Chapter 2
A Brief Overview on Intelligent Computing-Based Biological Data and Image Analysis

Mousomi Roy
University of Kalyani, India

ABSTRACT

Biological data analysis is one of the most important and challenging tasks in today's world. Automated analysis of these data is necessary for quick and accurate diagnosis. Intelligent computing-based solutions are highly required to reduce the human intervention as well as time. Artificial intelligence-based methods are frequently used to analyze and mine information from biological data. There are several machine learning-based tools available, using which powerful and intelligent automated systems can be developed. In general, the amount and volume of this kind of data is quite huge and demands sophisticated tools that can efficiently handle this data and produce results within reasonable time by extracting useful information from big data. In this chapter, the authors have made a comprehensive study about different computer-aided automated methods and tools to analyze the different types of biological data. Moreover, this chapter gives an insight about various types of biological data and their real-life applications.

INTRODUCTION

In the field of diagnosis and medical research, analysis of biological data is inevitable. We can explore and extract some precious information from various types of biological data. To mine some knowledge from these data requires the involvement of the domain experts. Although humans are highly intelligent, there are some inherent problems associated with the human observers. For example, experiments done by the humans are subject to errors. It may also happen that, same data when diagnosed two times by the same observer may produce the different results. It is purely subjective and depends on the present state of the observer (Chakraborty, Chatterjee, et al., 2017a; Hore et al., 2015).

DOI: 10.4018/979-8-3693-3026-5.ch002

In general, biological sources generates a huge amount of data which is often considered as the big data (Prabha, Rai, & Singh, n.d.). Detailed study of these data is required for generating any useful information from the raw form. Now, humans are hardly capable of handling such a huge amount of data in a stipulated amount of time. This is an another problem associated with the human observers. Moreover, the analysis of these data is significant in the diagnostic field to provide proper treatment. The inaccuracies in the diagnostic process can be very costly in terms of the life of the patients.

Automated systems are highly beneficial in analyzing and mining useful information from the biological datasets. The term 'data mining' is a well-known technology which is used to find hidden information from large databases and big data. The exploration process is often known as 'knowledge discovery'. Now a days, it is very hard to think the world without computers. The technical explosion in the automated systems makes the life easier. Automated systems are highly efficient and can mimic the diagnostic process as made by a human expert. Modern applications can effectively search the ocean of the data and discover useful information with high accuracy and in moderate time (Chakraborty, Chatterjee, Ashour, Mali, & Dey, 2017).

Most of the computer aided diagnostic systems are based on the artificial intelligence (AI) based tools. Artificial intelligence mimics the human intelligence and helps a machine to behave intelligently ("Artificial intelligence," n.d.). In general, it can induce the cognitive nature in a machine. AI based tools provides the power to a computer to perform certain tasks very efficiently like humans (even better in some situation) (Boden, 1998).

In recent years, a huge growth in Artificial intelligence based technologies can be observed that can change the standard of the life. Artificial intelligence provides a way to make a machine learn. This technology is known as machine learning which is the blessing of artificial intelligence. Machine learning algorithms are used to make a machine learn and to avoid programming for each and every problem instance (Mitchell, 1997). These algorithms follows a set of rules for learning as well as for producing results. The performance of these algorithms are measured in terms of the accuracy and some other parameters. The machine learning based approaches can be broadly categorized as given follows:

i. **Unsupervised Learning**: In this type of learning, there is no predefined knowledge about the class of the object. Algorithm tries to find out some pattern so that it can identify the class of the unknown object. The classification operation is based on some hypothesis. To draw a conclusion, all of the hypothesis must be evaluated (Karaboga & Basturk, 2007). There is no so called 'supervision' or guidance for the algorithms, It predicts the natural grouping of the objects.
ii. **Supervised Learning**: Here, some data are used whose classes are known to train the classifier. After the training phase, testing is performed on the known datasets. The deviation of the result from the actual one and predicted one is considered as the error. The main objective of these kind of algorithms is to train the classifier in such a way so that the error will be minimized (Chakraborty et al., 2017). There are several models proposed in the literature that follows the supervised learning methods. For example, Artificial Neural Networks, Support Vector Machine etc. Several hybrid models are found in literature which are based on the metaheurictic algorithms such as GA (Chakraborty, Seal, & Roy, 2015), SA (Chakraborty & Bhowmik, 2015; SChakraborty & Bhowmik, 2013, 2015), CS (S. Chakraborty, Chatterjee, et al., 2017a; Mousomi Roy et al., 2017) to adjust the input weights to enhance the accuracy.