Machine learning for precision medicine in cancer: Transforming drug discovery and treatment

Sachin Kumar Deshmukh*

Mitchell Cancer Institute, University of South Alabama, Mobile, AL 36604, USA

Received date: June 05, 2020
Accepted date: June 09, 2020


Machine learning (ML) is a branch of artificial intelligence that uses an algorithm to process the data, retrieve valuable information, learn from it, find a pattern, and make predictions. Manual data analyses suffer from several disadvantages including it is time-consuming and subject to error. The ML tools can repeatedly perform the task with precision. In the age of value-based healthcare, big data, and computer-aided assistance, clinical intervention has become crucial for disease management. Cancer is a major health challenge and the second leading cause of death globally behind cardiovascular diseases [1]. According to a recent study, cancer surpasses cardiovascular disease as the leading cause of death in some high-income and middle-income countries [2]. Cancer is continuously exerting emotional and financial burdens on individuals and families. Even with scientific advancement and improved therapeutic approaches, cancer treatment is still a major challenge. Heterogeneous natures of tumors, compromised drug efficacy and intrinsic and/or acquired resistance developed by tumors attribute to poor clinical outcomes. Remarkable progress in our understanding of cancer biology and insights gained over the year indicate that personalized treatment or precision medicine could lead to better clinical outcomes. Molecular profiling of an individual tumor is crucial for the personalized (precision) medicine. In the analyses of tumor genomic information, ML-based approaches have emerged as a valuable tool. Starting from disease detection, target validation, and identification of prognostic biomarkers, ML approaches are helping in treatment decision making.

A literature survey or text-mining is a helpful tool to get an idea about the field that bridges the knowledge gap between free-text and, valuable information. It also helps in understanding the mechanism(s) of existing drugs and possible reasons behind their compromised efficacy. Cancer cells possess unique characteristics termed as hallmarks that indicate the complexities of neoplastic disease. The cancer hallmarks that differentiate tumor cells from normal cell includes sustaining proliferative signaling, evading growth suppressor, avoiding immune destruction, avoiding apoptosis, inducing tumor-promoting inflammation, activating invasion and metastasis, inducing angiogenesis, genome instability, and mutations, and deregulating cellular energetic [3]. A supervised ML-based approach was utilized to develop the Cancer Hallmarks Analytics Tool (CHAT) which retrieves, organizes thousands of articles and, evaluates the scientific literature on cancer [4,5]. The CHAT arranges the information into different categories that allow the reader to easily analyze and understand the association between the research question and the hallmarks of cancer. To understand the interaction between genes and protein, and the association between environment and cancer several other ML-based tools like OncoCL [6] and OncoSearch [7] were also been developed.

A case study performed to understand the drug mechanisms, ML approaches were able to identify crucial genes (1% of the genome) among the 23,398 genes of the human genome that were altered in triple-negative breast cancer cells treated with metformin [8]. ML-based unsupervised methods identified a cluster of single-cell subpopulations that have considerable differential gene expression which correlated with anticancer pathways. Interestingly, the downregulation of one gene (CDC42) inhibited cancer cell proliferation and motility, indicating the potential of ML to...
identify biologically-relevant candidates [8]. Considering the large size of the human genome that requires resources, time, and cost for analyses, clustering important genes that can be a drug target with the help of ML-based algorithms can expedite the drug discovery program. Also, combinations of cancer drugs can lead to increased efficacy compared to single-drug treatments. Besides, the toxicity and adverse side effects due to off-target are also less as the doses of drug combinations are relatively lower. ML offers the possibility to explore the synergistic efficacies of drugs. The ML predictive models can be developed from high-throughput screening drug synergy data that provide an opportunity to validate them in in vivo settings. Several ML approaches including Random Forests and Naïve Bayes methods are being utilized to predict the synergy between anti-cancer drugs which predict signaling pathways and targets [9,10].

The sequencing of the human genome has improved our understanding of the signaling pathways associated with human diseases. Genome sequencing and data analyses are critical for precision medicine that is a promising area of modern science. ML has been proved its efficacy in automated analysis of patient gene and protein data. Huang et al. developed an ML support vector machine algorithm combined with standard recursive feature elimination to predict precision medicine responses. The model was developed from gene expression profiles and drug response data of the National Cancer Institute panel of 60 human cancer cell lines [11]. The model demonstrated its efficacy in predicting the drug responsiveness of a variety of cancer cell lines. Clinical trials demand time and cost; thus, predictive models are valuable resources. ML-based models help stratify patients, identify drug efficacy, and suggest the mechanisms of drug action.

Advancement in microscopy technology enabled the generation of a large amount of phenotypic data from the cells and tissue. Analyses of large, complex data demand automated and precise analytical strategies that can differentiate and pick altered cellular phenotype from the wild type phenotype. General statistical methods such as mean, mode, median often remain insufficient to illustrate the data complexity. Supervised or unsupervised ML strategies have shown their potential in analyzing the variations in phenotypic changes of single cells or populations of cells [12,13]. ML has been used for image-based genetic screening experiments to organize single cells based on their phenotypic profiles and biologically meaningful classes [14]. Kraus et al. used microscopy and automated image analysis with deep convolutional neural networks to differentiate S. cerevisiae proteins by their subcellular localization [15], suggesting that deep learning can be a handy tool for rapid data analyses.

A good amount of data coupled with algorithm validation is crucial for model performance and precise prediction. Availability of huge biological and medical data, along with sophisticated ML algorithms, the design of an automated drug development framework can now be envisioned. ML has demonstrated its potential in cancer risk prediction, cancer diagnosis, and predicting therapeutic outcomes. Also, ML methods have helped in identifying signaling pathways and targets critical for pharmacological modulation and drug development. ML-based approaches can successfully arrange genomic data of individual patient tumors in perspective and predict drug responses. The global healthcare community is continuously working to improve the ML performance by developing novel algorithms and innovative strategies. However, the success of ML depends on reliable data from different healthcare streams including research laboratories, drugs and pharmaceutical companies, clinical trial companies, FDA, and others. Collaboration, cooperation, and coordination between these organizations will further push the ML field forward. In the past decade, ML has moved from theoretical studies to real-world applications. We are slowly moving towards the era of precision medicine and ML approached seems to speed up the drug development program.

References


