## ISSN: 2229-8711

## **Open Access**

# **Bioinformatics and Deep Learning**

#### Annie Marrie\*

Faculty of Bioinformatics, Delft University of Technology, Delft, Netherlands

## Introduction

Three significant advancements in science and technology over the last decade have resulted in the regeneration of ANNs, primarily via DL. First, enormous amounts of data, primarily imaging and natural language data, have been generated in modern life. Other machine learning approaches have struggled with the complexity of deriving knowledge from such data, but ANNs have handled it admirably. High-throughput biological data, such as next-generation sequencing, metabolomics data, proteome data, and electron microscopic structure data, has also posed computational challenges. Second, processing power has been quickly expanding at low prices, thanks to the development of new computing technologies like GPUs and FPGAs.

These devices are suitable hardware platforms for models that require a lot of parallelism. The following are some issues in the discipline of bioinformatics that must be addressed. First, scientists need to be able to grasp how a model helps them address a biological problem, such as predicting DNA-protein binding. Second, the clinical anticipate accuracy of a computational model connected to healthcare or illness diagnosis is 98 percent to 99 percent, and achieving that level of accuracy is difficult [1,2].

## Description

#### **Principled DL's current trend**

Attention mechanism: Attention processes could be utilised to solve a variety of biosequence analysis challenges, including RNA sequence analysis and prediction, protein structure and function prediction from amino acid sequences, and enhancer–promoter interaction (EPI) discovery. EPIVAN's AUROC (area under the ROC) and AUPR (area under the PR curve) values are larger than those without the attention mechanism, indicating that the attention mechanism is more concerned with cell line-specific properties and may better capture hidden information from the perspective of sequences.

**Reinforcement learning:** Collective cell migration, DNA fragment assembly, and describing cell movement are all examples of where reinforcement learning can be used. DNA fragment assembly is an NP-hard optimization problem that seeks to rebuild the original DNA sequence from a large number of fragments by identifying the order in which the fragments must be joined back into the original DNA molecule.

**Few-shot learning:** Many issues in bioinformatics that have minimal data, such as protein function prediction and drug discovery, are well-suited to few-shot learning. The drug discovery task, for example, is to locate analogue compounds with increased pharmaceutical activity and optimise the candidate molecule that can affect critical pathways to produce therapeutic action. It is difficult to make good forecasts for novel compounds because of the scarcity of biological data [3,4].

\*Address for Correspondence: Annie Marrie, Faculty of Bioinformatics, Delft University of Technology, Delft, Netherlands; E-mail: marriean02@tudelft.nl

**Copyright:** © 2022 Marrie A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Received:** 01 April, 2022, Manuscript No.GJTO-22-66950; **Editor Assigned:** 04 April, 2022, PreQC No. P-66950; **Reviewed:** 09 April, 2022, QC No. Q-66950; **Revised:** 14 April, 2022, Manuscript No. R-66950; **Published:** 19 April, 2022, DOI: 10.37421/2229-8711.2022.13.289 **Deep generative models:** Protein structure design, 3D compound design, protein loop modelling, and DNA design are all problems that can be solved using deep generative models. Understanding protein structure and function is crucial to understanding biology at the molecular and cellular levels. However, there may be missing parts that must be reconstructed, and the loop-modelling problem is the prediction of those missing regions. For this challenge, a generative adversarial network (GAN) is used, which can capture the loop region's context and forecast the missing area. The 2D distance map represents the 3D protein structure, with each value representing an actual Euclidean distance between C atoms of two amino acids.

### Typical algorithms and applications

**Recurrent neural network:** Recurrent Neural Network (RNN) is a deep learning model that differs from traditional neural networks in that it can integrate previously learned status using a recurrent approach, namely backpropagation, whereas traditional neural networks typically output predictions based on the current layer's status.

**Convolutional neural network:** Convolutional neural networks (CNN or ConvNet) are well suited to processing data in multiple arrays. The general design philosophy of CNN is to minimise the parameters without affecting its learning power [2]. The backpropagation algorithm is used to train the parameters of each convolution kernel in CNN. CNN can handle pixel scanning and processing, which considerably speeds up the implementation of efficient algorithms in practise, especially in image-related applications.

Auto encoder: Auto encoder may compress and encode information from the input layer into a short code, which it can subsequently decode into an output that closely matches the original input after appropriate processing. Auto encoder is a typical artificial neural network that uses data driven learning to precisely extract coding or representation properties in an unsupervised manner. Because loading all raw data into a network is time-consuming and infeasible for high-dimensional data, dimension reduction or compression is a requirement in pre-processing raw data.

**Deep belief network:** Each hidden layer in subnetworks acts as a visible layer for the following layer in a Deep Belief Network (DBN), which is made up of many Restricted Boltzmann Machines (RBM) or auto encoders stacked on top of each other. DBN uses an unsupervised greedy technique to set network weights layer by layer; subsequently, during fine-tuning, it can use the wake-sleep or backpropagation algorithms. While classical backpropagation, which is utilised in fine-tuning, may experience numerous issues, DBN may not: 1) labelled data is required for training; 2) learning rate is slow; and 3) improper settings lead to the acquisition of a local optimum [5].

**Transfer learning in deep learning:** When there is not enough labelling information or dimensionality, transfer learning is typically used. Transfer learning has attracted attention from deep learning areas due to its transferability of high-level semantic categorization for deep neural networks, despite the fact that it does not belong to deep learning conceptually.

## Conclusion

We covered the basic but essential concepts and methods in deep learning, as well as its latest applications in a variety of biomedical studies, in this paper. By examining common deep learning models such as RNN, CNN, auto encoder, and DBN, we can see that the specific application scenario or context, such as data feature and model applicability, are the most important factors in designing a suitable deep learning approach to extract knowledge from data; thus, figuring out how to decipher and characterise data features is still a difficult task in the deep-learning workflow. Many modifications of classic network models, like the network models described above, have shown in recent deep learning studies that model selection affects the efficacy of deep learning applications.

# Acknowledgement

None.

# **Conflict of Interest**

The authors reported no potential conflict of interest.

## References

- 1. Li, Haoyang, Shuye Tian, Yu Li and Qiming Fang, et al. "Modern deep learning in bioinformatics." *J Mol cell Bio* 11 (2020): 823-827.
- Maljkovic Berry, Irina, Melanie C. Melendrez, Kimberly A. Bishop-Lilly and Wiriya Rutvisuttinunt, et al. "Next generation sequencing and bioinformatics methodologies for infectious disease research and public health: Approaches, applications, and considerations for development of laboratory capacity." *T J Inf Dis* 3 (2020): S292-S307.
- Azad, Rajeev K and Vladimir Shulaev. "Metabolomics technology and bioinformatics for precision medicine." Brief Bioinform 6 (2019): 1957-1971.
- Malathi, Kullappan and Sudha Ramaiah. "Bioinformatics approaches for new drug discovery: A review." Biotech Genetic Eng Rev 2 (2018): 243-260.
- Li, Yu, Chao Huang, Lizhong Ding, Zhongxiao Li, Yijie Pan, and Xin Gao. "Deep learning in bioinformatics: Introduction, application, and perspective in the big data era." *Methods* 166 (2019): 4-21.

How to cite this article: Marrie, Annie. "Bioinformatics and Deep Learning." Glob J Tech Optim 13 (2022): 289.