EXCITING ATTRIBUTES OF MACHINE LEARNING IN BIOMEDICAL RESEARCH AND TECHNOLOGY

C. Sandhya, S. Hemalatha*

Abstract

The role of Machine Learning in the field of Biomedical Research and Technology is of hot deliberation. This paper aims to discuss the various applications of machine learning in biomedical research and technology and how it can transform the three broad areas of it; personalized treatment and health monitoring, precision medicine, and clinical diagnostics. How machine learning can be employed in each area, as well as the opportunities and challenges in the current scenario is discussed. Once these challenges are overcome, Machine learning promises a future in which medical practitioners can provide personalized treatment and outcome-based medicines with high precision detection and accurate diagnosis. Various aspects of these techniques are to be continuously monitored and adapted according to the needs of individuals with environmental differences.

Keywords: Healthcare, Supervised Machine Learning, Disease Prediction

I INTRODUCTION

Machine learning is currently the most discussed terminology that has been put into various fields including Biomedical Technology. The vast amount of information is generated for a decade was quite phenomenal. Earlier the amount of data unearthed was frugal compared to the present situations where the complexity and volume of data have increased tremendously. This has necessitated the need for sophisticated computational algorithms that make repetitive and complex tasks possible. The overwhelming situation has been quite competently managed by the institution of Machine Learning and Artificial Intelligence. This enables the dawn of technological excellence which will shape the future of biomedical research.

Department of Computer Science,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
*Corresponding Author

According to Arthur Samuel, Machine learning (ML), is a "field of study that gives computers the ability to learn without being explicitly programmed".

A Machine Learning algorithm is able to learn from data. But, how do we define learning for a machine?

According to Tom M.Mitchell [1],a machine learning algorithm can be formally defined as "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

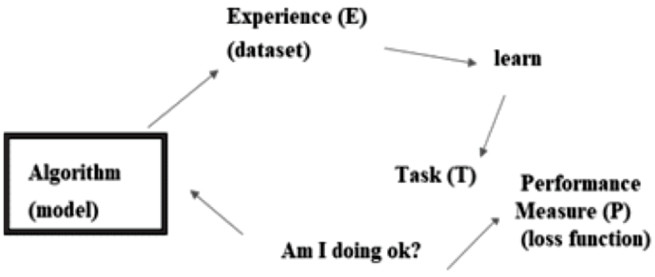

Fig1:The Mitchell Paradigm

The following narration exemplifies this situation. Consider our immune system which has special cells called T-cells and B-cells to protect us from outside invaders like viruses, bacteria, fungi, and other pathogens. When an antigen enters our body, a special cell in our immune system called Antigen-Presenting Cells (APC) detects and engulfs them. They form several fragments of the antigen and transport them to the periphery of the cell membrane.

Out of this, a few T-cells respond to APC by invoking B-cells which produce proteins called antibodies which are specific antidotes to the invading pathogens. These antibodies are bound to the surface of the invading pathogen which marks its destruction. After infection, pathogen-specific B-cells are present in the immune system to provide a quick response in case of reinfection. Just like our immune system which has learned to identify the antigens as pathogenic and use this information to protect our body in case of reinfection, the machines learn from their experience to perform tasks better in subsequent counters. The repeated exposure makes the more agile and responsive to perform better.

II IMPACT OF MACHINE LEARNING IN BIOMEDICAL RESEARCH

Machine learning (ML) is a rapidly advancing technology in the field of medicine for integrating computer science and statistics to solve medical problems. Machine learning can deal with a multitude of deterrent factors affecting data such as enormity of volume, variability, complexity, heterogeneity, etc., in the field of biomedicine which cannot be properly analyzed using conventional statistical methods. The reports from the Kinsey Global survey, big data, and machine learning are generating pet bytes of data worth $100 billion in the field of pharma and medicine [2]. This has opened new venues in biomedical research as a result of enhanced decision making, improving the efficiency of research methods, and using high-performance computational algorithms. Machine learning finds its application in the key areas of biomedicine as discussed.

Identification of Disease:

Machine Learning has proved to be a highly successful tool in identifying diseases. The statistical reports from major Pharmaceutical Research and Manufactures from America in 2015 have shown that more than 800 medicines and vaccines to treat cancer were on trial. Machine Learning has a great role to play in the diagnosis of major diseases like cancer, cardio vascular diseases, and neuro degenerative diseases. By identifying and summarizing meaningful patterns from heterogeneous datasets machine learning has proved to be a powerful tool that can help medical practitioners and biomedical scientists. IBM Watson Health in October 2016 set up a project “IBM Watson Genomics” in partnership with Quest Diagnostics [3].This paved the way for several achievements in precision medicine by concocting genome tumor sequencing and cognitive computing.

(a) Personalized Treatment: Another major application of Machine Learning is in Personalized Health Management. Personal data collected from various sources like homes, clinics, etc., were combined and subjected to predictive models that are built using Supervised Machine learning has enhanced the quality of treatment. Recent studies have revealed that the sensory data (e.g.,voice, accelerometer) collected from smart devices are processed using machine learning can be used to monitor symptoms and progression of neurodegenerative diseases like Parkinson’s disease [4].IBM Watson Oncology is a leading institution in this domain that makes use of patients’ personal information and history for optimizing the treatment option. Thus it drastically brings down the cost of healthcare.

(b) Drug Discovery: Machine learning has created several positive trends in the discovery of drugs. It engages R&D technologies like next-generation sequencing. Precision medicine helps in providing customized treatment for patients. Patient’s genetic and molecular information are used by this diagnostic testing to devise optimal therapies. The MIT Clinical Machine Learning Group is a pioneer in this domain. Their precision medicine research is focused on developing algorithms that can modify the treatment for Type2 diabetes.

© **Clinical Trial Research:** Machine learning has high potential in the field of clinical trial research. It helps in identifying candidates from various sources like social media, doctor visits, genetic information, etc., This will help in formulating the clinical trials and thereby minimizing the cost, time, and effort. Machine learning also facilitates the monitoring of real-time data from remote locations thus improving safety; for example, monitoring biological and other vital signs can help in preventing medical catastrophes.

(d) Epidemic Out Break Prediction: Data collected from web, satellites, social media and other sources are subjected to Machine Learning and Artificial Intelligence-based analysis for monitoring and predicting possible out breaks of the epidemic around the world, especially the third world countries.

These countries have limited health care infrastructure, educational avenues, and access to treatments. Pro MED-mail is an internet-based reporting program which has proved to be highly successful in monitoring the emerging diseases and providing out break reports in real-time. This has a significant emphasis in the present context of the pandemic Covid-19. Several lessons have been learned and are still for the coming.

III POTENTIAL CHALLENGES FOR MACHINE LEARNING IN BIOMEDICAL RESEARCH

Machine Learning has a prominent role in the Biomedical Research field but it has several limitations to overcome to make it worthwhile. The following are a few instances of the challenges faced by biomedical researchers.

1. Heterogeneity of Data: Heterogeneity of data owing to integration of multi-omics study poses a formidable challenge in the field of biomedical research. Earlier studies involved data from a single source (eg:genomics)and were examined using conventional statistical methods.

Integrative study of clinical data and multi-omics results in new advancements in the field of biomedical research and precision medicine. For example, Glioblastoma Multi form is a highly aggressive type of brain cancer. Its prognostic prediction can be enhanced by considering multiple factors [5] like clinical data, gene expression, DNA methylation, and copy number alterations.

But this data integration introduces new computational challenges associated with single-omics studies. Machine Learning encompasses a variety of efficient algorithms like Decision Trees, Random Forest (Tree-based learning) Bayesian, Multi-task MKL (Multiple- Kernel Learning), Clustering and Page Rank(CPR),Deep Neural Network Synergy model with Auto encoders(AuDNN synergy)[6] which are based on Deep Learning for integrative analysis of heterogeneous data. Even though it has contributed favorably in this complex analysis, it still has to undergo significant transformations to emerge as a novel method for accurate data prediction involving heterogeneous multi-faceted propositions.

2. Curse of Dimensionality: Compared to the biological sample of the study the integrative analysis involving multi-omics data has a greater number of variables or features. For example, in the study for classifying of different cancer patients that is based on factors like DNA methylation, and gene expression measurements, the number of variables is more than the number of samples (thousands of variables measured on just a few hundred patients).The increased dimensionality in the number of variables, with the same sample size, make the most machine learning methods prone to the problem of over fitting i.e., when the model fits perfectly with the training data.[7]. Some of the issues that arise due to high dimensional data manifest during the analysis and visualization of data for pattern identification. Few of them appear while training machine learning models[8]. This difficulty related to training machine

learning mode is due to high dimensional data is referred to as the "Curse of Dimensionality".

Feature Selection (FS) and Feature Extraction (FE)[9][10]methods are designed to curb this situation. The other problems associated with the application of Machine Learning in Biomedical Research include the lack of standardization, privacy issues, complexity in managing pattern recurrence and regularities, use of multiple platforms form measurement, interpretability, understandability, and reproducibility issues related to data management.

IV CONCLUSION AND FUTURE PERSECTIVES

Machine learning is the latest and most powerful tool in the field of Computer Science.

A multitude of complex and diverse data sets from various sources are analyzed to evolve patterns that can be employed in the health care sector especially in personalized medicine. Precision medicine and personalized treatment can be effectively achieved through a combination of machine learning techniques and data gathered through specialized biosensors [8].

The data retrieved should be comprehensive and of superior quality in order to attain higher levels of accuracy and precision. The short comings of machine learning in the multi-omics integrative analysis have to be addressed, remedial measures are to be invoked and novel machine learning approaches are to be designed to dispel the inadequacies to herald the rise of a new era in the field of biomedical technology.

REFERENCES

[1] Yoshua Bengio,Aaron Courville(2016). "Deep Learning". The MIT Press

[2] Dutta Pramanik, Pijush & Pal, Saurabh & Mukhopadhyay, Moutan (2020). "Big Data and Big Data Analytics for Improved Healthcare Service and Management" International Journal of Privacy and Health Information Management.8.13-51.10.4018/IJPHIM.2020010102

[3] "Bringing Precision Medicine to Community Oncologists." Cancerdiscoveryvol.7,1 (2017): 6-7.

[4] Arora et al., Espay et al., Ginis et al., Pereira et al., (2016) "To monitor symptoms and progression of Parkinson's disease"

[5] Zhang, Y.; Li, A.; Peng, C.; Wang, M (2016)"Improve glioblastomamulti-for meprognos is prediction by using feature selection and multiple kernel learning". IEEEACM Trans. Comput. Biol. Bio inform. 13,825–835

[6] Zhang, T.; Zhang, L.; Payne, P.R.; Li, F.(2018) "Synergistic Drug Combination Prediction by Integrating Multi-omics Data in Deep Learning Models"

[7] De Meulder, B.; Lefaudeux, D.; Bansal, A.T.; Mazein, A.;Chaiboonchoe, A.;Ahmed, H.; Balaur, I.; Saqi, M.; Pellet,J.;Ballereau, S(2018)"A computational framework for complex disease stratification from multiple large scale datasets".BMCSyst.Biol

[8] Goecks, Jeremyetal.(2020)"How Machine Learning Will Transform Biomedicine." Cell vol. 181,1:92 -101.

[9] Hira, Z.M.; Gillies, D.F. (2015) "Are view of features election and feature extraction methods applied on micro array data".Adv. Bioinform

[10] Guyon,I.;Elisseeff(2003) "An introduction to variable and feature selection".