticipate the missing components of a compromised protein structure. The data involves the utilisation of a three-dimensional protein structure, which is interpreted as the two-dimensional paired alpha-carbon size.

DNA sequencing

DNA sequencing is the process of finding the order of nucleotide sequences (A, T, G, and C) in a DNA strand. However, because each organism has its own nucleotide sequences, comprehensive DNA sequencing was not possible. Artificial intelligence is used by companies like Deep Genomics to assist researchers in understanding genetic diversity. Algorithms are developed based on patterns observed in large amounts of genetic data, which are then turned into computer models to assist people understand how essential cellular processes influence genetic variation [3]. In comparison to the old Sanger sequencing procedure, which took over a decade to complete when the human genome was first sequenced, Next-Generation Sequencing encompasses new DNA sequencing methods, allowing researchers to sequence a full human genome in a single day.

Protein classification

Polypeptides are simple amino acid chains that are found in proteins. By folding into the final three-dimensional shape, these chains form a functioning protein. Proteins are divided into different classes based on their biological function. Proteinclassification is difficult since most proteins have a similar core structure and share a shared evolutionary origin [4]. One method is to write a computer programme that compares an

undefined amino acid sequence to known protein sequences and outputs the appropriate protein classification. Protein identification and analysis are critical since they are required for many of an organism's primary functions.

Knowledge discovery in biological databases

Database processing technologies, artificial intelligence, and statistics are all used in Knowledge Discovery from Databases (KDD). Computational methods have been acknowledged as necessary for combining and analysing diverse knowledge in order to pick genetic variations and features of functional interactions that have a good effect on the biological result of the entire organism. Artificial Intelligence (AI) can aid in the deciphering of ever-increasing and dense biological data [5]. There is already a significant corpus of biological data, and their effective utilisation necessitates the recovery of relevant information.

Computer-Aided Drug Design (CADD)

For the identification, development, and research of medicines and associated physiologically active compounds, computer-aided drug design employs statistical methodologies. CADD is heavily reliant on information technology, databases, and computing tools. Those activities can be successfully handled by AI. At many stages of drug development, such as target recognition, the CADD can be utilised for simulated screening or lead optimization. With several different AI algorithms to choose from, it's become common for researchers to utilise computers that can recognise and exploit their datasets. Scientists want procedures that allow them to obtain data that is labelled with meaning, exact estimations, and descriptions in a logical manner. Artificial intelligence (AI) can be utilised in bioinformatics to model biological data and create new discoveries.

REFERENCES

1. Meiring GA, Myburgh HC. A review of intelligent driving style analysis systems and related artificial intelligence algorithms. Sensors. 2015;15(12):30653-30682.

2. McCain K. Neural networks research in context: A longitudinal journal cocitation analysis of an emerging interdisciplinary field. Scientometrics. 1998;41(3):389-410.

3. Nevo E. Genetic variation in natural populations: patterns and theory. Theor Popul Biol. 1978;13(1):121-177.

4. Lenski RE, Ofria C, Pennock RT, Adami C. The evolutionary origin of complex features. Nature. 2003;423(6936):139-144.

5. Webb RA. Use of the boundary line in the analysis of biological data. J. Hortic. Sci. 1972;47(3):309-319.