ELEVATING NEXT GENERATION GENOMIC SCIENCE AND TECHNOLOGY USING MACHINE LEARNING IN THE HEALTHCARE INDUSTRY

Editors:

Roohi Sille
Tanupriya Choudhury
Sonal Talreja
S. Balamurugan

Bentham Books

Applied Machine Learning for IoT and Data Analytics

(Volume 4)

Elevating Next Generation Genomic Science and Technology using Machine Learning in the Healthcare Industry

Edited by

Roohi Sille, Tanupriya Choudhury,
Sonal Talreja
School of Computer Science
University of Petroleum and Energy Studies (UPES)
Bidholi Campus
Dehradun, Uttarakhand
India

&

S. Balamurugan
Research and Development, Intelligent Research
Consultancy Services (iRCS), Coimbatore
Tamil Nadu, India

Applied Machine Learning for IoT and Data Analytics

(Volume 4)

Elevating Next Generation Genomic Science and Technology using Machine Learning in the Healthcare Industry

Editors: Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan

ISBN (Online): 979-8-89881-549-3

ISBN (Print): 979-8-89881-550-9

ISBN (Paperback): 979-8-89881-551-6

Published by Bentham Science Publishers Pte. Ltd. Singapore, in collaboration with
Eureka Conferences, USA.

First published in 2026.

BENTHAM SCIENCE PUBLISHERS LTD.

End User License Agreement (for non-institutional, personal use)

This is an agreement between you and Bentham Science Publishers Ltd. Please read this License Agreement carefully before using the ebook/echapter/ejournal (**"Work"**). Your use of the Work constitutes your agreement to the terms and conditions set forth in this License Agreement. If you do not agree to these terms and conditions then you should not use the Work.

Bentham Science Publishers agrees to grant you a non-exclusive, non-transferable limited license to use the Work subject to and in accordance with the following terms and conditions. This License Agreement is for non-library, personal use only. For a library / institutional / multi user license in respect of the Work, please contact: permission@benthamscience.org.

Usage Rules:

1. All rights reserved: The Work is the subject of copyright and Bentham Science Publishers either owns the Work (and the copyright in it) or is licensed to distribute the Work. You shall not copy, reproduce, modify, remove, delete, augment, add to, publish, transmit, sell, resell, create derivative works from, or in any way exploit the Work or make the Work available for others to do any of the same, in any form or by any means, in whole or in part, in each case without the prior written permission of Bentham Science Publishers, unless stated otherwise in this License Agreement.
2. You may download a copy of the Work on one occasion to one personal computer (including tablet, laptop, desktop, or other such devices). You may make one back-up copy of the Work to avoid losing it.
3. The unauthorised use or distribution of copyrighted or other proprietary content is illegal and could subject you to liability for substantial money damages. You will be liable for any damage resulting from your misuse of the Work or any violation of this License Agreement, including any infringement by you of copyrights or proprietary rights.

Disclaimer:

Bentham Science Publishers does not guarantee that the information in the Work is error-free, or warrant that it will meet your requirements or that access to the Work will be uninterrupted or error-free. The Work is provided "as is" without warranty of any kind, either express or implied or statutory, including, without limitation, implied warranties of merchantability and fitness for a particular purpose. The entire risk as to the results and performance of the Work is assumed by you. No responsibility is assumed by Bentham Science Publishers, its staff, editors and/or authors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products instruction, advertisements or ideas contained in the Work.

Limitation of Liability:

In no event will Bentham Science Publishers, its staff, editors and/or authors, be liable for any damages, including, without limitation, special, incidental and/or consequential damages and/or damages for lost data and/or profits arising out of (whether directly or indirectly) the use or inability to use the Work. The entire liability of Bentham Science Publishers shall be limited to the amount actually paid by you for the Work.

General:

1. Any dispute or claim arising out of or in connection with this License Agreement or the Work (including non-contractual disputes or claims) will be governed by and construed in accordance with the laws of Singapore. Each party agrees that the courts of the state of Singapore shall have exclusive jurisdiction to settle any dispute or claim arising out of or in connection with this License Agreement or the Work (including non-contractual disputes or claims).
2. Your rights under this License Agreement will automatically terminate without notice and without the

need for a court order if at any point you breach any terms of this License Agreement. In no event will any delay or failure by Bentham Science Publishers in enforcing your compliance with this License Agreement constitute a waiver of any of its rights.

3. You acknowledge that you have read this License Agreement, and agree to be bound by its terms and conditions. To the extent that any other terms and conditions presented on any website of Bentham Science Publishers conflict with, or are inconsistent with, the terms and conditions set out in this License Agreement, you acknowledge that the terms and conditions set out in this License Agreement shall prevail.

Bentham Science Publishers Pte. Ltd.
No. 9 Raffles Place
Office No. 26-01
Singapore 048619
Singapore
Email: subscriptions@benthamscience.net

CONTENTS

PREFACE

Genome mapping, a critical aspect of genomics, has seen transformative advancements with the advent of machine learning techniques. These innovative methodologies have redefined the way we analyse, interpret, and leverage genetic data. The book, "Elevating Next Generation Genomic Science and Technology using Machine Learning in the Healthcare Industry," aims to provide an in-depth exploration of the intersection of machine learning and genomics. It serves as a comprehensive guide for researchers, students, and professionals looking to harness the power of artificial intelligence in genomics.

The revolutionary capabilities offered by the field of Machine learning make it a powerful tool in genomic mapping. A human genome, made up of billions of base pairs, is a repository of vast amounts of information. This information plays a crucial role in developing a knowledge base for understanding diseases, genetic traits, and evolutionary processes. Handling the complexity and scale of genomic data has been a challenging task for most of the traditional methods.

In the context of genome mapping, machine learning algorithms are capable of processing genomic sequences, identifying patterns, and extracting meaningful insights in a more efficient and accurate manner than conventional approaches, offering enhanced accuracy, improved speed, and enabling personalised medicine.

Chapter 1: Overview of Next Generation Genomic Science and Technology Utilising Machine Learning in Healthcare

This chapter focuses on presenting an overview of genomic science and technology. It highlights the utilisation of various ML techniques in this field. It also delves into the transformation observed in healthcare with the introduction of various ML techniques in genomic science.

Chapter 2: A Review of AI-Driven Genomic Approaches for Cancer Detection and Personalized Medicine

The exponential increase in the demand for better and more effective ways to detect cancer demands exploration of new techniques. Artificial Intelligence is proving to be a strong candidate in offering such support. This chapter presents a review of the integration of AI into genomic research to enhance cancer detection, predict cancer risks, and provide personalized treatment plans.

Chapter 3: Machine Learning Approaches for Microbiome Analysis and Applications

ML techniques can play a major role in the improvement of human health. It can be used to analyse the structure of human microbiomes and how it can be used to enhance human health. This chapter focuses on the role of ML techniques for microbiome analysis and applications.

Chapter 4: Machine Learning Model for Annotation and Assembly of Genomes

This chapter focuses on the role of machine learning in genome annotation and assembly. It emphasizes how the ML models are used to decode complex genomic sequences, identifying genes and mutations. This understanding will greatly enhance the accuracy and efficiency of genome research.

Chapter 5: Extending Bayesian Classification to Predict Phase Transition in Biopolymer

This chapter explores the extension of Bayesian Classification techniques to predict the phase transitions in biopolymers. It discusses how this framework can help model the behaviour of biopolymers, which in turn can help the researchers gain better understanding of the structural changes for future drug development.

Chapter 6: Ethical Aspects of Analysing Genomic Data for Medical Analysis: A Comparative Study with a Focus on India

This chapter discusses the various ethical aspects of analysing genomic data for medical analysis. It covers various ethical topics, including patient consent and data privacy, and focuses on the regulatory landscape in India to ensure ethical practices in genomic medicine.

Chapter 7: Integrating Pharmacogenomics and Machine Learning in Personalised Treatment Strategies for Parkinson's Disease

In this chapter, the focus is on the integration of pharmacogenomics and machine learning techniques in suggesting personalised treatment strategies for diseases like Parkinson's. It also explains how ML algorithms can help analyse genetic information for better drug prediction while minimising side effects. This can lead to more effective management of this disease.

Chapter 8: Analysing Microbiome Metagenomics to Detect Anomaly using Variational Autoencoders for Predicting Crohn's Disease

This chapter focuses on the use of variational autoencoders for analysing microbiome metagenomics data, which can help in predicting Crohn's disease. It discusses various models that can potentially help in the early diagnosis of the disease.

Chapter 9: Leveraging ML Models for the Future of Genomes

The AI-driven approaches have enabled the rapid analysis of vast genomic datasets, improved diagnosis of various diseases, and drug discoveries leading to personalised medicines. This chapter explains the role of AI and ML in genetics, exploring their applications along with their benefits, challenges, and future prospects.

Chapter 10: Conclusion: The Transformative Role of Machine Learning in Genomic Science and Healthcare

This chapter presents the book's overall conclusion, including a summary of all previous chapters and how machine learning techniques have played a transformative role in Genomic Science and Healthcare.

Roohi Sille, Tanupriya Choudhury, Sonal Talreja
School of Computer Science
University of Petroleum and Energy Studies (UPES)
Bidholi Campus
Dehradun, Uttarakhand
India

&

S. Balamurugan
Research and Development, Intelligent Research
Consultancy Services (iRCS), Coimbatore
Tamil Nadu, India

Acknowledgements

The editorial team extends heartfelt gratitude to their institution for providing an encouraging research atmosphere that laid the groundwork for this proposal. We are deeply appreciative of the diverse group of contributors from various nations, and we offer special thanks to the reviewers worldwide who have diligently scrutinised each chapter to uphold the book's high standards. Their insightful feedback has been indispensable. Our sincere appreciation goes out to all involved parties for their dedication and readiness to undertake tasks that stretched them beyond their usual comfort zones. We eagerly anticipate reuniting with you in the forthcoming edition of our publication.

List of Contributors

Aditya Vardhan	Department of Physics and Computer Science, Dayalbagh Educational Institute, Agra, UP, India
Akhilesh Singh	School of Computer Science, UPES, Dehradun, India
Amarjeet Singh Chauhan	Department of Physics and Computer Science, Dayalbagh Educational Institute, Agra, UP, India
Ankit Vishnoi	Department of Computer Science and Engineering, Graphic Era Deemed to Be University, Dehradun, India
Charu Kathuria	Department of Computer Science and Engineering, PES University, Bengaluru, Karnataka, India
Deepika Raina	Graphic Era Hill University, Dehradun, Uttarakhand, India
Deepti Mehrotra	Jaypee Institute of Information Technology, Noida, India
G. Pius Agbulu	Department of Mechatronics Engineering, Chinhoyi University of Technology, Chinhoyi, Zimbabwe
Garima Sharma	Department of Botany, SMS Govt. Model Science College, Gwalior, India
Gutta Geethika	Department of AIML, New Horizon College of Engineering, Bengaluru, India
Luxmi Sapra	Graphic Era Hill University, Dehradun, India
Navnit Kumar Misra	Department of Physics, Brahmanand College, The Mall, Kanpur, India
Nikita Kukreti	School of Computer Science, UPES, Dehradun, India
P.S. Akshatha	Department of AIML, New Horizon College of Engineering, Bengaluru, India
Parul Madan	Department of Computer Science and Engineering, Graphic Era Deemed to Be University, Dehradun, India
Prachi Mall	Graphic Era Hill University, Dehradun, Uttarakhand, India
Raghavi D. Kumar	Department of AIML, New Horizon College of Engineering, Bengaluru, India
Ramesh Prasad	Department of AIML, New Horizon College of Engineering, Bengaluru, India
Roohi Sille	School of Computer Sciences, University of Petroleum and Energy Studies (UPES), Dehradun, Uttarakhand, India
S. Balamurugan	Research and Development, Intelligent Research Consultancy Services (iRCS), Coimbatore, Tamil Nadu, India
S.M. Dilip Kumar	Department of CSE, University Visvesvaraya College of Engineering, Bengaluru, India
Sagar Sharma	George Brown College, Toronto, Canada
Samir Jaiswal	School of Computer Science, UPES, Dehradun, India
Samrath Prakash	School of Computer Science, UPES, Dehradun, India

Sanjay Saini	Department of Physics and Computer Science, Dayalbagh Educational Institute, Agra, UP, India
Sonia D'souza	Department of AIML, New Horizon College of Engineering, Bengaluru, India
Sumit Bhardwaj	(Global Analytics), Concentrix Daksh Services India Pvt Ltd, Chandigarh, India
Tanupriya Choudhury	School of Computer Sciences, University of Petroleum and Energy Studies (UPES), Dehradun, Uttarakhand, India
Varun Sapra	ASET, Amity University Punjab, Mohali, India

CHAPTER 1

Overview of Next Generation Genomic Science and Technology Utilising Machine Learning in Healthcare

Garima Sharma[1,*]

[1] *Department of Botany, SMS Govt. Model Science College, Gwalior, India*

Abstract: Genomics is an interdisciplinary field of molecular biology that focuses on studying the structure and function of the gene as well as its relation to gene mapping and tailoring of the genome. The techniques of genomics have led to a revolution in the area of clinical medicine and public health. Genomics plays a critical role in health science by identifying targeted genes, precision in treatment, and drug therapies for patients. The tremendous progress in genomics in the healthcare field is due to the high speed, low cost, increased accuracy, and improved sensitivity of its techniques. The next generation of genomics is marked by recent advances in the field, which utilise computational tools, like artificial intelligence and machine learning. Machine Learning (ML) allows computers to learn, filter, and classify data without human intervention. The Next Generation genomic science technology has major applications in the fields of early detection of tumour gene analysis, non-invasive prenatal screening, and genomic tests for childhood and rare disorders. In this article, we have discussed some recent advances in Next-Generation Machine Learning techniques (NGML) in healthcare and summarised their applications in some recent genomic science techniques like Nanopore sequencing, CRISPR-Cas system, and Spatial transcriptomes technique.

Keywords: Artificial neural network, CRISPR-Cas system, Deep learning, Machine learning, Medical image analysis, Spatial transcriptosome.

INTRODUCTION

The study of genes of an individual and its analytical applications in healthcare are the most advanced fields of medicine, collectively known as 'Genomic Medicine.' Genomic medicine helps with the diagnosis, treatment, and prevention of disease, and it is gradually revolutionizing the scenario in healthcare from pre-

---

[*] **Corresponding author Garima Sharma:** Department of Botany, SMS Govt. Model Science College, Gwalior, India; E-mail: sharmagarimas@gmail.com

Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan (Eds.)

dicting diseases to finding cures. Especially after the COVID-19 outbreak, genomic medicine has proved its marvellous role in delivering next-generation breakthroughs in designing diagnostic kits, drug delivery systems, and gene therapies.

Currently, healthcare is experiencing a paradigm shift as there is a continuous rise in the availability of health records of patients in electronic forms known as Electronic Health Records (EHRs) and the growing advances in techniques of medical imaging. 'Machine learning' in healthcare has become one of the most popular buzzwords these days as it helps healthcare professionals to utilise artificial intelligence for providing better care for patients and managing their clinical data. Machine learning is showing a positive impact in this field, although its full potential is still not realised. In the coming years, the use of Machine learning in healthcare will become even more essential and revolutionary due to continuously growing clinical data sets.

Machine Learning (ML) has emerged as a key enabler of this transformation, with applications ranging from disease diagnosis to population health management. ML is the process of programming and training computers to learn and analyse the given clinical details, to understand the patterns of disease, and to recommend treatments. The healthcare professionals are gradually adapting and recognising the potential of machine learning to reduce workload, cost of treatment, risks, and improve decision-making in this field. The demand for a skilled workforce in the area of machine learning for healthcare industries is continuously on the rise.

The two terms, "artificial intelligence" and "machine learning," might seem to be the same. Artificial Intelligence (AI) is a specialised part of computational technology that creates computer machines that can partially or fully replicate the intelligence of the human brain in various fields. The cleaning robots, self-driven cars, *etc.*, are examples of AI. Machine learning is a type of AI application in which a computer is trained to learn and improve from observations of patterns and correlations without direct instruction. In this way, ML can allow a computer to make predictions and take decisions like human experts. Machine learning applications improve with use and become more and more accurate as they get access to more data. With the advent of technology, expansion in medical data and its analysis tools, the machine learning technique is, day by day, providing exciting opportunities in healthcare.

The following are some of the benefits of machine learning applications in the healthcare industry:

Improving Diagnosis: Machine learning can be used to design and develop better diagnostic tools to make quicker and more accurate analysis of medical images like X-rays or MRI. With the help of machine learning, various patterns in the scan are identified that indicate a particular disease. This is very important in case of fast-progressing diseases to improve patient care.

Developing New Treatment Methods: With the help of machine learning, healthcare organisations and pharmaceutical companies can identify relevant information by analysing patterns in medical data of unknown/ rare side effects of drugs, thus improving patient care, drug discovery, safety, and effectiveness of drugs or medical procedures.

Reducing Healthcare Costs: Machine learning technologies have significantly improved the efficiency of healthcare services by reducing the time and resources required for diagnosis. Additionally, better algorithms are being developed for managing patient consultations and health records.

Increasing Patient Data Security: The privacy of patients' Electronic Health Records (EHR) is of paramount importance. For this, Machine learning is very useful as it can detect cyber threats in real-time and manage them. New algorithms can identify abnormal/ non-specific patterns to identify any breach to cybersecurity, thus ensuring the protection of patient data.

The key areas of healthcare in which machine learning has played an important role are shown in Fig. (**1**) [1]:

Digital services	**Smart Documents**	**Smart Care**	**Smart Units**
• AI cloud • Cloud data system	• Smart reports • Discharge dicuments	• Reduced cost • Better follow up	• Fit bits • smart watches

Fig. (1). Key areas of healthcare affected by machine learning.

NEXT GENERATION MACHINE LEARNING (NGML) TECHNIQUES

The next generation of ML is poised to further transform the field of healthcare by leveraging advances in data analytics, computing power, and algorithmic complexity. Many techniques that may be included in next-generation machine learning, and their applications in healthcare, are discussed in the following section:

A. Deep Learning (DL): It is an advanced type of machine learning that can be used to analyse a bulk of complex data and patterns with minimal human intervention. A traditional machine learning has the following steps: a. data

CHAPTER 2

A Review of AI-Driven Genomic Approaches for Cancer Detection and Personalised Medicine

P.S. Akshatha[1,*], **Gutta Geethika**[1], **Raghavi D. Kumar**[1], **Ramesh Prasad**[1], **Sonia D'souza**[1], **S.M. Dilip Kumar**[2] and **G. Pius Agbulu**[3]

[1] *Department of AIML, New Horizon College of Engineering, Bengaluru, India*

[2] *Department of CSE, University Visvesvaraya College of Engineering, Bengaluru, India*

[3] *Department of Mechatronics Engineering, Chinhoyi University of Technology, Chinhoyi, Zimbabwe*

Abstract: Cancer is a genetically driven disease characterised by mutations that lead to abnormal cell growth and division. Genomic science is essential in identifying these mutations. However, recent advances in AI (artificial intelligence) and DL (deep learning) have significantly enhanced complex genomic data analysis, making cancer detection more precise and efficient. AI is not just a tool but a beacon of hope, enabling early detection, diagnosis, and personalized treatment strategies. This survey comprehensively reviews state-of-the-art AI methodologies applied in cancer genomics, focusing on supervised learning, unsupervised learning, and transfer learning techniques. We examine how AI-driven approaches significantly enhance the identification of cancer-related mutations, improve gene expression profiling, and support the development of targeted therapies. Moreover, the survey underscores the potential of AI not only to reduce false positives and negatives but also to streamline diagnostic workflows and, most importantly, predict patient outcomes in real-time. This real-time prediction capability of AI is a significant step forward in cancer care, providing a high level of reassurance to both patients and healthcare professionals about the accuracy of AI-driven cancer diagnostics. However, it's essential to be aware of the challenges, such as limited data availability, model transparency, and integration into clinical settings, which remain significant obstacles to the widespread adoption of AI in cancer genomics. This review highlights current trends, identifies key challenges, and explores future directions in the application of AI to cancer genomic research. The goal is to emphasize the role of AI in accelerating the development of precision medicine, a promising future for cancer care that will significantly improve patient outcomes.

Keywords: Artificial intelligence, Cancer diagnosis, Deep learning, Genome science, Healthcare, Machine learning, Precision medicine.

---

[*] **Corresponding author P.S. Akshatha:** Department of AIML, New Horizon College of Engineering, Bengaluru, India; E-mail: akshatha.ps@gmail.com

Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan (Eds.)

INTRODUCTION

The Role of Genomics in Cancer Research

Genomics is one of the most important parts of cancer research, which has dramatically changed our view on this disease and its treatment over the last few years. By using high-throughput sequencing technology, researchers can also identify specific genetic mutations and changes in tumors, which can then be managed through personalised treatments that address those unique characteristics. This personalised precision medicine approach improves therapy efficacy and reduces off-target adverse effects by targeting pharmacological agents to the right patient groups. Bioinformatics has also already enabled the detection of high-risk cancer patients using only a blood sample, work that could lead to early-stage diagnosis long before tumors form — hence the current hot topics like liquid biopsies detecting tumour-related mutations. In addition, tumor heterogeneity [1] and altered genetic landscapes are also well captured by genomic analysis, which aids in making a better estimate of treating resistant tumors. This genetic research has identified a group of markers that are implicated in Crohn's, improving patient detection.

Challenges in Traditional Cancer Detection Techniques

Conventional methods used to detect cancer, which enjoy a long history in oncology, come with a number of limitations that can compromise the process of diagnosis. Diagnosis: Many current diagnostic methods, such as imaging studies (X-rays, CT scans, *etc*) and biopsies, can be invasive and frequently result from observations due to symptoms or the palpable presence of tumors, leading often to late-stage diagnoses when treatment options are limited. These tactics can be invasive, painful, and hazardous to patients. Additionally, the sensitivity and specificity of some imaging modalities may be suboptimal, resulting in false positives or negatives, which can complicate management decisions. Tumors are also heterogeneous, so some types of cancers are harder to detect with this approach than others, which a standardized protocol would help address. Moreover, the typically used traditional approaches fall short in giving a comprehensive account of tumor molecular characteristics, an essential trigger for disease progression, and a guide to personalised treatment. Therefore, there is a requirement for alternative diagnosis techniques that can overcome existing limitations and help in early detection as well as the efficacy of the treatment.

Emergence of AI in Genomic Analysis for Cancer

AI in genomic analysis: The adoption of AI in the context of genomic analysis has markedly improved cancer research by facilitating processing and interpretation for large data sets. Machine Learning (ML) and DL models, AI algorithms to be precise, enable analysing enormous datasets at speeds never before matched by human researchers, and could unveil even the smallest of patterns and correlations. These technologies allow for the detection of genetic mutations, structural variants, and epigenetic changes underlying different cancers to guide improved biomarkers as well as targeted therapy. These AI-driven tools can combine genomic data with clinical history, enhancing risk stratification and patient prognosis. Automate data analysis: With automation in surfacing useful insights faster, AI can help quicken the pace of genomic research by saving time and money on analysing heaps of data for scientists to focus on hypothesis generation & experimental planning. By combining AI with genomic technologies, we can not only speed up discoveries in cancer biology but also enable the progression towards a personalised medicine paradigm, which should hopefully translate into more efficient and targeted treatment solutions for patients. Given their continued evolution, we are excited about the potential for these tools to transform our understanding of cancer and drive clinical improvement.

AI for Personalised Treatment Planning and Ethical Considerations

AI is increasingly being used not only in the interpretation of genomic data but also in clinical decision-making, particularly for personalised treatment planning. AI models can predict individual patient responses to therapies, thereby minimizing the traditional trial-and-error approach in cancer treatment. This advancement enables the design of treatment strategies tailored to the unique genetic and physiological profiles of patients, enhancing therapeutic efficacy and reducing adverse effects.

Foundational preprocessing techniques such as Robust Feature Extraction from Omnidirectional Outdoor Images for Computer Vision Applications, Light Propagation through Biological Tissue: Comparison of Monte Carlo Simulation with Deterministic Models, Artificial Intelligence in Radiation Oncology, and Near-Infrared Optical Imaging of the Breast play a critical role in supporting high-quality data extraction, simulation, and imaging. These processes are crucial for generating accurate and meaningful input for AI models used in personalised oncology care.

However, the integration of AI into healthcare systems introduces important ethical and technical challenges. Ensuring the privacy and security of patient data, maintaining transparency and explainability in AI model decisions, and

CHAPTER 3

Machine Learning Approaches for Microbiome Analysis and Applications

Aditya Vardhan[1,*], **Amarjeet Singh Chauhan**[1], **Sanjay Saini**[1] and **Sagar Sharma**[2]

[1] *Department of Physics and Computer Science, Dayalbagh Educational Institute, Agra, UP, India*

[2] *George Brown College, Toronto, Canada*

Abstract: Microbiome analysis has become an important area of study in biomedical and environmental sciences, offering knowledge regarding microbial communities and their interactions. Machine Learning (ML) and Deep Learning (DL) have also played a major role in improving the study of microbiomes by supporting effective processing and classification of high-dimensional metagenomic data. Here, the application of DL methods, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Autoencoders, is explored in the context of disease prognosis, microbial community typing, environmental microbiome research, and personalised medicine. The challenges, such as limited labelled data, model explainability, and the absence of generalisability, persist despite all the progress. The integration of multi-omics data, Explainable AI (XAI), and Federated Learning becomes unavoidable in overcoming these bottlenecks. Future research needs to be focused on developing standardised datasets and scalable AI models for real-time monitoring of the microbiome. Overwhelming these encounters, ML and DL will continue to revolutionise microbiome analysis, foremost to innovation in accurate medicine and biotechnology, besides ecological sustainability.

Keywords: ANN, Autoencoder, Biotechnology, CNN, Deep learning, Machine learning, Microbiome.

INTRODUCTION

Microbiomes are multifaceted communities of bacteria, fungi, and viruses, besides archaea, living within environments such as the human host, soil, water, and atmosphere. Recently, **there has been extensive interest** in **studying** microbiomes due to their role in human disease, agriculture, biotechnology, and

---

[*] **Corresponding author Aditya Vardhan:** Department of Physics and Computer Science, Dayalbagh Educational Institute, Agra UP, India; E-mail: vardhana3098@gmail.com

Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan (Eds.)

environmental sustainability. High-throughput sequencing knowledge, like 16S rRNA sequencing and metagenomics, has, over the historical decades, shaped vast microbiome data. However, it is computationally expensive to process this data and needs sophisticated Machine Learning (ML) and Deep Learning (DL) approaches for successful clarification, as shown in Fig. (**1**) [1].

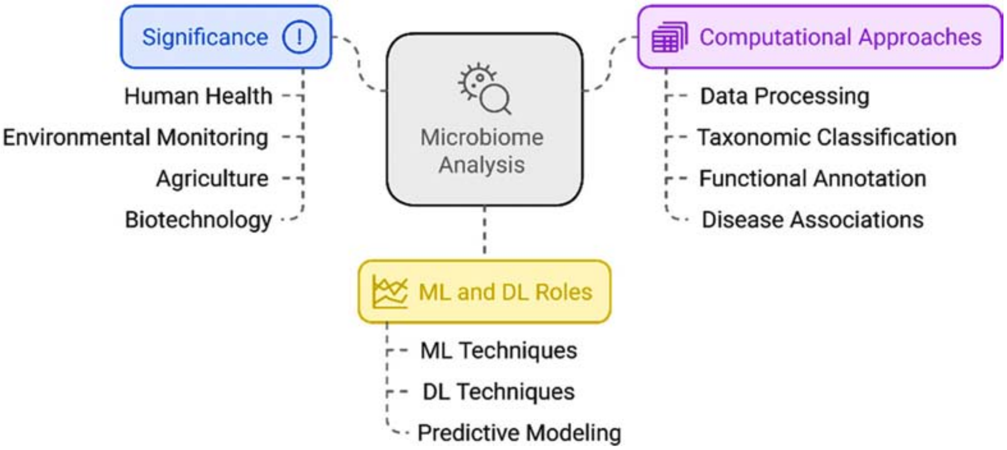

Fig. (1). Microbiome analysis: Significance and computational approaches.

Overview of Microbiome Analysis and its Significance

Microbiome analysis includes recognising, classifying, and understanding the functions of microbial groups [2]. This study is critical for:

- **Human Health:** Microbiomes influence digestion, immunity, and disease susceptibility. Modifications in gut microbiota consumption remained linked to settings such as diabetes, obesity, Inflammatory Bowel Disease (IBD), and cancer [3].
- **Environmental Monitoring:** Soil and water microbiomes play key roles in nutrient cycling, pollution degradation, and ecosystem health [4].
- **Agriculture:** Microbial interactions affect soil fertility, plant health, and crop yields.
- **Biotechnology and Industry:** Microbes are used in fermentation, biofuel production, and pharmaceutical development.

Despite its suggestion, microbiome analysis is motivating due to the massive diversity of microbial periods, the dynamic nature of microbial systems, and the high dimensionality of sequencing data [5].

Importance of Computational Approaches in Handling Microbiome Data

With the increasing availability of microbiome datasets, computational methods are essential for:

- **Data Processing:** Handling raw sequencing information, quality control, and standardisation.
- **Taxonomic Classification:** Identifying, in addition to categorising, microbial classes.
- **Functional Annotation:** Forecasting microbial purposes in relation to metabolic pathways.
- **Pattern Recognition and Disease Associations:** Identifying microbiome cross-supplementary diseases.

Traditional statistical approaches are challenged by the high dimensionality, as well as the intricacy of microbiome data. Machine Learning (ML) and Deep Learning (DL) have emerged as influential methods of processing and examining microbiome datasets more accurately and professionally [6, 7].

Role of Machine Learning (ML) and Deep Learning (DL) in Microbiome Research

Machine Learning systems, such as supervised and unsupervised learning, have also been extensively used for microbiome prediction, demonstrating clustering and classification. The ML algorithms, Random Forest, Support Vector Machines (SVM), and k-means clustering support the discovery of microbial patterns and their correlation with diseases. Deep Learning, an expansion of ML, has transfigured microbiome research by leveraging neural systems to uncover latent patterns in multifaceted microbiome data. Convolutional Neural Networks and Recurrent Neural Networks have proven to be extremely effective for metagenomic sequence examination with accurate class classification, functional prediction, and modelling of microbial communications. As microbiome research advances, merging ML and DL methods will play a vital role in discovering microbial functions in health, disease, and environmental balance. The subsequent sections provide insight into basic microbiome analysis methods, ML models, and their uses in metagenomics research [8].

FUNDAMENTALS OF MICROBIOME ANALYSIS

Microbiome analysis is the study of microbial populations dwelling in various habitats, including the human body, soil, water, and extreme environments. Microbial populations play key roles in health, agriculture, and environmental

CHAPTER 4

Machine Learning Model for Annotation and Assembly of Genomes

Aditya Vardhan[1,*], **Amarjeet Singh Chauhan**[1], **Sanjay Saini**[1] and **Sagar Sharma**[2]

[1] *Department of Physics and Computer Science, Dayalbagh Educational Institute, Agra, UP, India*

[2] *George Brown College, Toronto, Canada*

Abstract: The new advances in genome sequencing technologies have produced exponentially greater genomic data, which need scalable and accurate computational methods for genome annotation and assembly. Traditional annotation and assembly procedures are not scalable, accurate, and computationally efficient, which allows the utilisation of Machine Learning (ML) approaches. This chapter thoroughly reviews ML-based procedures for genomic data processing, annotation, and assembly. It describes some of the various ML approaches, from conventional algorithms to deep learning and combination models, and how they improve sequence alignment, feature extraction, and functional analysis of genetic components. The chapter also describes some of the challenges of integrating ML into genomics, including data heterogeneity, model interpretability, and scalability issues. In addition, a comparative assessment of ML-based and traditional genome annotation tools is illustrated, emphasising the superiority of ML models in terms of accuracy and efficiency.

Keywords: Artificial intelligence, Computational genomics, Deep learning, Genome annotation, Genome assembly, Machine learning.

INTRODUCTION

Genome assembly and annotation are fundamental genomics operations by which researchers can decipher the organisation and function of genetic sequences. Genome annotation refers to a procedure by which functional components within a genome, such as coding areas, regulatory units, and non-coding RNA, are delineated and labelled. It is a key process toward unravelling gene function, evolutionary background, and potential biomedical uses. Genome assembly, on the other hand, is the reconstruction of full genome sequences from short sequencing reads. Due to technological restrictions, existing sequencing techno-

---

[*] **Corresponding author Aditya Vardhan:** Department of Physics and Computer Science, Dayalbagh Educational Institute, Agra UP, India; E-mail: vardhana3098@gmail.com

Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan (Eds.)

logies generate millions of short DNA sequences, which must be correctly aligned and integrated into a full genome. Quality of genome annotation and assembly significantly affects downstream analyses in functional genomics, comparative genomics, and clinical research [1, 2].

Importance of Accurate Genomic Feature Identification

Correct identification of genomic features is required for many applications, including disease gene discovery, drug discovery, and evolutionary biology. Errors in genome annotation will result in inaccurate gene predictions, which will affect biomedical research and personalised medicine. Genome assembly errors will result in fragmented or misassembled sequences, lowering genomic study reliability. With increasing genomic data sets, scalable and automated methods must be offered to ensure the accuracy and effectiveness of genomic feature discovery [3].

Limitations of Traditional Computational and Manual Methods

- Legacy genome assembly and annotation are based on both rule-based computational algorithms and manual curation. Legacy methods are subject to various challenges:
- High Computational Cost: As genomic data increases, conventional algorithms are inefficient and costly.
- Error Propagation: Human annotation is susceptible to error, and rule-based approaches fail to deal with new sequences that differ from previously seen patterns.
- Limited Scalability: Traditional approaches do not generalise across species because they do not account for differences in genome structure and complexity.
- Gene Prediction Ambiguity: Genes overlapping, alternative splicing, and non-coding factors make genome annotation more complicated, and thus, conventional methods become less dependable.

Role of Machine Learning in Genomics

Machine Learning (ML) is becoming an important means of enhancing genome annotation and assembly through large-scale genomic datasets (Fig. **1**) [3]. ML models can:

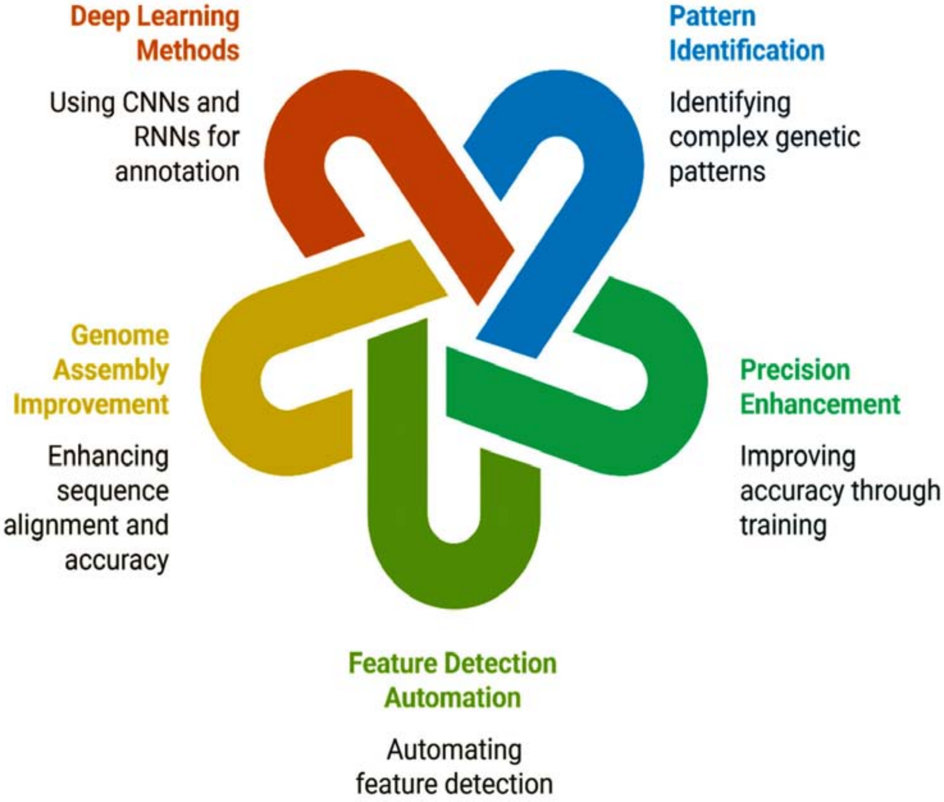

Fig. (1). Enhancing genomics with machine learning.

- Identify intricate patterns in genetic sequences that are challenging to model with conventional computational techniques.
- Enhance precision by being trained on annotated data and modifying themselves to fit new sequences.
- Facilitate feature detection automation, lessening the need for manual curation and increasing scalability.
- Improve genome assembly through the improvement of sequence alignment, error correction, and scaffold assembly.

Deep learning methods, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have proven to be highly successful at genome annotation and have been used to identify coding regions, regulatory elements, and functional motifs at high accuracy. Likewise, machine learning models improve genome assembly by predicting sequence overlap, correcting sequence errors, and enhancing the assembled genome's accuracy [1, 4].

CHAPTER 5

Extending Bayesian Classification to Predict Phase Transition in Biopolymer

Charu Kathuria[1], **Deepti Mehrotra**[2,*] and **Navnit Kumar Misra**[3]

[1] *Department of Computer Science and Engineering, PES University, Bengaluru, Karnataka, India*

[2] *Jaypee Institute of Information Technology, Noida, India*

[3] *Department of Physics, Brahmanand College, The Mall, Kanpur, India*

Abstract: Machine learning is an essential tool for predicting systems' behavior. Many machine learning algorithms have recently been used to predict the system's changing behavioral pattern. Phase transition is an exciting phenomenon in physics, observed in many molecules and systems, where a sharp structural change occurs with a minor change in an environment like pressure or temperature. It is a significant process to understand the structure of a molecule. The Bayesian algorithm has diverse applications, and in this chapter, its usage is incorporated with the approach of one-class classification to predict the phase transition in biopolymers. In this study, the algorithm proposed predicts the phase transition of L-alanine polypeptide. This algorithm corroborates the detection of transition parameters or changepoints in temperature-dependent significant frequencies with experimental results. The observed variation of Amide frequencies of the polypeptide is used to confirm the occurrence of a phase transition that causes structural change with the effect of temperature change.

Keywords: Amide frequencies, Change-point, Infrared spectroscopy, Machine learning, Naïve Bayes, One-class classification, Phase transition, Probability, Temperature variability.

INTRODUCTION

Phase transition is an essential concept of physics, and its understanding as well as exploration is a challenging task. Many physical systems exhibit a sudden change in symmetry when there is a change in the external environment, such as a change in some parameters like temperature or pressure. This change can be continuous or discontinuous across a transition parameter. This transition parameter, detecting how a change occurred in a system, is essential in many application areas. Recently, machine learning algorithms have been explored to

---

[*] **Corresponding author Deepti Mehrotra:** Jaypee Institute of Information Technology, Noida, India;
E-mail: mehdeepti@gmail.com

Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan (Eds.)

explain the concept of phase transition and have been applied in various case studies [1 - 4].

Machine learning, a subset of artificial intelligence, allows computers to act without being explicitly programmed. It turns information into knowledge and finds hidden patterns within data. This knowledge and hidden patterns from data predict complex decisions or events in the future. In today's technological era, Machine learning has progressed dramatically in applications of broader domains [5 - 9]. Broadly, the machine learning algorithms are categorized as supervised and unsupervised learning. In supervised learning, the data is labeled with prior knowledge, which can be categorical or discrete values, leading to further categorization in classification or prediction problems. In unsupervised learning, there is only input data with no prior output information, leading to clustering problems.Therefore, the data type plays a vital role in deciding which algorithms can be implemented for a given set of values.

The traditional classification problem works to categorise the unseen data into one of the given sets of labels. However, confusion occurs if this unseen data does not belong to any predefined class or label. Therefore, in such cases, the classification task is not to fit this new object in any predefined class; instead, it should decide whether it belongs to a particular class or not [10]. So this concept led to the advent of One Class Classification (OCC) models. Here, one of the target classes is considered positive, while all the other categories are treated as unfavorable or outliers. The OCC model primarily creates a boundary that encloses the available samples [11]. Many techniques can decompose multiclass problems suitable for one-class classification [12].

The data involved in time series analysis determines the system's behavior concerning sequence or time. A change-point is a concept mainly applied to time series data to determine breakpoints where the transition has occurred. The concept has covered several application areas, such as speech recognition [13], image analysis [14], climate change detection [15], and many more. For decades, models have been designed to implement this concept as follows: Julious (2001) determined a single change point [16], Muggeo (2003) illustrated a segmentation model for multiple change points [17], Yang *et al.* (2006) developed the CUSUM model for detecting heart rate [18], Reddy *et al.* (2010) used a hidden Markov model for transportation mode [19], and Han *et al.* (2012) implemented a Gaussian mixture model based on multimodal sensors in a smartphone [20]. Apart from these models, various other models and packages are available in the literature. The time series approach can be extended to understand the impact of the variability of environmental factors on the given system. Thus, mathematical modeling applied to interpret the change point analysis in a given time series pattern can also be used to describe a given system's transition by varying environmental factors like temperature, pressure, and pH.

This chapter designs an algorithm to predict the phase transition in biopolymers. The biopolymers show a change in structure with variations in temperature or pressure. The proposed algorithm uses the OCC machine learning model and a change-point concept to detect phase transition in a peptide (L-alanine), considered a case study. With the increase in temperature, the L-alanine shows a transition from the alpha-helical structure to the random coil. The order of transition reflects the sharpness of the phase transition. The selected case study validates the proposed model. The chapter further describes the literature review in the next section, continuing with the proposed algorithm in section three. The fourth section considers a case study that implements the algorithm and discusses the results. Finally, the conclusion is described in the fifth section.

LITERATURE REVIEW

The proposed algorithm extended the one-class classification approach to explain this phase transition phenomenon. In this section, we have studied the recent research work where machine learning algorithms are used along with the phase transition phenomena. In the following section, the one classification approach is explained to justify why this approach can be suitably used to develop the algorithm.

Phase Transition with Data Mining Concepts

Phase transition is an essential physics concept that has recently gained much importance in data mining and other fields. Giordana and Saitta, in 2000, used the phase transition concept in matching problems based on the Monte Carlo algorithm and showed that phase transition could occur in real-world problems [21]. Richard *et al.*, in 2004, presented a signal classification approach using the Gaussian Mixture model to capture the reconstructed phase space [22]. Broecker *et al.*, in 2016, demonstrated the use of Convolutional Neural Networks in quantum fermion systems to locate phase transitions [23]. Ohtsuki and Ohtsuki, in 2016, proposed a deep learning approach for two-dimensional quantum phase transitions [24], and Ohtsuki and Ohtsuki, in 2017, showed a phase diagram as well as used CNN for three-dimensional quantum phase transitions [25]. Tanaka and Tomiya (2017) designed a CNN that studied the correlation between the two-dimensional Ising model spin configuration and the temperature [26]. Schindler *et al.* (2017) demonstrated a transition in many-body quantum systems using simple neural networks [27]. Li *et al.* (2018) determined phase transition in ferroelectrics using a machine learning approach and demonstrated the presence of phase boundaries [1].

CHAPTER 6

Ethical Aspects of Analysing Genomic Data for Medical Analysis: A Comparative Study with a Focus on India

Ankit Vishnoi[2], **Luxmi Sapra**[3], **Varun Sapra**[1,*] and **Parul Madan**[2]

[1] *ASET, Amity University Punjab, Mohali, India*

[2] *Department of Computer Science and Engineering, Graphic Era Deemed to Be University, Dehradun, India*

[3] *Graphic Era Hill University, Dehradun, India*

Abstract: The analyses of genomic data hold great promise for personalised medicine by allowing predictive diagnosis and treatment. However, this also brings ethical issues on privacy, consent, ownership of data, and genetic discrimination into question. This work discusses the above issues in the background of the evolving legal framework in India and compares the same with international regulations such as HIPAA, GINA, and GDPR. The sensitive nature of genomic data requires robust protection for privacy. However, India does not have any specific guidelines regarding the privacy and ownership of such data, thereby making people very vulnerable to misuse. Issues of transparency in the commercial use of genetic data are raised by the 23andMe case. Informed consent, too, is particularly challenging, especially in rural areas of India, where many participants may not understand the consequences of divulging their genetic information. The ownership disputes that were encountered in the case of Henrietta Lacks and DeCODE Genetics emphasize the need for more refined sets of regulations. Further, even in the insurance sector in India, discrimination based on genetics is often witnessed, which again calls for immediate protection similar to GINA in the U.S. This study thus recommends the following changes in the existing legal framework: introducing genomic-specific legislation in India, working out models of dynamic consent, transparency in data use, and non-genetic discrimination. Such a filling of the gaps shall allow innovation to take place in genomic research, but will guarantee the rights and privacy of the individual.

Keywords: DNA, Genetic discrimination, Genome, GINA, HIPAA.

---

[*] **Corresponding author Varun Sapra:** ASET, Amity University Punjab, Mohali, India; E-mail: varun.sapra@gmail.com

Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan (Eds.)

INTRODUCTION

The analysis of genomic data ushers medicine into a revolutionary phase wherein treatment protocols can be planned with unprecedented precision. In other words, genomic data provides immense detail regarding the DNA and heredity.

The constitution of a person aids health care providers in predicting not only which diseases the patient is most likely to develop but also how he or she could respond to various treatments. More appropriately termed personalised or precision medicine, the concept has the potential to revolutionise the mode of delivery in healthcare. Targeting a patient's unique genetic profile could permit physicians to offer more effective and personalised interventions that reduce trial-and-error approaches when it comes to medication and can improve patient outcomes. These scientific advances come hand-in-hand with some very complex and multi-layered ethical issues. The nature of genomic data in itself, reaching into very private and sensitive information about one's health, ancestry, and even possible future medical conditions, raises serious privacy concerns. In an environment where data protection mechanisms are weak or underdeveloped, individuals become exposed to possible breaches of privacy and the potential misuse of their personal genetic information. Ethical challenges of genomic data do not stop at privacy issues. At the forefront are issues related to informed consent, especially in view of the complexity of genomic information. The patients and research participants may not understand what they are signing up for when giving consent for the analysis or storage of their genomic data, especially where such data may be shared with third parties for research or commercial purposes. In addition, data ambiguities raise the question of ownership in relation to who is the controller of the genetic information: whether it is from the individual on whom the collection of data was made, the researchers analysing, or companies offering services in genomic testing. Ambiguity is, however, very critical where data might be commercialised, raising questions as to whether an individual should be remunerated when his or her genetic data is used to generate profit. Perhaps the most sensitive consequence of genomic data analysis is the possibility of genetic discrimination. When one's genetic tendencies for specific diseases or conditions that might not have manifested are known, there is a chance that the individual might be discriminated against based on such information. For example, insurance companies can reject insurance or charge higher amounts based on genetic markers without considering the real illness of the person. Similarly, employers can use such information to discriminate against any employee or a potential employee who, according to them, may become sick someday. This chapter points to ethical issues related to genomic data with a specific relation to India's data protection regime.

The rapidly expanding biomedical research and healthcare sector in India throws up unique challenges related to the regulation of the collection, storage, and use of genomic information. While the genomic testing services market is expanding and the number of research studies employing data from genes is also on the rise, it is relevant to probe whether India's current legal and regulatory environment is positioned to meet the complex moral dilemmas because of the use of such data. Furthermore, this chapter compares the approach of India to global standards, especially in areas like the United States and Europe, which have, of late, given much importance to data protection laws over the last few decades. For example, HIPAA in the United States has partial broad protection for medical information, while GINA bars the use of genetic information in certain health insurance underwriting and employment decisions. In Europe, the General Data Protection Regulation sets one of the most comprehensive data protection regimes globally. It regards genetic data as a special class of personal data, and consent for collection and use is based on explicit consent. Comparing India's evolving data protection laws, such as the Digital Personal Data Protection Act (DPDPA) 2023, against established frameworks like HIPAA, GINA, and GDPR, the chapter examines how well India's regulations address ethical issues thrown up by the analysis of genomic data. The specific research questions to be explored in this chapter are whether the regulatory environment in India is prepared to match the rapid growth of the field in genomic research and personalised medicine.

PRIVACY AND CONFIDENTIALITY OF GENOMIC DATA

It is expected that privacy and confidentiality become key issues in handling genomic data, for such data is, by its nature, intimate and sensitive. No other medical information speaks more of existing health status or future health risk and familial lineage than does genomic data. The unique features of genomic data as sensitive information are discussed here, together with a summary of relevant regulatory initiatives globally and in India, aimed at preventing its misuse [1].

Genomic Data as Sensitive Information

Genomic information, by its nature, encompasses one's total biological being. It should be able to predict most of one's personal information, including susceptibility to certain illnesses, ancestry, and other biological relationships. This becomes valuable due to its predictive power. For example, with the help of the analysis of some genetic markers, it is possible to estimate the risk for a person to develop cancer, Alzheimer's disease, or heart disease. As much as this type of information has become so significant for tailored medicine, it also poses more and more serious challenges regarding privacy.

CHAPTER 7

Integrating Pharmacogenomics and Machine Learning in Personalised Treatment Strategies for Parkinson's Disease

Prachi Mall[1] and **Deepika Raina**[1,*]

[1] *Graphic Era Hill University, Dehradun, Uttarakhand, India*

Abstract: Parkinson's Disease (PD) ranks second among the several neurodegenerative diseases. It is characterised by gradual and progressive motor and non-motor symptoms, with highly variable patient responses to pharmacological treatments. Conventional therapeutic procedures frequently result in less-than-ideal outcomes, calling for a more customised approach. The highly tailored treatments provide the right set of circumstances to customise the treatment plans based on the individual's genetic, lifestyle, and environmental factors. Pharmacogenomics is the study that focuses on how individual gene variations influence drug response. It holds great potential to optimise the therapeutic outcome of Parkinson's Disease (PD). The amalgam of the genetic information with clinical practice can do wonders. However, it also remains a challenge due to its variability and complexity.Moreover, highly sophisticated instruments that can handle and process the enormous volumes of genetic and clinical data are needed to utilise pharmacogenomics fully.

As a branch of artificial intelligence, Machine Learning (ML) offers the potential to comprehend large datasets, identify trends, and offer a viable way through the complexity. ML models can be trained to predict individual responsiveness to a particular medication, optimize dosing regimes, and minimize adverse effects by augmenting the pharmacogenomic data. To provide individualised treatment plans for Parkinson's disease, the study investigates the combination of pharmacogenomics and machine learning. The chapter focuses on recent developments in the area, emphasizes the possible advantages of these technologies, and discusses implementation difficulties such as model interpretability, clinical adoption, and data integration. The study also suggests a paradigm for incorporating ML-driven pharmacogenomics into standard clinical practice to improve therapeutic effectiveness and raise patient satisfaction.

Keywords: Machine learning, Neuro-degenerative disease, Parkinson's disease, Personalized treatment, Pharmacogenomics.

---

[*] **Corresponding author Deepika Raina:** Graphic Era Hill University, Dehradun, Uttarakhand, India; Tel: +91-8791591364; E-mail: deepika709@gmail.com

Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan (Eds.)

INTRODUCTION

Overview of Parkinson's Disease

Parkinson's disease is the second most prevalent, gradually progressive neurodegenerative disorder. It affects the dopamine-producing neurons in the substantia nigra region of the brain. PD symptoms can be widely categorised as motor and non-motor, which can vary from patient to patient [1]. Researchers are continuously working to investigate the biomarkers for the early detection of the disease and individualised treatment to slow the disease progression. Biomarkers are quantifiable indicators that provide information about biological processes, physiological states, or disease states in living organisms [2]. At present, there is no identified treatment for the disease; however, several pharmacological therapies and surgeries can largely provide symptomatic relief only if detected in the early phase. PD is primarily idiopathic, affecting dopaminergic neurons in the brain's substantia nigra region [3]. It is primarily caused by host sensitivity and environmental factors. Physiologically, Parkinson's disease-related symptoms lead to the reduction of several neurotransmitters, particularly dopamine. In Parkinson's, the dopaminergic connection is lost with the neostriatum putamen. Dopamine is a neurotransmitter that transmits chemical signals across the synapse. It is secreted by membrane storage vesicles in the presynaptic membrane and attached to the postsynaptic membrane, activating dopamine receptors [4]. Dopamine is absorbed back into the presynaptic cell, and extra dopamine is released into the synapse. When transferred, MAO and COMT enzymes break down dopamine, rendering it inactive. A therapeutic approach involves MAO inhibitors that disrupt MAO enzyme function, allowing more dopamine to remain in the synapse and reducing binding to the post-synaptic membrane [5].

The Role of Personalized Medicine in Neurodegenerative Disorders

Personalized medicine is revolutionizing the way neurodegenerative disorders are treated and managed by adjusting medical care to the unique characteristics of each patient, including their environment, lifestyle, and genetics [6]. Due to the complexity and variety of disorders like ALS, Huntington's disease, Parkinson's disease, and Alzheimer's disease, traditional medicines frequently prove ineffective in treating these conditions. This technique, however, offers promising pathways for treating these conditions [7].

Some of the key aspects of personalised medicine in neurodegenerative disorders are summarised as follows and depicted in Fig. (**1**).

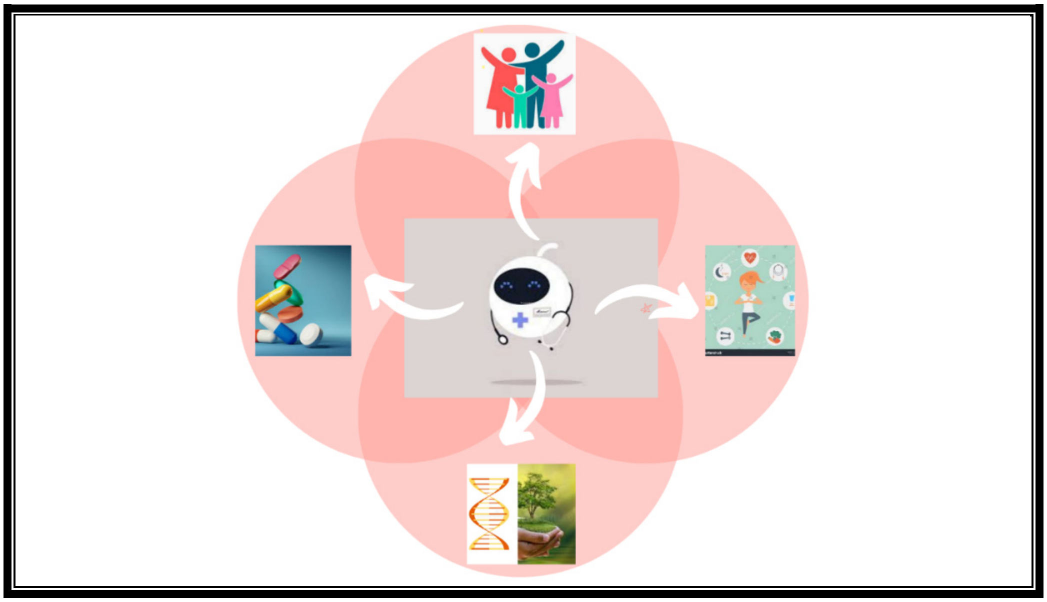

Fig. (1). Factors affecting the role of personalized medicine in PD.

1. **Genomic Makeup:** There are genetic components to many neurodegenerative disorders. For instance, PD is associated with mutations in the LRRK2, PARK7, PINK1, PRKN, or SNCA gene [8]. Genomic testing is used to determine risk factors and genetic variants. It also enables the creation of specialized treatments to deal with particular genetic mutations and provides families with information about hereditary risks through genetic counselling [9]. **Development of Biomarker:** Biomarkers are utilised to customise medicine to detect diseases earlier, even before clinical symptoms appear (*e.g.*, amyloid and tau biomarkers in Alzheimer’s and Lewy bodies in Parkinson's), track the course of the illness and the response to treatment and segregate patients into groups so that the best treatments can be prescribed to them [11, 12].
2. **Targeted Therapy:** Conventional therapies work on the principle of “one-siz--fits-all” and frequently fall short in addressing the distinct disease mechanisms that each patient has [12]. Personalised medicine aids in the creation of focused treatments by adjusting treatment plans based on each patient's unique genetic profile. Employing immunotherapies that are formulated depending on the patient’s immune response [13].
3. **Pharmacogenomics:** Pharmacogenomics is the study of how genes affect an individual's response to drugs [14]. Pharmacogenomic data are used in personalized medicine in the context of neurodegenerative diseases to adjust medication dosages to best suit the patient's metabolic profile. It helps

CHAPTER 8

Analysing Microbiome Metagenomics to Detect Anomaly using Variational Autoencoders (VAE) and Generative Adversarial Network (GAN) for Predicting Crohn's Disease

Samrath Prakash[1,*], **Samir Jaiswal**[1], **Nikita Kukreti**[1], **Akhilesh Singh**[1] and **Roohi Sille**[1]

[1] *School of Computer Science, UPES, Dehradun, India*

Abstract: The microbiome is a group of microorganisms that live inside the human gut. Imbalances in this intestinal microbiome can lead to the development of Inflammatory Bowel Disease (IBD), because of which it is essential to detect the anomalies in the microbiome metagenomics, which can help in better diagnosis of diseases like Crohn's disease. Various anomaly and feature selection techniques are available, such as statistical methods (Z-score & IQR, Principal Component Analysis (PCA)) and machine learning techniques (Random Forest, Support Vector Machine (SVM), and K-means clustering). However, due to the high dimensionality of the genomics data, it becomes difficult for these techniques to capture the non-linear relationships of the data, as there is a lot of probabilistic uncertainty associated with it. In this research, we have tried to suggest a study for feature extraction and anomaly detection methods using Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN) to tackle the problem of high dimensionality, proposing a better way to capture non-linear relationships and handle probabilistic uncertainty. The expected outcome for using the VAE and GAN approach is that it can help in better diagnosing and predicting the diseases that are caused by genetic factors, such as Crohn's disease.

Keywords: Crohn's disease, Generative Adversarial Network (GAN), Metagenomics, Microbiome, Variational Autoencoders (VAE).

INTRODUCTION

The human body is a complex symbiotic organism that is composed of many symbiotic microorganisms along with its own cells [1]. Intestinal microecology,

---

[*] **Corresponding author Samrath Prakash:** School of Computer Science, UPES, Dehradun, India; E-mail: Samrath.14606@stu.upes.ac.in

Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan (Eds.)

being the most important in the human body, is responsible for digestion, immunity regulation, absorption of nutrients, and most importantly, creating a balance between harmful and beneficial bacteria. Any imbalance in the intestinal microecology can lead to systemic diseases. The microbiome is the most crucial microorganisms that live inside the human gut. These microbiomes are dynamic in nature and are affected by different microorganisms and the way they interact with each other.

Crohn's disease is a chronic inflammatory bowel condition that causes inflammation and ulcers and affects the digestive tract, leading to symptoms such as diarrhea, stomach aches, blood in stool, weight loss, mouth ulcers, fever, and skin problems. Although the exact causes of Crohn's disease are still not clear, it is thought to be caused by a combination of factors. Since the human gut microbiome is an ecosystem of trillions of microorganisms, it plays an important role in digestion, immune regulation, and overall health. However, there can be disruptions in this microbiome, which are termed dysbiosis, and have been linked with various gastrointestinal disorders [2]. Crohn's disease is characterised by an altered gut microbiome composition, which causes a reduction in the beneficial bacteria and an increase in pathogenic species [3]. No particular cure has been found yet for Crohn's disease, but early detection and treatments can help in decreasing inflammation in the intestines, which could help in preventing future complications.

Traditional methods of analyzing microbiomes, such as 16S rRNA sequencing and shotgun metagenomics, provide a huge amount of microbial composition and functional data. However, detecting disease-related anomalies in such a high-dimensional dataset is a challenge [4] where statistical methods often fail to capture the complexity of microbial interactions, which vary across different individuals. Machine learning and deep learning strategies, however, offer promising solutions for microbiome analysis. Generative Adversarial Networks (GANs) are becoming very popular for generating information from the given data. It uses the provided trained data and generates new data that is similar to the training data. GAN contains two neural networks: a generator that creates fake data that is similar to the original data and a discriminator that compares data and differentiates between the real and the fake data. It is an emerging technique that bridges the gap between real and fake data. Through refinement and competition, GANs can be used to generate artificial data that is of high quality, having applications in various fields [5].

As depicted in Fig. (**1**), in 2019, nearly 4.9 million cases worldwide were found to be related to Inflammatory Bowel Disease (IBD). Most of the cases were found to be in China and the USA. From 1990 to 2019, the worldwide age-standardised rates for prevalence, mortality, and disability-adjusted life-years (DALYs) declined (EAPCs=-0.66, -0.69, and -1.04, respectively). Nonetheless, the age-adjusted prevalence rate rose in 13 of the 21 Global Burden of Disease (GBD) regions. Out of 204, nearly 147 countries have seen an increase in age-standardised rates for prevalence. Deaths related to IBD cases have risen more among females than among males from 1990 to 2019. IBD has become a major health burden due to the increasing number of cases, the related deaths, and DALY [6].

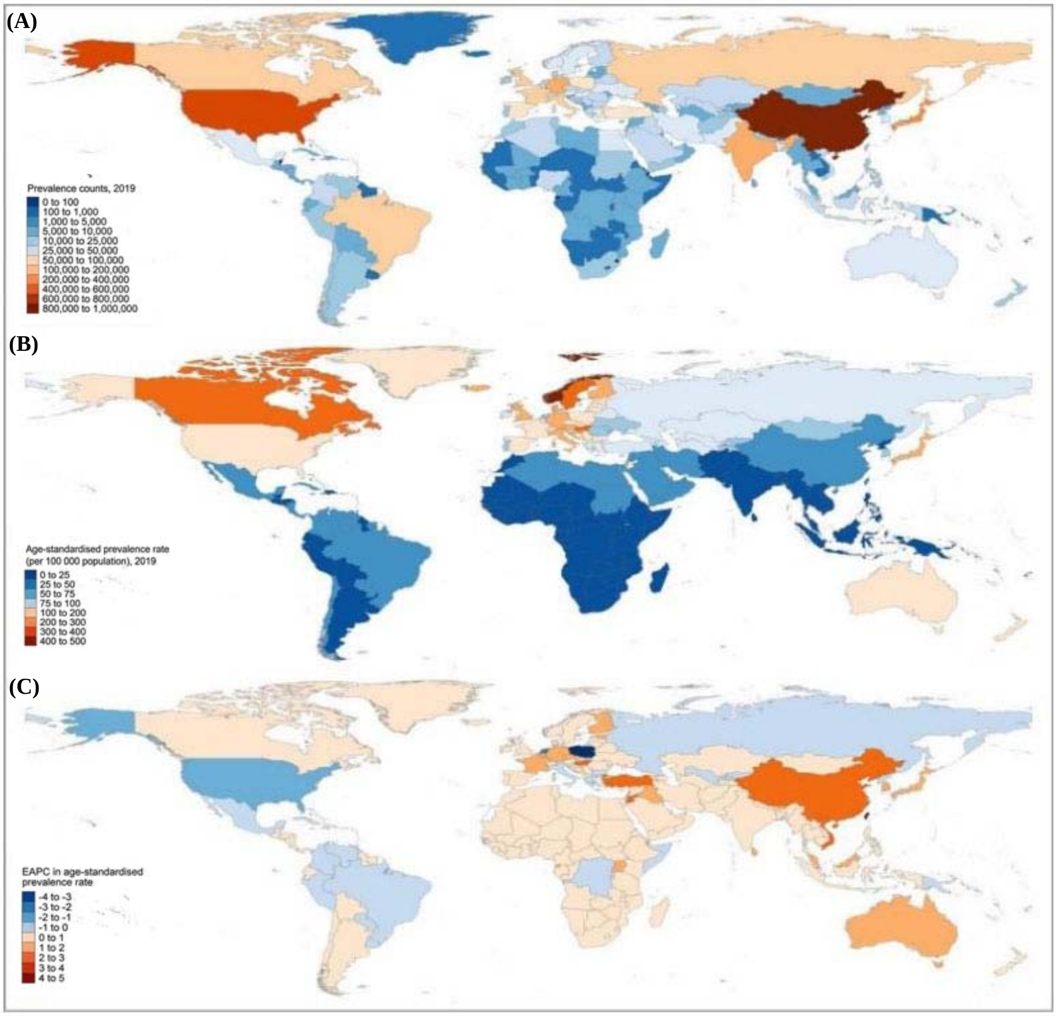

Fig. (1). (**A**) The total number of IBD prevalent cases in 2019. (**B**) The age-standardised occurrence rate (per 100,000 population) of IBD in 2019. (**C**) The EAPC of the age-standardised occurrence rate for IBD between 1990 and 2019. EAPC, estimated annual percentage change; IBD, inflammatory bowel disease [6].

CHAPTER 9

Leveraging ML Models for the Future of Genomes

Sumit Bhardwaj[1,*]

[1] *(Global Analytics), Concentrix Daksh Services India Pvt Ltd, Chandigarh, India*

Abstract: Artificial Intelligence (AI) that includes Machine Learning (ML) and Deep Learning (DL) has revolutionised the various domains of science and genetics. The AI-driven approaches have enabled the rapid analysis of vast genomic datasets, improved the diagnosis of various diseases and drug discoveries, leading to personalised medicines. This chapter explains the role of AI and ML in genetics, exploring their applications along with their benefits, challenges, and future prospects.

Keywords: Artificial intelligence, *De novo* mutations, Deep learning, Genetics, Insertions and deletions, Machine learning, Single nucleotide polymorphisms.

INTRODUCTION

Artificial Intelligence is an emerging field of computer science that enables machines to perform tasks that typically require human intelligence. The main branches of AI are Machine Learning (ML) and Deep Learning (DL). Machine Learning consists of AI models that allow machines to learn from available data and improve their predictions over time. In contrast, Deep Learning is a subset of Machine Learning that uses neural networks for the analysis of complex patterns in large datasets. For example, various diseases based on genetic mutations can be predicted with the help of AI that can analyse the DNA sequences instantly [1 - 6].

The field of AI-incorporated genetics has revealed new possibilities to understand complex genetic interactions and, hence, predict disease risks with further tailored treatments. The advancement in the areas of ML, DL, and neural networks has enabled computational tools to process extensive genomic data that can identify patterns beyond human capability. This chapter focuses on highlighting the role of AI in transforming genetics, finding its extensive applications in potential future developments.

---

[*] **Corresponding author Sumit Bhardwaj:** (Global Analytics), Concentrix Daksh Services India Pvt Ltd, Chandigarh, India; E-mail: sumitb460@gmail.com

Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan (Eds.)

The concept of machine learning is widely used in genomics for various purposes, such as the detection of various mutations with the help of Single Nucleotide Polymorphisms (SNPs) and indel classification, predicting the risks of various diseases called the polygenic risk scores, and functional annotations that predict if a mutation is harmful or not. The most common ML techniques that are used in genetics are as follows:

1. Random Forest & Gradient Boosting: This technique is used for the classification of mutations as benign or pathogenic.
2. Deep Neural Networks (DNNs): This technique uses tools like DeepVariant to improve the variant calling accuracy.
3. Graph Neural Networks (GNNs): Such networks are used to map the gene interactions that prioritize the disease-causing mutations.

From the literature survey, it has been found that the human genome contains a billion base pairs, which can make manual analysis an infeasible process. To overcome this problem, ML algorithms, either supervised or unsupervised learning models, are found to be instrumental in processing the genetic data. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) analyse the DNA sequences that help to detect the mutations in the structural variations. Single-cell RNA sequencing done with the help of ML techniques categorises cell types and identifies the various biomarkers in complex genetic expressions [7 - 9].

AI has set a landmark in the diagnosis and prediction of various incurable diseases. AI models identify the correlations between genetic variants and diseases. AI-driven models have made it possible to early detect and diagnose the genetic disorders of cancer genomics. AI algorithms analyse the mutations in oncogenes and tumour suppressor genes to predict the susceptibility to cancer. The available ML tools assist in the identification of genetic risk factors of neurodegenerative diseases such as Alzheimer's and Parkinson's. AI aids in the diagnosis of rare genetic disorders and conditions like cystic fibrosis and Duchenne muscular dystrophy with the help of pattern recognition. AI models predict the potential drug targets of pharmacogenomics based on genetic markers. ML ensures precise medicine by predicting the individual responses to the drugs based on the technique of genetic makeup. Moreover, AI upgrades the accuracy of CRISPR-Cas9 gene-editing techniques, reducing the off-target effects [10 - 12].

However, despite its potential, AI faces a lot of challenges in the field of genetics. The data privacy, security, and handling of sensitive genetic data mark up the ethical concerns. The algorithmic bias in AI models can often lead to inaccurate genetic predictions, thus affecting the healthcare outcomes. Additionally, complex AI models require improved interpretability for clinical applications.

The advancements in the field of AI will continue to drive the genetic research and hence improve the diagnostic accuracy with treatment efficacy. The future developments will continue to focus on enhancing the model transparency by addressing the ethical issues and integrating AI with the next-generation sequencing technologies. The role of AI in genetics holds immense promise for revolutionising medicine and human health.

ROLE OF MACHINE LEARNING AND AI IN THE VARIOUS FIELDS OF GENETICS

Single Nucleotide Polymorphisms (SNPs)

SNPs, also pronounced as "snips," are found to be one of the most common types of genetic variations in humans, in which a single nucleotide, such as A, T, C, or G, can be altered at a particular position inside the genome. The types of SNPs include neutral traits and diseases, or impact traits and diseases. One of the advantages of SNPs is their Genome-Wide Association Studies, also known as GWAS, which help to link genes to various diseases. SNPs are used in a wide variety of applications, including ancestry tracing and the development of personalised medicines. SNPs are used in a wide variety of ancestry tracing and in the development of personalised medicines. For example, if a single SNP changes A to G in a gene, then it will potentially affect the functions of a protein.

Insertions and Deletions (Indels)

Indels are referred to as the small insertions or deletions of nucleotides in DNA sequences, in which the base pairs usually range from 1 to 50. Such a type of mutation can lead to disruptions in the gene functions that occur in the coding regions. Indels can alter the way a gene is read, known as frameshift mutations. Indels find applications in the fields of forensic genetics as well as evolutionary studies. For example, if two nucleotides in a gene are deleted, it could result in a mutated protein that becomes nonfunctional, ultimately leading to a genetic disorder.

De Novo Mutations

De novo mutations are the new kind of genetic changes that have not been inherited by an individual from either parent but appear spontaneously in their genes only. Such mutations occur mainly in the sperm, eggs, or during early embryonic development. Sometimes such mutations become the major cause of

CHAPTER 10

Conclusion: The Transformative Role of Machine Learning in Genomic Science and Healthcare

Ankit Vishnoi[1,*], **Parul Madan**[1], **S. Balamurugan**[2], **Roohi Sille**[3] and **Tanupriya Choudhury**[3]

[1] *Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India*

[2] *Research and Development, Intelligent Research Consultancy Services (iRCS), Coimbatore, Tamil Nadu, India*

[3] *School of Computer Sciences, University of Petroleum and Energy Studies (UPES), Dehradun, Uttarakhand, India*

Abstract: This final chapter reiterates the revolutionary contribution of machine learning to genomic medicine and healthcare. It points to the growing role played by AI-based methodologies, especially deep learning architectures such as CNNs, in speeding genomic data analysis, increasing diagnostic accuracy, and making personalised medicine possible. Other methods like reinforcement learning and GANs also continue to revolutionise drug discovery and modeling complex biological interactions. Machine learning applies to a wide range of applications across realms such as cancer diagnosis and the study of rare genetic diseases, with companies such as DeepMind and 23andMe at the forefront of innovation. With technologies such as next-gen sequencing and CRISPR converging with AI, the possibilities for accurate, affordable, and personalised healthcare expand exponentially. The author foresees the synergy between genomics and AI transforming medicine into a more predictive and preventive approach, opening the doors to a new era of healthcare solutions across the lifespan.

Keywords: Annotation, Gene expression analytics, Genome assembly, Genomic ML models, Next-gen sequencing, Non-coding RNA.

INTRODUCTION: THE CHANGING LANDSCAPE OF GENOMICS AND MACHINE LEARNING

If we look at how genomic science and machine learning have evolved over the past decade, we can see that we are at the edge of a scientific revolution. The fusion of these two fields has completely changed how we understand the human

---

[*] **Corresponding author Ankit Vishnoi:** Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India; E-mail: Vishnoi.ankit@gmail.com

Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan (Eds.)

genome, diagnose diseases, develop treatments, and create personalized healthcare solutions. What was once impossible, such as predicting genetic disorders before they manifest or designing drugs specifically for an individual's genetic makeup, is now within reach, thanks to advancements in artificial intelligence and machine learning.

This book titled, *Elevating Next Generation Genomic Science and Technology Using Machine Learning in the Healthcare Industry*, extensively explores the relationship between machine learning and genomic science. Through twelve well-structured chapters, we have seen how machine learning is applied in genome sequencing, gene expression analysis, personalised medicine, epigenomics, non-coding RNA research, microbiome studies, and even ethical considerations surrounding genomic data analysis.

Now, as we conclude, let's take a deeper look at the most critical takeaways from this book, the challenges that still exist, and what the future holds for this field.

THE MAJOR INSIGHTS FROM THIS BOOK

The book was structured to take readers from fundamental concepts to real-world applications, ethical concerns, and prospects. Here's a breakdown of the main insights we have gained.

Understanding the Role of Genomics in Healthcare

The book starts with an introduction to genomic science and its significance in medicine. As it was set out, sequencing the human genome was a historic achievement, but interpreting the data is always the true challenge. The magnitude of genetic data is unfathomable, and processing it by hand is infeasible. This is where Machine Learning (ML) enters the picture, analysing complex patterns in genomic sequence to predict genetic diseases and even recommend targeted therapies based on an individual's unique genetic profile.

The book introduced us to the basics of machine learning in genomics. The data used in training these ML models is a large genomic data set, from which ML models learn the patterns that humans are unable to see. Supervised learning to estimate the likelihood of a disease, unsupervised learning to uncover new gene functions, and deep learning to understand complicated systemic interactions **are examples of how** machine learning has become a central factor in genomics.

The Application of Machine Learning in Genomic Analysis

Once we understood the basics, the book progressed to real-world illustrations. One of those areas is genome assembly and annotation. However, the labour involved in annotating genes and their functions is boring, prone to error, and computationally expensive. Machine learning models are now automating this process, and they are doing it much faster and with greater accuracy.

In this article, we looked at gene expression analysis. Determining which genes are activated or silenced under different conditions is of tremendous importance in medicine, particularly in cancer and neurodegenerative diseases. The large amount of gene expression data is used with machine learning to understand the most important regulatory elements potentially involved in health and disease.

That led to an engaging discussion on personalized medicine and pharmacogenomics. This is where some of the most advanced work in ML is being done. It envisions a world in which medications are tailored to your genetic profile, where side effects are minimized, and efficacy is maximized. Machine Learning (ML) is bringing this vision a step closer by **enabling researchers to** map out how individuals will react to various drugs.

I found a discussion on epigenomics and gene regulation interesting.While classical genetics is grounded in relatively immutable stretches of DNA, epigenomics examines how our environment influences which genes are expressed. Machine Learning (ML) is helping scientists to unravel these complex relationships, as well as to figure out how lifestyle choices such as diet and exercise mold activity in our genes.

Machine Learning was also shown to integrate a hidden layer function in the non-coding RNA analysis data. Scientists had long thought that the non-coding parts of DNA were useless. Nonetheless, these regions are relevant for gene regulation, cancer progression, and brain function according to ML.

One of the discussions that opened my eyes was the application of ML in microbiome. Our bodies are host to trillions of bacteria, viruses, and fungi that help us digest food, power our immune systems, and optimise our mental health. The microbiome has proven difficult to study because of its complexity, but Machine Learning (ML) has assisted researchers in detecting microbial patterns associated with different diseases, potentially paving the way for a new era of breakthroughs in the field of probiotics for gut health, digestive health, and disease prevention.

Machine learning is used in genomic analysis in various contexts, including genome assembly, whole-genome association, and analysis of microbiome data. Datasets and computational power have been driving the progress behind these fields. The details of each dataset, including their sources, primary applications, and commonly used machine learning techniques, for various genomic

SUBJECT INDEX

Roohi Sille, Tanupriya Choudhury, Sonal Talreja & S. Balamurugan (Eds.)

Roohi Sille

Dr. Roohi Sille is an academic and researcher in Computer Science with over 12 years of teaching and research experience. She holds a PhD in Computer Science, an M.Tech in Artificial Intelligence and Neural Networks, and a B.Tech in Biomedical Engineering. Her interdisciplinary background enables her to connect computing technologies with practical applications, particularly in artificial intelligence and medical data analysis. Her research focuses on artificial intelligence, deep learning, and medical image analysis, aiming to develop intelligent computational solutions for healthcare and data-driven decision-making. She has contributed to academic research, scholarly publications, and student mentoring in emerging areas of computer science. Dr. Sille is actively involved in teaching and curriculum development, promoting critical thinking, innovation, and problem-solving skills among students. Her teaching approach emphasizes experiential and research-oriented learning, along with the integration of modern technologies. She also participates in international academic collaborations, supporting global research partnerships, student mobility, and knowledge exchange. Committed to academic leadership and lifelong learning, she continues to strengthen research ecosystems and contribute to advancements in artificial intelligence and computing technologies.

Sonal Talreja

Dr. Sonal Talreja completed her Ph.D. in Computer Science from Mohanlal Sukhadia University in 2017. She is currently an Assistant Professor at the School of Computer Science, UPES, Dehradun. Her research primarily focuses on computational science, particularly the study of electronic and optical properties of advanced computer hardware materials using first-principles calculations and modern computational techniques. In recent years, she has expanded her work to include Artificial Intelligence and Machine Learning (AI/ML) methods for analyzing and predicting material properties. By integrating AI/ML with traditional computational approaches, she has developed an interdisciplinary research framework that enhances materials discovery and analysis. Dr. Talreja has published numerous research papers in SCI-indexed journals with high impact factors, contributing significantly to scientific literature. She has strong experience in interdisciplinary collaborations and research teamwork. A gold medalist in both her undergraduate and postgraduate studies, she has received several awards, including the Bhamashah Award and a Certificate of Appreciation from the Collector and District Magistrate of Udaipur. Her work reflects a strong commitment to research excellence and academic growth.

Tanupriya Choudhury

Dr. Tanupriya Choudhury is an accomplished academic with extensive experience in teaching and research in Computer Science. He holds a bachelor's degree from West Bengal University of Technology, a master's degree from Dr. MGR University, and a Ph.D. completed in 2016. He has over 17 years of experience in areas such as human computing, soft computing, cloud computing, and data mining. He currently serves in a senior academic role at UPES Dehradun. His research contributions are significant, with over 250 published papers in national and international forums. He has also filed 25 patents and received 26 copyrights for software innovations. Dr. Choudhury has received several recognitions, including the Global Outreach Education Award for Excellence (Best Young Researcher Award, 2018). He actively participates in academic conferences as a TPC member and session chair. He is a lifetime member of IETA, a senior member of IEEE, and a member of IET (UK). Additionally, he serves as a Technical Adviser to several organizations and holds leadership roles in professional bodies. His work reflects strong dedication to research, innovation, and academic excellence.

S. Balamurugan

Dr. S. Balamurugan, PhD, SMIEEE, is the Director of Intelligent Research Consultancy Services (iRCS), a technology research and consulting firm based in Coimbatore, India. He earned his B.Tech., M.Tech., and Ph.D. degrees from PSG College of Technology, Anna University. With over 20 years of research experience, he specializes in emerging technologies such as artificial intelligence, machine learning, robotics, data science, AR/VR, and IoT. He has edited more than 75 books and contributed approximately 300 research articles and 300 patents. His expertise supports technology forecasting and strategic decision-making for industries and startups. Dr. Balamurugan is actively involved in academic and professional communities, serving as Vice-Chairman of the Renewable Energy Society of India and contributing to several international journals as an editor. He has received numerous national and international awards, including the Rashtriya Vidhya Gourav Gold Medal, Best Educationalist Award, Dr. A.P.J. Abdul Kalam Sadhbhavana Award, and multiple Lifetime Achievement Awards. His career reflects a strong commitment to research, innovation, and advancing technological development.