Chapter 24
Crow–ENN:
An Optimized Elman Neural Network with Crow Search Algorithm for Leukemia DNA Sequence Classification

Rehan Ullah

https://orcid.org/0000-0003-0738-9269

The University of Agriculture, Peshawar, Pakistan

Abdullah Khan

https://orcid.org/0000-0003-1718-7038

The University of Agriculture, Peshawar, Pakistan

Syed Bakhtawar Shah Abid

The University of Agriculture, Peshawar, Pakistan

Siyab Khan

https://orcid.org/0000-0003-4455-9466

The University of Agriculture, Peshawar, Pakistan

Said Khalid Shah

Department of Computer Science, University of Science and Technology, Bannu, Pakistan

Maria Ali

The University of Agriculture, Peshawar, Pakistan

ABSTRACT

DNA sequence classification is one of the main research activities in bioinformatics on which, many researchers have worked and are working on it. In bioinformatics, machine learning can be applied for the analysis of genomic sequences like the classification of DNA sequences, comparison of DNA sequences. This article proposes a new hybrid meta-heuristic model called Crow-ENN for leukemia DNA sequences classification. The proposed algorithm is the combination of the Crow Search Algorithm (CSA) and the Elman Neural Network (ENN). DNA sequences of Leukemia are used to train and test the proposed hybrid model. Five other comparable models i.e. Crow-ANN, Crow-BPNN, ANN, BPNN and ENN are also trained and tested on these DNA sequences. The performance of models is evaluated in terms of accuracy and MSE. The overall simulation results show that the proposed model has outperformed all the other five comparable models by attaining the highest accuracy of over 99%. This model may also be used for other classification problems in different fields because it can achieve promising results.

1. INTRODUCTION

In the past, the speed of producing and sharing scientific knowledge was never been so fast as compared to the present era. New disciplines are being raised by combining different fields of science. One such newly arisen field is bioinformatics, which uses statistics, mathematics and computer science in molecular biology to store, analyze and retrieve biological data. Bioinformatics is growing very fast and it has made itself a basic part of any biological research work. Bioinformatics can serve a biologist to excerpt meaningful information from biological data using different kind of web or computer-based tools, most of which are available freely (Mehmood et al., 2014). Among all these computational techniques machine learning is the most common procedure for analyzing data in the form of protein and DNA sequences. Machine learning is a subfield of AI, which is concerned with designing and development of computer algorithms which get improved with experience. The field of machine learning makes computers capable to aid humans in analyzing complex and large problems. In bioinformatics, machine learning can have applied for the analysis of genomic sequences like the classification of DNA sequences, comparison of DNA sequences, Identification of Unknown DNA sequences etc. Supervised and Unsupervised learning are two broad methodologies which are used commonly in machine learning. (Libbrecht et al., 2015). Classification is kind of supervised machine learning which is used to classify every element in a dataset into one of the predefined set of groups or classes. Classification is a function of data mining which assigns elements/items in a collection/dataset to some target classes or categories based on some similarities. Classification is aimed to accurately predict a target group or class for every item in a dataset. There are many techniques used for classification like Support Vector Machine (SVM), Decision Trees, Naive Bayes Classification, Artificial Neural Networks (ANN), Bayesian Networks, etc. (Kesavaraj et al., 2013).

This research has combined two machine learning algorithms namely Crow Search Algorithm (CSA) and Elman Neural Network (ENN). The simple ENN has the problem of being stuck in the local minima. It was not able to reach to a global optimum. So, by merging CSA with ENN, this problem is solved.

This research aims to construct a hybrid technique by combining Crow Search Optimization Algorithm (CSA) with Elman neural network (ENN) for leukemia DNA sequences classification. The proposed model is called Crow-ENN. The objective of the proposed hybrid model to construct the proposed optimized machine learning classification model for Leukemia DNA sequences classification and the performance evaluation of the proposed hybrid algorithm by comparing its Mean Square Error (MSE) and accuracy with the existing models.

This study is aimed to classify Leukemia DNA sequences using the proposed hybrid Crow-ENN model. Datasets of Leukemia DNA sequences are taken from NCBI (National Center for Biotechnology Information) database. Two measures are used for the performance evaluation of the proposed model, which are; MSE and Accuracy.

2. RELEVANT LITERATURE

Sarhan, (2009) developed a stomach cancer identification model that is based on ANN and DCT. The proposed methodology extracts feature from stomach microarrays for classification through Discrete Cosine Transform (DCT). The features which are extracted by DCT coefficients are then used by Artificial Neural Network (ANN) to classify whether it is tumor or non-tumor. Next in 2015, (Adetiba et al.,

Related Content

Techniques for Named Entity Recognition: A Survey
Girish Keshav Palshikar (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications (pp. 400-426).*
www.irma-international.org/chapter/techniques-named-entity-recognition/76075

Classification Tree
(2011). *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations (pp. 53-67).*
www.irma-international.org/chapter/classification-tree/53897

A Latent Feature Model Approach to Biclustering
José Caldasand Samuel Kaski (2016). *International Journal of Knowledge Discovery in Bioinformatics (pp. 11-28).*
www.irma-international.org/article/a-latent-feature-model-approach-to-biclustering/172003

Infer Species Phylogenies Using Self-Organizing Maps
Xiaoxu Han (2010). *International Journal of Knowledge Discovery in Bioinformatics (pp. 29-49).*
www.irma-international.org/article/infer-species-phylogenies-using-self/45164

Caring for our Aging Population: Using CPOE and Telehomecare Systems as a Response to Health Policy Concerns
Sama Al-Khudairy (2014). *Research Perspectives on the Role of Informatics in Health Policy and Management (pp. 153-166).*
www.irma-international.org/chapter/caring-for-our-aging-population/78695