# Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA

**abstraction :** DNA is a biological macromolecule. Its main function is store information,  At present, sequencing technology had caused DNA sequence data to be known, which has also pushed the study of DNA sequences in the wave of big data.  machine learning is a technique for analyzing largescale data and to gain knowledge. It has been widely used in DNA sequence data analysis and made a lot of achievements. Firstly, the review introduces the development process of sequencing technology, on the concept of DNA sequence data structure and similarity Then we analyze the basic process of data mining, summary machine learning algorithms, and put the challenges faced by machine learning algorithms. Then we review four typical applications of machine learning in sequencing data. DNA sequence alignment, DNA sequence classification, DNA sequence clustering, and DNA pattern mining. We analyze their corresponding biological application background. Finally, we summarize the content of the review and look into the future of research directions for the next step.

**Introduction:** DNA, or deoxyribonucleic acid, is the hereditary material in humans and almost all other organisms. Nearly every cell in a person’s body has the same DNA. Most DNA is located in the cell nucleus (where it is called nuclear DNA), but a small amount of DNA can also be found in the mitochondria .DNA sequencing is a laboratory technique used to determine the exact sequence of bases (A, C, G, and T) in a DNA molecule. The DNA base sequence carries the information a cell needs to assemble **protein** and RNA molecules. DNA sequence information is important to scientists investigating the functions of genes.   
For people experiencing a health-impacting condition, **DNA sequencing** can provide a precise diagnosis which might affect the medical management of symptoms, or provide treatment options. Another **advantage of genome sequencing** is that information regarding drug efficacy or adverse effects of drug use can be obtained.