Deep Learning: The future of IT

Dr. P. Kiran Sree, M. S. Sudheer

Department of CSE,

Shri Vishnu Engineering College for Women,

Bhimavaram.

The development of a potential classifier with Deep Learning towards addressing Promoter, Protein Coding and Protein Structure Prediction was identified and the plan to achieve it was formulated. This research creates an intuition, how interrelated and logical problems in Bioinformatics like protein coding regions, promoter regions, protein structure prediction etc. can be addressed with a common framework.

The proposed approach aims at developing four independent classifiers which can perform their operations individually and also in a combined form. These classifiers AIS-MACA series uses the basic framework of Cellular Automata (CA) and features of the modified CLONAL classifier like self monitoring and non uniformity which is potential, versatile and robust. We use 16-NeighborhoodHybrid Cellular Automata to design and develop our classifiers.

SIGNIFICANCE OF RESEARCH

- Identifying the protein coding region plays vital role in understanding these genes.
- If we identify the promoter region, we can extract information regarding gene expression patterns, cell specificity and development.
- Identifying the quaternary protein structure helps in drug design.

The classifiers that are proposed to achieve our goals are given below.

- AIS-MACA-PCR classifier aims to predict the protein coding regions from 54,108,162,252 and354 length DNA sequences. It has is trained with Fickett and Toung data sets for the DNAsequences of length 54,108 and 162. It is trained with MMCRI data sets for 252 and 354 lengthDNA sequences.
- AIS-MACA-PR classifier is used to predict the promoter regions in Humans, drosophila andArabidopsis thaliana. It is trained and tested with DBTSS, EID, UTRdb data sets for humans, Berkeley Drosophila datasets for drosophila and TAIR Arabidopsis thaliana datasets forArabidopsis thaliana.
- Integrated Classifier IN-AIS-MACA predicts both promoter and protein coding regions in a given DNA sequence. This classifier is thoroughly trained and tested for accuracy with DBTSS, EID,UTRdb, and MMCRI datasets.
- AIS-MACA-PSP classifier is used to predict the quaternary structure of protein from amino acid sequences. This classifier is trained with PDB data sets.