Active Deep Learning Techniques for Addressing Logical Problems in Bioinformatics

Kiran Sree P[1]* and Usha Devi N[2]

[1]Department of CSE, Shri Vishnu Engineering College for Women, India

[2]Department of CSE, University College of Engineering, India

Corresponding author: Dr P. Kiran Sree, Professor, Department of CSE, Shri Vishnu Engineering College for Women, Bhimavaram, India, Email: drkiransree@gmail.com

Editorial

Biological data is exceptionally immense and unpredictable in nature. This biological data has tremendously helped scientific research immensely; this enormous data must be saved and examined. The most important challenge of researchers and scientists in this field is to examine and interpret the DNA sequences which contain adenine (A), thymine (T), cytosine (C) and guanine (G).

Living organisms consists of cells which consist of numerous complex substances like chromosomes, deoxyribonucleic acid(DNA) and genes that shape the character/identity(heredity) which could be passed on to the different cells. This character/identify serves to figure out certain qualities like what may be cause for disease, hair color, what may be the reason for having six fingers in hand instead of five etc. All these characteristics are referred as genes and the investigation of the genes is known as genetics.

Cell is the fundamental building block of life. Every human body contains more than thirty trillion cells. A huge number of genetic instructions are held in these cells which will make proteins. DNA holds the complete hereditary information which can be used in drug discovery and diagnosis.

A Deep Learning performs reckonings in a distributed manner on a spatially enlarged grid. It contrasts from the conventional approach [1] to parallel processing in which a problem is divided into independent sub tasks; each one is solved by different processor. The result of sub tasks are joined together to yield the final result. The complex nature of DL arises from bottom up manner from the spatially extended grid whose interactions are locally defined. A programmer can specify local rules for interaction among neighbors and can study the consequences [2] of the rules with respect to fitness. The selection of the rules which can be applied to solve a particular problem is challenging in DL. An active learning mechanism [3] can address various problems in Bioinformatics.

The active deep learning mechanism can address various problems like promoter region prediction, protein coding region prediction, translation, protein structure prediction and RNA mapping etc as shown in Figure 1.

This DL approach is the first to integrate AIS with Cellular Automata. Genetic Algorithm with AIS (Clonal Algorithm) is used for evolving MACA. AIS-MACA is the first algorithm to handle DNA sequences of 252,354. AIS-MACA and AIS-PRMACA will give the exact boundary for predicting the protein coding region [4] and promoter region as it is trained to handle real values. Integrated AIS-MACA with AIS-PRMACA will give very faster and give accurate output. This is the first algorithm which addresses the complete transcription phase. Theoretical frame work of HMACA is presented which can address more problems in bioinformatics [5]. The integrated algorithm is trained to predict the corresponding class even in mixed and overlapped sequences. The proposed classifier can handle large data sets.

Figure 1: Logical Structure of Deep Learning Problems.
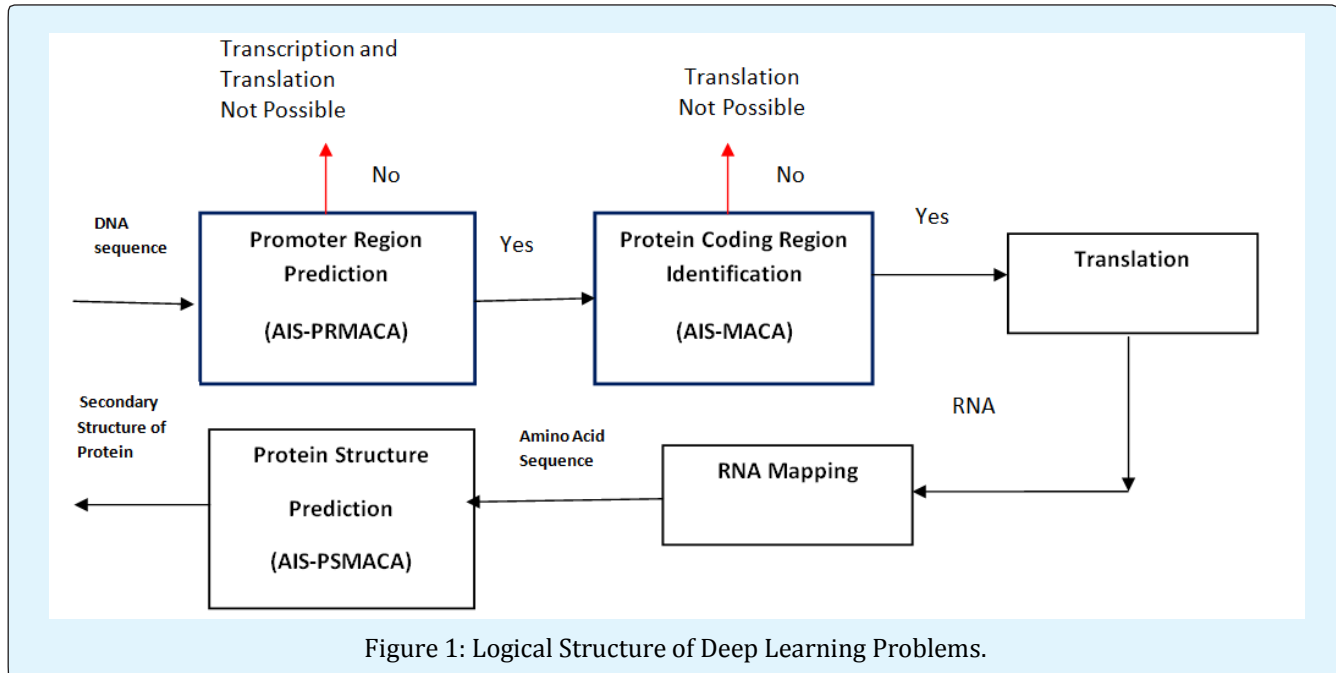
References

1. Kiran Sree P, Ramesh Babu I, Usha Devi N (2009) Investigating an Artificial Immune System to Strengthen the Protein Structure Prediction and Protein Coding Region Identification using Cellular Automata Classifier. International Journal of Bioinformatics Research and Applications 5(6): 647-662.

2. Kiran Sree, Ramesh Babu (2010) Identification of Promoter Region in Genomic DNA Using Cellular Automata Based Text Clustering. The International Arab Journal of Information Technology (IAJIT) 7(1): 75-78.

3. Kiran Sree P, Usha Devi N (2016) Additive Cellular Automata Augmented with Deep Learning for Pattern Reorganization. J Adv Res Comp Tech Soft Appl 3: 1-3.

4. Kiran Sree P, Ramesh Babu I, Usha Devi N (2014) A Fast Multiple Attractor Cellular Automata with Modified Clonal Classifier for Coding Region Prediction in Human Genome, Journal of Bioinformatics and Intelligent Control 3(2): 128-136.

5. Kiran Sree P, Ramesh Babu I, Usha Devi N (2014) A Fast Multiple Attractor Cellular Automata with Modified Clonal Classifier Promoter Region Prediction in Eukaryotes. Journal of Bioinformatics and Intelligent Control 3(2): 123-127.