Classifying Cancer Using Machine Learning in Order for CRISPR/Cas9 Technology to Be More Effective

Arora, Shreeya (School: duPont Manual High School)

Artificial Intelligence (AI) is the science of creating machines that can perform tasks which normally require human abilities (e.g. visual perception, speech recognition, decision-making). The goal of this engineering project was to determine the most accurate machine learning (ML) method, which is a type of AI, to use when classifying cancer using gene expression profile data. It was suspected that clustering data with AI could be more efficient than manual methods. ML models are able to analyze patterns and learn them, allowing models to predict accurate results. After hundreds of gene expression profiles were tested with the four coded models, an accuracy score was established. The goal was to determine the number of profiles the models could label correctly. The experiment yielded that the K Nearest Neighbors algorithm, with an accuracy of 96.2%, was the most effective.   The use of AI is a paradigm shift in how cancer classification in the future will be conducted. Based on this experiment, it is possible to use AI to identify which portion of a gene's DNA causes cancer. Research about CRISPR/Cas9, a specific, efficient, and versatile gene-editing technology used to modify, delete or correct precise DNA regions, was conducted. The application of classification techniques using AI can be used in order to increase the accuracy of identifying the precise location at which the DNA should be cut. Using AI to increase the efficiency of this process could potentially aid in the elimination of cancerous conditions from the genome.