Edited by

Bruno Carpentieri • Paola Lecca

Big Data Analysis and Artificial Intelligence for Medical Sciences

WILEY

Big Data Analysis and Artificial Intelligence for Medical Sciences

Big Data Analysis and Artificial Intelligence for Medical Sciences

Edited by

Bruno Carpentieri
Free University of Bozen-Bolzano
Bozen-Bolzano
Italy

Paola Lecca
Free University of Bozen-Bolzano
Bozen-Bolzano
Italy

WILEY

This edition first published 2024
© 2024 John Wiley & Sons Ltd.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at http://www.wiley.com/go/permissions.

The right of Bruno Carpentieri and Paola Lecca to be identified as the author of the editorial material in this work has been asserted in accordance with law.

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty
While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data applied for
Hardback ISBN 9781119846536

Cover Design: Wiley
Cover Image: © Yuichiro Chino/Getty Images

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

Contents

List of Contributors *xiii*
Preface *xix*

1 Introduction *1*
Bruno Carpentieri and Paola Lecca
1.1 Disease Diagnoses *4*
1.2 Drug Development *6*
1.3 Personalized Medicine *6*
1.4 Gene Editing *7*
Author Biographies *9*
References *9*

2 Fuzzy Logic for Knowledge-Driven and Data-Driven Modeling in Biomedical Sciences *17*
Paolo Cazzaniga, Simone Spolaor, Caro Fuchs, Marco S. Nobile and Daniela Besozzi
2.1 Introduction *17*
2.2 Fuzzy Logic *18*
2.2.1 Fuzzy Sets *19*
2.2.2 Linguistic Variables *19*
2.2.3 Fuzzy Rules *20*
2.2.4 Fuzzy Inference Systems *21*
2.2.5 Simpful *22*
2.3 Knowledge-Driven Modeling *22*
2.3.1 Dynamic Fuzzy Modeling *23*
2.3.2 Application 1: Maximizing Cancer Cells Death with Minimal Drug Combinations *25*
2.3.3 FuzzX: A Hybrid Mechanistic-Fuzzy Modeling and Simulation Engine *27*
2.3.4 Application 2: Analyzing Oscillatory Regimes in Signal Transduction Pathways *29*
2.4 Data-Driven Modeling *30*
2.4.1 pyFUME: Automatic Generation of Fuzzy Inference Systems *31*
2.4.2 Application 3: Assessing Tremor Severity in Neurological Disorders *33*

2.5 Discussion *35*
Author Biographies *36*
References *37*

3 Application of Machine Learning Algorithms to Diagnosis and Prognosis of Chronic Wounds *43*
Mai Dabas and Amit Gefen
3.1 Background *43*
3.1.1 Chronic Wounds *43*
3.1.2 Implementation of AI Methodologies in Wound Care and Management *43*
3.2 Clinical Visual Assessment of Wounds Supported by Artificial Intelligence *44*
3.2.1 Predicting the Formation and Progress of Wounds Based on Electronic Health Records *46*
3.2.2 Predicting the Formation and Evolution of Wounds Based on a Dynamic Evaluation of Wound Characteristics and Relevant Physiological Measures *48*
3.2.3 Feasible Implementation of AI Solutions For Wound Care Delivery and Management *49*
3.2.4 Types of Data Modalities for Diagnosis, Detection, and Prediction of Chronic Wounds *50*
3.3 Smartphone and Tablet Use in Wound Diagnosis and Management *51*
3.4 Conclusions *53*
Acronyms *54*
Author Biographies *55*
References *55*

4 Deep Learning Techniques for Gene Identification in Cancer Prevention *59*
Eleonora Lusito
4.1 The Next-Generation Era of Cancer Investigation *59*
4.1.1 Cancer at Its First Definitions *59*
4.1.2 Attempts to Sequence Nucleic Acids Over the Years *60*
4.1.3 From the First to the Third-Generation Sequencing *61*
4.1.4 Applications of NGS in Clinical Oncology *62*
4.2 Deep Learning Approaches for Genomic Variants Identification in Cancer *63*
4.2.1 Cancer Causing Factors *63*
4.2.2 The Contribution of Germline Alterations to Cancer *64*
4.2.3 Somatic Mutations and Cancer *64*
4.2.4 Calling Variants from Sequence Data *65*
4.2.5 Computational Approaches for Variant Discovery *65*
4.2.6 Convolutional Neural Networks (CNNs): Basic Principles *66*
4.2.7 Application of CNNs to Variant Calling *67*
4.2.8 A Typical CNN Architecture for Variant Calling *68*
4.2.9 The Activation Function *69*
4.2.10 Dropout and L1–L2 Regularization *71*
4.2.11 Advantages of Deep Learning Over the Existing Techniques *72*

4.2.12 Residual Neural Networks (ResNet)-Inspired CNN in Genomic Variants Detection *73*
4.3 Deep Learning in Cancer Transcriptomics *74*
4.3.1 Gene Expression and Cancer *74*
4.3.2 Analytical Approaches to Deal with Gene Expression Data *76*
4.3.3 Stacked Denoising Autoencoders (SDAEs) for Dimensionality Reduction *76*
4.3.4 The Variational Autoencoder (VAE) *79*
4.3.5 VAEs to Integrate Gene Expression and Methylation Data *81*
4.3.5.1 DNA Methylation: the Epigenetic Regulation of Gene Expression *81*
4.3.5.2 Preprocessing Input Data of Different Sources *82*
4.3.5.3 A VAE Architecture for Multimodal Data *82*
4.4 Conclusions *84*
Acronyms *86*
Author Biographies *87*
References *87*

5 Deep Learning for Network Biology *97*
Eleonora Lusito
5.1 Types of Interactions Between Genes and Their Products *97*
5.2 Deep Learning Methods with Graph-input Data *99*
5.2.1 Graph Embedding *99*
5.2.1.1 Random Walk-Based Graph Embedding *100*
5.2.1.2 Proximity-Based Graph Embedding *101*
5.2.2 Graph Convolutional Networks (GCNs) *102*
5.3 Applications of GNNs to Infer Biological and Pharmacological Interactions *104*
5.3.1 Proteomics *104*
5.3.2 Drug Development and Repurposing *104*
5.3.3 Drug–Drug Interaction Prediction *105*
5.3.4 Disease Classification and Outcome Prediction *106*
Author Biography *107*
References *107*

6 Deep Learning-Based Reduced Order Models for Cardiac Electrophysiology *115*
Stefania Fresca, Luca Dedè and Andrea Manzoni
6.1 Overview of Cardiac Physiology *115*
6.1.1 Atrial Tachycardia and Atrial Fibrillation *117*
6.1.2 Mathematical Models for Cardiac Electrophysiology *118*
6.2 Reduced Order Modeling *121*
6.2.1 Problem Formulation *123*
6.2.2 Nonlinear Dimensionality Reduction *123*
6.3 Decreasing Complexity in Cardiac Electrophysiology *124*
6.3.1 POD-Enhanced Deep Learning-Based ROMs *125*
6.3.1.1 POD-DL-ROM Architecture and Algorithms *128*

6.4 Numerical Results *130*
6.4.1 Test 1: Two-Dimensional Slab with Figure of Eight Reentry *131*
6.4.2 Test 2: Three-Dimensional Left Ventricle Geometry *133*
6.4.3 Test 3: Left Atrium Surface by Varying the Stimuli Location *135*
6.4.4 Test 4: Reentry Breakup *137*
6.5 Conclusions *139*
Author Biographies *140*
References *140*

7 The Potential of Microbiome Big Data in Precision Medicine: Predicting Outcomes Through Machine Learning *149*
Silvia Turroni and Simone Rampelli
7.1 The Gut Microbiome: A Major Player in Human Physiology and Pathophysiology *149*
7.2 Machine Learning Applied to Microbiome Research *151*
7.2.1 Case Study 1: Obesity *151*
7.2.2 Case Study 2: Cancer *153*
7.2.3 Case Study 3: Personalized Nutrition *154*
7.2.4 Case Study 4: Exploiting the Meta-Community Theory for New Machine Learning Approaches *155*
7.3 Conclusions and Perspectives *155*
Author Biographies *156*
References *156*

8 Predictive Patient Stratification Using Artificial Intelligence and Machine Learning *161*
Thanh-Phuong Nguyen, Thanh T. Giang, Quang T. Pham and Dang H. Tran
8.1 Overview of Artificial Intelligence for Patient Stratification *161*
8.2 A RPCA and MKL Combination Model for Patient Stratification *164*
8.2.1 Robust Principal Component Analysis *164*
8.2.2 Dimensionality Reduction and Features Extraction Based on RPCA *166*
8.2.3 Predictive Model Construction Based on Multiple Kernel Learning *168*
8.2.4 Materials *169*
8.2.4.1 Cancer Patient Datasets *169*
8.2.4.2 Alzheimer Disease Patient Datasets *170*
8.2.5 Experiment Design *171*
8.2.5.1 Experiment of Stratifying Cancer Patients *171*
8.2.5.2 Experiment of Stratifying Alzheimer Disease Patients *171*
8.2.6 Results and Discussions *171*
8.2.6.1 Application of Stratifying Cancer Patients *172*
8.2.7 Application of Stratifying Alzheimer Disease Patients *174*
8.3 Conclusion *175*
Author Biographies *175*
References *176*

9 Hybrid Data-Driven and Numerical Modeling of Articular Cartilage *181*
Seyed Shayan Sajjadinia, Bruno Carpentieri and Gerhard A. Holzapfel
9.1 Introduction *181*
9.2 Knee and Cartilage *182*
9.2.1 Main Joint Substructures *182*
9.2.2 Load-Bearing Cartilage Phases *183*
9.3 Physics-Based Modeling *185*
9.3.1 Numerical Modeling *185*
9.3.2 Constitutive Modeling *188*
9.4 AI-Enhanced Modeling *191*
9.4.1 Deep Learning *191*
9.4.2 Surrogate Modeling *192*
9.5 Discussion and Conclusion *194*
Author Biographies *194*
References *195*

10 A Hybrid of Differential Evolution and Minimization of Metabolic Adjustment for Succinic and Ethanol Production *205*
Zhang N. Hor, Mohd S. Mohamad, Yee W. Choon, Muhammad A. Remli and Hairudin A. Majid
10.1 Introduction *205*
10.2 Method *206*
10.2.1 Differential Evolution (DE) *206*
10.2.2 Mutation *206*
10.2.3 Crossover *207*
10.2.4 Selection *208*
10.2.5 Minimization of Metabolic Adjustment *208*
10.2.6 A Hybrid of Differential Evolution and Minimization of Metabolic Adjustment *209*
10.3 Experiments and Discussion *209*
10.3.1 Dataset *209*
10.3.2 Parameter Setting *209*
10.3.3 Experimental Results *210*
10.3.4 Comparative Analysis *214*
10.4 Conclusion *214*
Acknowledgment *215*
Author Bibliographies *215*
References *216*

11 Analysis Pipelines and a Platform Solution for Next-Generation Sequencing Data *219*
Víctor Duarte, Alesandro Gómez and Juan M. Corchado
11.1 Introduction *219*
11.2 NGS Data Analysis Pipeline and State of the Art Tools *220*

11.2.1 Quality Assessment *220*
11.2.2 Alignment *221*
11.2.3 Post-alignment and pre-variant Calling Processing *222*
11.2.4 Variant Calling *223*
11.2.5 Variant Annotation *228*
11.3 Nanopore Sequencing Data Analysis *229*
11.3.1 Base-Calling *230*
11.3.2 Quality Control and Preprocessing *230*
11.3.3 Error Correction *231*
11.3.4 Alignment *231*
11.3.5 Variant Calling *231*
11.4 Machine Learning Approaches in Variant Calling *232*
11.5 Next-Generation Sequencing Data Analysis Frameworks *233*
11.6 DeepNGS *235*
11.6.1 Pipeline *235*
11.6.2 DeepNGS Main Features *236*
11.6.2.1 Power and Speed *236*
11.6.2.2 Optimized Workflow *236*
11.6.2.3 Intuitive Design and Interactive Charts *237*
11.6.2.4 Extended Information *237*
11.6.2.5 Artificial Intelligence and Machine Learning *237*
11.7 Conclusions *240*
Author Biographies *241*
References *241*

12 Artificial Intelligence: From Drug Discovery to Clinical Pharmacology *253*
Paola Lecca
12.1 Artificial Intelligence and the Druggable Genome *253*
12.2 Feature-Based Methods *257*
12.3 Similarity/Distance-Based Methods *257*
12.4 Matrix Factorization *258*
12.4.1 Causal K-Nearest-Neighborhood *261*
12.4.2 Causal Random Forests *263*
12.4.3 Causal Support Vector Machine *264*
12.5 Opportunities and Challenges *265*
Author Biography *266*
References *266*

13 Using AI to Steer Brain Regeneration: The Enhanced Regenerative Medicine Paradigm *273*
Gabriella Panuccio, Narayan P. Subramaniyam, Angel Canal-Alonso, Juan M. Corchado and Carlo Ierna
13.1 The Challenge of Brain Regeneration *273*
13.2 The *Enhanced* Regenerative Medicine Paradigm *274*

13.3 The Case of Epilepsy *276*
13.4 AI to Understand Epilepsy *279*
13.4.1 Commonly Applied Learning Algorithms for Basic Neuroscience and Clinical Application in Epilepsy *282*
13.4.2 Seizure and Epilepsy Type Classification *284*
13.4.3 Seizure Onset Zone Localization *284*
13.4.4 Seizure Detection *285*
13.4.5 Seizure Prediction *285*
13.4.6 Signal Feature Extraction for Seizure Detection and Prediction *288*
13.4.7 Network Interactions and Evolving Dynamics in the Epileptic Brain: The Eye of AI *290*
13.5 Artificial Intelligence to Guide Graft-Host Dynamics in Epilepsy *292*
13.6 Challenges and Limitations *294*
13.6.1 From AI to Explainable AI *295*
13.7 A Philosophical Perspective on Enhanced Brain Regeneration *297*
Acknowledgments *299*
Acronyms *299*
Author Biographies *300*
References *300*

14 Towards Better Ways to Assess Predictive Computing in Medicine: On Reliability, Robustness, and Utility *309*
Federico Cabitza and Andrea Campagner
14.1 Introduction *309*
14.2 On Ground Truth Reliability *311*
14.2.1 Weighted Reliability *314*
14.2.2 Example Application *316*
14.3 On Utility Metrics to Evaluate ML Performance *318*
14.3.1 Weighted Utility *318*
14.3.2 Example Application *321*
14.4 On the Replicability of Clinical ML Models *322*
14.4.1 Dataset Size *323*
14.4.2 Dataset Similarity *325*
14.4.3 Meta-Validation Procedure *325*
14.4.4 Example Application *328*
14.5 Conclusions and Future Outlook *331*
Author Biographies *332*
References *333*

15 Legal Aspects of AI in the Biomedical Field. The Role of Interpretable Models *339*
Chiara Gallese
15.1 Introduction *339*
15.2 Data Protection *340*
15.3 Transparency Principle *343*

15.3.1 Right of Explanation *343*
15.3.2 Right of Information *348*
15.3.3 Informed Consent Requirements *349*
15.4 Accountability Principle *350*
15.5 Non-discrimination Principle and Biases *351*
15.6 High-Risk Systems and Human Oversight *353*
15.7 Additional Requirements of the AI Act Proposal *354*
15.8 Interpretability as a Standard *355*
15.9 Conclusion *358*
Author Biography *358*
References *359*

16 The Long Path to Usable AI *363*
Barbara Di Camillo, Enrico Longato, Erica Tavazzi and Martina Vettoretti
16.1 Promises and Challenges of Artificial Intelligence in Healthcare *363*
16.2 Deployment of Usable Artificial Intelligence Models *367*
16.2.1 Case Study: Predicting the Cardiovascular Complications of Diabetes via a Deep Learning Approach *368*
16.3 Potential and Challenges of Employing Longitudinal Clinical Data in AI *375*
16.3.1 Case Study: Modeling the Progression of Amyotrophic Lateral Sclerosis Through a Dynamic Bayesian Network *378*
16.3.2 Case Study: Investigating Amyotrophic Lateral Sclerosis Progression Trajectories Leveraging Process Mining *381*
16.4 Enhancing the Applicability of AI Predictive Models by a Combined Model Approach: A Case Study on T2D Onset Prediction *386*
16.4.1 The Problem of Type 2 Diabetes Prediction *386*
16.4.2 Potential Applications of T2D Predictive Models *387*
16.4.3 Barriers to the Adoption of T2D Predictive Models *387*
16.4.4 Addressing Practical Issues by Combining Multiple T2D Predictive Models *388*
16.4.5 The Combined Model Achieves High Prediction Performance with High Coverage *390*
16.5 Conclusions and Future Outlook *391*
Author Biography *392*
References *393*

Index *399*

List of Contributors

Daniela Besozzi

Department of Informatics
Systems and Communication
University of Milano-Bicocca
Milano
Italy

and

Bicocca Bioinformatics Biostatistics and
Bioimaging Centre - B4
University of Milano-Bicocca
Vedano al Lambro
Italy

Federico Cabitza

Department of Computer Science
Systems and Communication
University of Milano-Bicocca
Milan
Italy

and

IRCCS Istituto Ortopedico Galeazzi
Milan
Italy

Barbara Di Camillo

Department of Information Engineering
University of Padova
Padua
Italy

and

Department of Comparative Biomedicine
and Food Science
University of Padova
Padua
Italy

Andrea Campagner

IRCCS Istituto Ortopedico Galeazzi
Milan
Italy

Angel Canal-Alonso

BISITE Research Group
University of Salamanca
Salamanca
Spain

and

Institute for Biomedical Research of
Salamanca
University of Salamanca
Salamanca
Spain

Bruno Carpentieri

Faculty of Engineering
Free University of Bozen-Bolzano
Bolzano
Italy

Paolo Cazzaniga

Department of Human and Social Sciences
University of Bergamo
Bergamo
Italy

and

Bicocca Bioinformatics Biostatistics and Bioimaging Centre - B4
University of Milano-Bicocca
Vedano al Lambro
Italy

Yee W. Choon

Artificial Intelligence Lab
Institute for Artificial Intelligence and Big Data
Universiti Malaysia Kelantan
Kota Bharu
Kelantan
Malaysia

and

Faculty of Data Science and Computing
Universiti Malaysia Kelantan
Kota Bharu
Kelantan
Malaysia

Juan M. Corchado

BISITE Research Group
University of Salamanca
Salamanca
Spain

and

Institute for Biomedical Research of Salamanca
University of Salamanca
Salamanca
Spain

and

Air Institute
IoT Digital Innovation Hub
Salamanca
Spain

and

Department of Electronics
Information and Communication
Faculty of Engineering
Osaka Institute of Technology
Osaka
Japan

Mai Dabas

Department of Biomedical Engineering
Faculty of Engineering
Tel Aviv University
Tel Aviv
Israel

Luca Dedè

MOX - Dipartimento di Matematica
Politecnico di Milano
Milano
Italy

Víctor Duarte

BISITE Research Group
University of Salamanca
Salamanca
Spain

and

Air Institute
Salamanca
Spain

Stefania Fresca

MOX - Dipartimento di Matematica
Politecnico di Milano
Milano
Italy

Caro Fuchs

Department of Industrial Engineering and Innovation Sciences
Eindhoven University of Technology
Eindhoven
The Netherlands

Chiara Gallese

Department of Electrical Engineering
Eindhoven University of Technology
Eindhoven
Netherlands

Amit Gefen

Department of Biomedical Engineering
Faculty of Engineering
Tel Aviv University
Tel Aviv
Israel

and

Skin Integrity Research Group (SKINT)
University Centre for Nursing and Midwifery
Department of Public Health and Primary Care
Ghent University
Ghent
Belgium

and

Department of Mathematics and Statistics
Faculty of Sciences, Hasselt University
Hasselt
Belgium

Thanh Trung Giang

Tay Bac University
VNU University of Engineering and Technology
Hanoi
Vietnam

Alesandro Gómez

BISITE Research Group
University of Salamanca
Salamanca
Spain

and

Air Institute
IoT Digital Innovation Hub
Salamanca
Spain

Gerhard A. Holzapfel

Institute of Biomechanics
Graz University of Technology
Graz
Austria

and

Department of Structural Engineering
Norwegian University of Science and Technology
Trondheim
Norway

Zhang N. Hor

Faculty of Computing
Universiti Teknologi Malaysia
Johor
Malaysia

Carlo Ierna

Radboud University Nijmegen
Faculty of Philosophy, Theology, and Religious Studies
Center for the History of Philosophy and Science
Nijmegen
The Netherlands

Paola Lecca

Faculty of Engineering
Free University of Bozen-Bolzano
Bolzano
Italy

Enrico Longato

Department of Information Engineering
University of Padova
Padua
Italy

Eleonora Lusito

San Raffaele Telethon Institute for Gene Therapy (SR-Tiget)
IRCCS San Raffaele Scientific Institute
Milan
Italy

Hairudin A. Majid

Faculty of Computing
Universiti Teknologi Malaysia
Johor
Malaysia

Andrea Manzoni

MOX - Dipartimento di Matematica
Politecnico di Milano
Milano
Italy

Mohd S. Mohamad

Health Data Science Lab
Department of Genetics and Genomics
College of Medical and Health Sciences
United Arab Emirates University
Al Ain
Abu Dhabi
United Arab Emirates

and

Big Data Analytics Center
United Arab Emirates University
Al Ain
Abu Dhabi
United Arab Emirates

Thanh-Phuong Nguyen

Innovation Business Unit
Dennemeyer TechSys, Luxembourg
Luxembourg

Marco S. Nobile

Department of Environmental Sciences
Informatics and Statistics
Ca' Foscari University of Venice
Venice
Italy

and

Bicocca Bioinformatics Biostatistics and Bioimaging Centre - B4
University of Milano-Bicocca
Vedano al Lambro
Italy

Gabriella Panuccio

Istituto Italiano di Tecnologia
Enhanced Regenerative Medicine
Genova
Italy

Quang Trung Pham

Tay Bac University
Sonla
Vietnam

Simone Rampelli

Unit of Microbiome Science and Biotechnology
Department of Pharmacy and Biotechnology
University of Bologna
Bologna
Italy

Muhammad A. Remli

Artificial Intelligence Lab
Institute for Artificial Intelligence and Big Data
Universiti Malaysia Kelantan
Kota Bharu
Kelantan
Malaysia

and

Faculty of Data Science and Computing
Universiti Malaysia Kelantan
Kota Bharu
Kelantan
Malaysia

Seyed Shayan Sajjadinia

Faculty of Engineering
Free University of Bozen-Bolzano
Bozen-Bolzano
Italy

Simone Spolaor

Microsystems, Eindhoven University of Technology
Eindhoven
The Netherlands

Narayan P. Subramaniyam

Faculty of Medicine and Health Technology and BioMediTech Institute
Tampere University
Tampere
Finland

Erica Tavazzi

Department of Information Engineering
University of Padova
Padua
Italy

Dang Hung Tran

Hanoi National University of Education
Hanoi
Vietnam

Silvia Turroni

Unit of Microbiome Science and Biotechnology
Department of Pharmacy and Biotechnology
University of Bologna
Bologna
Italy

Martina Vettoretti

Department of Information Engineering
University of Padova
Padua
Italy

Preface

Our modern society is characterized by an unprecedented ability to generate vast amounts of data. The use of *big data* in science is driving the development of a new scientific paradigm. Smart search and learning computer algorithms are utilized to extract meaningful patterns from large datasets and generate new knowledge that can be applied to model the behavior of complex real-world systems much faster than by using traditional scientific laws and theories. The name of the new game is *machine learning, deep learning, and artificial intelligence*. These data-driven research methodologies are already paving the way for advanced discoveries in numerous scientific disciplines. In healthcare industry and research, the data-driven modeling approach is opening new frontiers, for example, enabling to produce more accurate diagnoses, to facilitate the design of drugs, to innovate treatment protocols and prevent diseases, to produce personalized treatments and reduce medical costs, thereby significantly advancing medical research.

Computational science, the scientific investigation and solution of complex problems in science and engineering through modeling and simulation on computers, may be considered the third pillar of science, complementing theory and experimenting. The conventional simulation approach refers to the theory-based approach in which a model is built using laws and predictive statements from physics, chemistry, biology, and other fields that describe the causal relationships between a set of controllable inputs and a set of target variables. The underlying physics is described by mathematical models such as systems of ordinary and/or partial differential equations. The model is then solved numerically unless a closed-form solution to the resulting set of equations is available, which is rarely the case in practice. Theory-based simulation models are generally very powerful in understanding the behavior of the system. However, they sometimes fail to accurately reveal the properties of a complex system due to lack of theory and simplified assumptions, large numbers of variables and parameters involved in the simulation and, sometimes, a lack of robust numerical solvers. In these circumstances, data-driven models can be used to identify correlations between two sets of controlled input and output variables without the need to explicitly describe their causal pips. Data-mining techniques, for instance, can be used to predict future data patterns by analyzing the properties of existing datasets. Genetic algorithms and artificial neural networks can map relationships between two datasets by reducing a cost or error function and then predicting the future behavior of the target system.

Several scientific disciplines, such as computer vision and image recognition, self-driving cars, natural language processing, website recommendations, solid-state materials science,

finance, bioinformatics, and chemistry, to name a few, have adopted machine learning algorithms in the past decade or so. Thanks to their general applicability to different applications, data-driven models are computationally attractive for use in the medical sciences and the healthcare industry, where they are becoming increasingly popular. Examples of recent studies using data-driven neural network models include better image classification of coronary angiography X-rays, localization of brain tumors from MRI image slices, and strabismus recognition. Support vector machines have assisted in the detection of common pneumothorax by analyzing binary patterns in chest X-ray images, as well as in the prediction of fractures in hip bones and vertebrates using datasets acquired from random and cluster-based under-sampling methods. Blood pressure has been predicted using principal component analysis. The application of fuzzy logic and ontological reasoning allows for more precise, personalized recommendations of antidiabetic drugs for individual patients. Projects like IBM's Watson Oncology, Microsoft's Hanover, and Google's DeepMind are only a few examples of the many programs that companies are developing to leverage big data in the healthcare industry.

This book provides an overview of the current state of the art in the use of artificial intelligence in medicine and biology. It collects chapters written by international experts in the field of medical and biological research. Their studies are the result of years of interdisciplinary collaborations with clinicians as well as computer scientists, mathematicians, and engineers. The aim of the book is to demonstrate the efforts made in the fields of computational biology and medical sciences to design and implement robust, accurate, and efficient computer algorithms for modeling the behavior of complex biological systems much faster than using traditional modeling approaches solely based on theory. Through the authors' contributions in the various chapters, we aim to highlight the difference between traditional computational approaches to data processing (those of mathematical biology) and the new way knowledge is extracted from data, and the experiment-data-theory-model-validation cycle is being implemented. The style of the book is not that of a typical textbook. We believe that understanding these new trends, the difficulties that have arisen as a result of these changes, and the potential future directions these changes may take, directly through the authors' reports of scientific work expressed in simple but rigorous language, may add a remarkable breadth. It may be of great benefit not only to professional scholars but also to MSc or PhD program students who are the future and those who will take up the baton to continue the race in scientific and technological research.

Bolzano
8 July 2023

Bruno Carpentieri and *Paola Lecca*
The Editors

1

Introduction

Bruno Carpentieri and Paola Lecca

Faculty of Engineering, Computer Science and Artificial Intelligence Institute, Free University of Bozen-Bolzano, Bolzano, Italy

The concept of intelligent machines is frequently attributed to Alan Turing, who published a seminal paper titled "Computing Machinery and Intelligence" in 1950, in which he developed a simple test known as the "Turing test" to assess whether a machine can demonstrate human-like intelligence. Six years later, in 1956, during the Dartmouth Conference, an influential event in the history of AI, the term "artificial intelligence" (AI) was coined by emeritus Stanford Professor John McCarthy, known as the "Father of AI" to characterize "the science and engineering of creating intelligent machines." The Turing test has had a significant impact on the development of modern AI by establishing a standard for measuring progress in AI research. Nevertheless, AI encompasses a broader spectrum of methods, concepts, and technologies. Using techniques, such as machine learning (ML), natural language processing (NLP), computer vision, and others, entail the study and development of systems that can perform tasks that typically require human intelligence. Early basic AI systems relied on explicitly coded rules based on a simple set of "if, then" or symbolic reasoning approaches, in which particular conditions would trigger specific actions to make judgments and to perform tasks. These early models necessitated considerable manual rule programming, which was time-consuming and difficult to scale to complex problems. As a result of these limitations, widespread adoption of early AI models proved difficult, particularly in complicated domains such as medicine. Advances in AI research have led to the development of more sophisticated algorithms that function similarly to the human brain and have helped address some of these challenges and opened up new possibilities for AI applications. ML has evolved into a field known as deep learning (DL), which consists of techniques for creating and training artificial neural networks (ANNs) with multiple layers of interconnected nodes, also known as neurons, capable of learning and making decisions independently, similar to the human brain. These neural networks are inspired by the structure and operation of biological neural networks in the human brain, although they do not completely replicate the human brain's complexities and mechanisms. By iteratively adjusting the weights and biases of the interconnected neurons, DL algorithms are able to recognize complex patterns, extract meaningful representations from large amounts of raw data, and make decisions or predictions across multiple domains. This has produced

Big Data Analysis and Artificial Intelligence for Medical Sciences, First Edition.
Edited by Bruno Carpentieri and Paola Lecca.
© 2024 John Wiley & Sons Ltd. Published 2024 by John Wiley & Sons Ltd.

extraordinary progress in numerous fields, including computer vision, NLP, speech recognition, and medical sciences.

The significant breakthroughs of DL methods can be attributed to the early 2000s owing to the availability of large datasets, increased computational power, and advancements in parallel computing, in particular with the advent of graphics processing units (GPUs), which played a crucial role in training deep neural networks on a larger scale. DL is now a dominant approach at the forefront of AI research, with applications in a variety of disciplines. In the medical field, it has shown the potential to revolutionize healthcare and pave the way for personalized medicine (Gilvary et al. 2019). The use of predictive models, advanced data analytics, and DL algorithms can provide valuable insights for healthcare applications such as diagnosis, treatment selection, and therapy response prediction. The ability to analyze vast quantities of patient data, including medical records, genetic information, imaging data, and real-time sensor data, is one of the primary benefits of AI in medicine (Zeng et al. 2021; Liu et al. 2021b; Ahmad et al. 2021; Hamet and Tremblay 2017). This data can guide interventions and preventive measures to reduce risks and promote proactive healthcare, enhance clinical workflow, and procedure precision. On the basis of the analysis of multiple risk factors, it can be possible to assess an individual's likelihood of developing specific diseases. In the context of medicine and healthcare, however, data-driven models present significant computational challenges. When the model is too complex or the training dataset is too small relative to the model's capacity, it may begin to capture noise or oddities that are specific to the training data, and perform exceptionally well on the training data but poorly on new, unseen data. Such models are called "overfit." Noise and unpredictability are common features of complex healthcare datasets. A DL model that overfits to these details when applied to new patient data may produce erroneous or unreliable predictions. In healthcare, researchers are actively investigating methods to reduce overfitting and to ensure the robustness and dependability of DL models across diverse patient populations and situations. The availability of larger datasets, transfer learning techniques, and advances in model architecture and regularization methods are important factors to mitigate overfitting concerns and facilitate the adoption of DL in the medical field.

Convolutional neural networks (CNNs) were an additional significant advancement and a subclass of DL algorithms used in image processing that were designed specifically for analyzing visual data, such as images. Inspired by the structure and operation of the human visual cortex, they imitate the activity of networked neurons by employing layers of interconnected nodes, known as "convolutional layers," which learn spatial hierarchies of features from the input data. Convolutional layers apply filters or kernels to input images, extracting and preserving local features and spatial relationships. In subsequent layers, these extracted features are combined and further processed to capture increasingly complex patterns and structures. Typically, the final layers of a CNN are composed of fully connected layers and are responsible for making predictions based on the learned features. CNNs have revolutionized image processing and computer vision tasks, outperforming traditional machine learning approaches in image classification, object detection, segmentation, and other tasks. Their ability to automatically acquire features from raw image data has made them extremely valuable in numerous applications, including autonomous vehicles, surveillance and medical imaging, and others. One of the first successful CNN architectures was LeNet-5 (LeCun et al. 1998), introduced by Yann LeCun et al. in 1998,

primarily designed for handwritten digit recognition. Other popular CNN models are AlexNet (Krizhevsky et al. 2017) (developed by Alex Krizhevsky et al.), that made a breakthrough in the field by significantly lowering error rates; VGGNet (Simonyan and Zisserman 2014) (developed at the University of Oxford by the Visual Geometry Group); GoogLeNet (Szegedy et al. 2015) (introduced by Christian Szegedy et al. from Google), that used parallel convolutional operations at different scales; ResNet (He et al. 2016) (proposed by Kaiming He et al.), that enabled the successful training of networks with hundreds or even thousands of layers; DenseNet (Huang et al. 2017) proposed by Gao Huang et al., and MobileNet (Howard et al. 2017) introduced by Andrew G. Howard et al. in 2017. These are only a few examples of popular CNN architectures. Numerous other CNN models have been developed over the years to address various applications, performance demands and computational constraints, and demonstrate great potential in the field of medicine. Their ability to analyze and interpret medical images, such as X-rays, computerized tomography (CT) scans, magnetic resonance imaging (MRI), and pathology slides, can assist in the diagnosis, planning of treatment, and monitoring of disease. In recent years, CNNs have been used for a variety of medical imaging applications, including image classification (classify medical images to identify different types of tumors, lesions, or diseases), segmentation (segment medical images to identify regions or specific structures of interest, such as organs, tumors, or blood vessels with the purpose of surgical planning, radiation therapy, disease progression study), object detection (for detecting abnormalities, nodules, or lesions within medical images), and disease prediction and prognosis (predict the likelihood of disease occurrence and its progression based on medical images and other clinical data), only to name a few.

As a result of these advancements, today, we are entering a new era in medicine in which risk assessment models can be implemented in clinical practice to improve diagnostic accuracy and operational efficiency. Kaul et al. coined the acronym "AIM" which stands for "Artificial Intelligence in Medicine" in a paper published in 2020 on gastrointestinal endoscopy, titled "History of artificial intelligence in medicine" (Kaul et al. 2020), an eloquent fact of the emergence of a new strand in computational science applied to the life sciences. According to Kaul and coauthors in that research, the critical advances came in the last two decades although AIM has undergone significant change during the last five decades. Watson, an open-domain question–answering system developed by IBM in 2007, competed against human contestants on the television game show Jeopardy! in 2011. Unlike conventional systems, which relied on either forward reasoning (following rules from data to conclusions), backward reasoning (following rules from conclusions to data), or manually created "if-then" rules, this technology, known as DeepQA (Ferrucci et al. 2010), used NLP and a variety of searches to analyze data over unstructured content and produce likely answers. This approach was less complicated and less expensive to use, and easier to maintain. Using IBM Watson, a novel RNA-binding protein that was changed in amyotrophic lateral sclerosis, was successfully discovered by Bakkar et al. in 2017. DeepQA technology could be applied to give evidence-based medicine solutions using data from a patient's electronic medical record and other electronic sources. This opened up new opportunities for evidence-based clinical decision-making. Digitalized medicine became more widely accessible thanks to advances in computer hardware and software, and AIM rapidly expanded as a result

of this momentum. Chatbots were originally created for surface-level communication (Eliza), but NLP has transformed them into useful conversation-based interfaces. This technology was utilized to create Apple's Siri and Amazon's Alexa in 2011 and 2014, respectively. Mandy was launched in 2017 as an automated patient intake technology for a primary care practice, while Pharmabot was created in 2015 to assist with medication instruction for pediatric patients and their parents (Ni et al. 2017; Comendador et al. 2015).

The use of AI in the medical field is rapidly expanding. CardioAI (the first Arterys product (Arterys 2018)) analyses cardiac MRI and provides details like the cardiac ejection percent. The tool incorporates non-contrast CT pictures of the head, chest, and musculoskeletal systems, as well as liver and lung imaging. In 2017, the US Food and Drug Administration approved Arterys (currently acquired by Tempus Radiology) as the first clinical cloud-based DL application in healthcare. DL can help to locate lesions, make differential diagnoses, and generate automated medical reports. With fivefold cross-validation, Gargeya and Leng (2017) employed DL to screen for diabetic retinopathy in 2017, reaching a 94% sensitivity and 98% specificity (area under the curve). Esteva et al. similarly trained a CNN to differentiate between nonmelanoma and melanoma skin cancers, and the results indicate that the CNN's performance is comparable to that of specialists. Weng et al. (2017) demonstrated that a CNN can be utilized to predict cardiovascular risk in cohort populations. Astonishingly, AI has been found to improve the accuracy of cardiovascular risk prediction compared to the standard methodology specified by the American College of Cardiology. It was also used to consistently predict the progression of Alzheimer disease by analyzing amyloid imaging data and precisely predicting drug therapy response in this disease (Mathotaarachchi et al. 2017; Fleck et al. 2017).

The literature is so extensive, despite the recent emergence of AI in this field, that it is difficult to compile an exhaustive and summarizing compendium. There are numerous review and perspective articles, blog posts, and journal portfolios that discuss the medical applications of AI, see e.g. Liu et al. (2021b), Suh et al. (2022), Briganti and Le Moine (2020), and Malik et al. (2019). The majority of these meta-reviews identify four application areas: (i) disease diagnose; (ii) drug development; (iii) personalized therapies; and (iv) gene editing, as depicted in Figure 1.1 and suggested, for instance, by Markus Schmitt, head of data science at Data Revenue (https://datarevenue.com).

1.1 Disease Diagnoses

For accurate disease diagnosis, years of medical training are required. Even so, the process of diagnosis is frequently laborious and lengthy. In many fields, the demand for expertise greatly exceeds the available supply. Consequently, doctors are under pressure, and critical patient diagnoses are frequently delayed. Recent advances in machine learning, particularly in DL algorithms, have significantly enhanced the accuracy and accessibility of disease diagnosis. Using machine learning, algorithms can learn to recognize patterns in the same way that doctors do. An important distinction, however, is that learning algorithms require

Figure 1.1 The four main applications of artificial intelligence in knowledge extraction and interpretation of biological and biochemical data for applications in the clinic and medicine. The rapid advancement of AI technologies will add further application domains in the near future, formalizing AI as an integral part of modern healthcare.

thousands of concrete examples. In addition, because robots cannot read between the lines in textbooks, these examples must be neatly digitalized. Consequently, machine learning (including DL) is particularly advantageous in fields where the diagnostic data a doctor considers has already been digitized, such as:

- utilizing CT images to diagnose strokes and lung cancer (Chiu et al. 2022; Aydín et al. 2021; Zhou and Xin 2022),
- evaluating the risk of sudden cardiac death or other heart disorders using cardiac MRI and electrocardiogram (ECG) data (Haq et al. 2020; Klein et al. 2022; Ledziński and Grześk 2023; Martínez-Sellés and Marina-Breysse 2023; Karatzia et al. 2022; Yasmin et al. 2021; Kabra et al. 2022; Madan et al. 2022; Argentiero et al. 2022; Jone et al. 2022),
- identifying skin disorders from photographic images (Goyal et al. 2020; Thieme et al. 2023; Son et al. 2021; Ahmad et al. 2023; Combalia et al. 2022; Liopyris et al. 2022; Nigar et al. 2022; Sreekala et al. 2022),
- and recognizing diabetic retinopathy in photographs of the eyes (Sheng et al. 2022; Huang et al. 2022; Padhy et al. 2019; Bader Alazzam et al. 2021; Lim et al. 2023; Mohan et al. 2022; Babenko et al. 2022; Muchuchuti and Viriri 2023; Sun et al. 2023).

Because there is a vast quantity of reliable data available in these medical areas, algorithms are enhancing their diagnostic capabilities to match those of specialists. The algorithm's ability to generate conclusions in a fraction of a second and its economic replicability on a global scale make up the difference. On this basis, it is anticipated that everyone, everywhere will soon have access to affordable radiological diagnostic services of the same high quality. More sophisticated AI diagnosis is being developed. Machine learning in diagnostics is still in its early stages; more ambitious systems will combine a number of data sources (such as CT, MRI, genomics, proteomics, patient data, and even handwritten documents) to evaluate an illness or its progression. It is important to recognize, however, that it is unlikely that AI will completely replace doctors. Instead,

AI tools will assist the doctors to focus on signal interpretation, e.g. to identify potentially malignant tumors and hazardous cardiac patterns.

1.2 Drug Development

Understanding a disease's fundamental causes (technically, "the pathways") and resistance mechanisms is the first step toward designing a treatment. This earliest stage of drug discovery is also known as "target identification." Methods for identifying targets, such as genes involved in disease pathophysiology, include genome-wide association studies (GWAS), risk gene identification, and data mining of published literature. The next phase is to identify effective disease targets (typically proteins). The amount of data available for identifying feasible target pathways has significantly expanded thanks to the widespread use of high-throughput methods like short hairpin RNA (shRNA) screening and deep sequencing. However, integrating a huge number and variety of data sources, and then identifying relevant patterns, remains difficult using traditional methods. All of the available data can be analyzed more easily by machine learning algorithms, which can even be trained to recognize good target proteins automatically (You et al. 2022; Zeng et al. 2020; Najm et al. 2021; Xu et al. 2021; Liu et al. 2021a; Dezsö and Ceccarelli 2020). After identifying a drug's target, the next stage is to find a substance that can interact with the target molecule in the appropriate manner. This entails screening a large number of candidate compounds for their affinity toward the target as well as their toxicity (unintended side effects). These substances may be synthetic, bioengineered, or natural.

It requires a significant amount of time to eliminate false positives and inaccuracies, which may result in a large number of undesired recommendations (false positives). Machine learning techniques can be beneficial in this scenario because they can be trained to predict the suitability of a molecule using structural fingerprints and molecular descriptors (Arnold 2023; Paul et al. 2021; Brown et al. 2020). Then, scientists rapidly sift through millions of potential molecules to identify the most promising candidates – those with the fewest adverse effects. Ultimately, this expedites the drug design process. Finally, it is important to note that machine learning may speed up clinical trials by autonomously selecting qualified applicants and ensuring the correct distribution among participant groups. Using algorithms, one can identify patterns that distinguish between excellent and bad candidates. In addition, they can serve as an early warning system for a clinical study that is not producing reliable results, allowing researchers to intervene sooner, and potentially save the development of the drug.

1.3 Personalized Medicine

Precision medicine is regarded as crucial for the treatment of complex diseases, including systemic autoimmune diseases such as rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), and psoriatic arthritis (PsA). Despite the remarkable number of novel molecules being developed for the treatment of these diseases, the growing understanding of their pathogenesis, and the advances in early diagnosis, it is clear that the clinical and

serological heterogeneity of these diseases, as well as the large number of comorbidities that can affect them, continue to limit the ability to tailor the treatment for these patients. There are often few therapy options available for these disorders, even when many organs are involved. Treat-to-target therapy is one of the available therapeutic modalities, and it remains the optimal approach for the majority of rheumatic disorders. However, patients have varying responses to medications and treatment plans. Therefore, there is an enormous potential for personalized care to lengthen patients' lives. One issue is that it is extremely difficult to determine which characteristics should influence therapy selection. ML can automate this tedious statistical work by cross-referencing similar patients and comparing their treatments and outcomes to determine a patient's likely response to a particular therapy (Peng et al. 2021; Fröhlich et al. 2018; Quazi 2022; Papadakis et al. 2019; Emmert-Streib and Dehmer 2019; Gaur et al. 2022; Sahu et al. 2022). The resultant outcome projections make it much simpler for clinicians to formulate the optimal treatment plan. In a recent perspective paper by Sebastiani et al. (2022), it is highlighted, for instance, that treatment and identification of immune-mediated disorders have undergone significant advancements over the past decade. For the treatment of these conditions, an increasing number of novel monoclonal antibodies and small compounds have been developed. Parallel to this, a large number of novel genetic or serological markers have been identified that enhance our ability to detect autoimmune diseases at an early stage. Due to advances in AI and ML, the treatment and follow-up of certain diseases, including cancer, have significantly improved over the past decade. However, the authors of Sebastiani et al. (2022) caution that our understanding of autoimmune systemic diseases is still quite limited. Despite the significant progress in our understanding, it is currently believed that we are still a long way from providing patients with true precision medicine.

1.4 Gene Editing

Clustered regularly interspaced short palindromic repeats (CRISPR), and more specifically the CRISPR-Cas9 system for gene editing, represents an important development in our ability to accurately and economically modify DNA, much like a surgeon. This technique uses short guide RNAs (sgRNA) to target and modify a specific region of DNA. However, the guide RNA can bind to multiple DNA sites, which may have undesired consequences (off-target effects). The careful selection of guide RNA with minimal negative side effects is one of the primary obstacles to the widespread use of the CRISPR system. ML techniques have been shown to make the best predictions for a specific sgRNA's level of guide-target interactions and off-target effects (Liu et al. 2020; Das et al. 2023; Vora et al. 2022; Fong and Wong 2023; Abadi et al. 2017; Aktas et al. 2019; Wang et al. 2020).

Aim of the book is to offer a portrait of the current state of the use of AI methodology in medicine and biology, of the new contributions in terms of techniques and algorithms, and of their integration with traditional disciplines and philosophies of thought typical of other fields such as mathematics and the more classical algorithmic approaches of computer science. Recently, AI techniques have now innervated these domains to the extent that they have become integral elements of them, described, narrated, and hence, conceptualized

in terms of an AI-specific language and pattern of thought. The chapters dealing specifically with algorithmic techniques and methodological approaches address the following two consolidated topics.

1) **Data-driven and knowledge-driven modeling**: *Fuzzy Logic for Knowledge-Driven and Data-Driven Modeling in Biomedical Sciences* by Paolo Cazzaniga, Simone Spolaor, Caro Fuchs, Marco S. Nobile, and Daniela Besozzi; *Application of Machine Learning Algorithms to Diagnosis and Prognosis of Chronic Wounds* by Mai Dabas and Amit Gefen; *Deep Learning Techniques for Gene Identification in Cancer Prevention* by Eleonora Lusito; and *Deep Learning-Based Reduced Order Models for Cardiac Electrophysiology* by Stefania Fresca, Luca Dedè, and Andrea Manzoni.
2) **Data analytics: technologies and methods for data interpretation and new knowledge inference**: *Deep Learning for Network Biology* by Eleonora Lusito; *Analysis Pipelines and a Platform Solution for Next Generation Sequencing Data* by Víctor Duarte, Alesandro Gómez, and Juan Manuel Corchado; *The Potential of Microbiome Big Data in Precision Medicine: Predicting Outcomes Through Machine Learning* by Silvia Turroni and Simone Rampelli; *Hybrid Data-Driven and Numerical Modeling of Articular Cartilage* by Seyed Shayan Sajjadinia, Bruno Carpentieri, and Gerhard A. Holzapfel; *A Hybrid of Differential Evolution and Minimization of Metabolic Adjustment for Succinic and Ethanol Production* by Zhang Neng Hor, Mohd Saberi Mohamad, Yee Wen Choon, Muhammad Akmal Remli, and Hairudin Abdul Majid; *Predictive Patient Stratification Using Artificial Intelligence and Machine Learning* by Thanh-Phuong Nguyen, Thanh Trung Giang, Quang Trung Pham, and Dang Hung Tran.

There is no need to emphasize how the fields of data analytics and data modeling are intertwined and cooperate to accelerate industrial and decision-making processes in the fields of medicine, biology, pharmacology, and recently, medicinal chemistry (Struble et al. 2020; Bajorath 2021; Tyrchan et al. 2022). Alongside the relevant areas of data science in informatics and mathematics, the book contains an innovative counterpoint pertinent to the increasing support that AI techniques and methodologies are providing to the field of biomedical engineering, e.g. the chapter *Using AI to Steer Brain Regeneration: The Enhanced Regenerative Medicine Paradigm* by Gabriella Panuccio, Narayan Puthanmadam Subramaniyam, Angel Canal-Alonso, Juan Manuel Corchado, and Carlo Ierna. The book also offers new visions and perspectives on the current state of the art, performance, and industrial applications of AI techniques in the life sciences, e.g. in the chapters *Toward Better Ways to Assess Predictive Computing in Medicine: On Reliability, Robustness and Utility* by Federico Cabitza and Andrea Campagner; *Artificial Intelligence: From Drug Discovery to Clinical Pharmacology* by Paola Lecca.

Alongside the undisputed benefits that the use of AI is bringing to medicine and healthcare, there are also new problems. A recent review by Naik et al. (2022) outlines the most pressing ones stating that privacy and surveillance, interpretability of the results, bias or discrimination, and potentially the philosophical problem are among the legal and ethical issues that society faces as a result of AI (Ahmad et al. 2021; Gruson et al. 2019). As a result of their use, there are concerns that modern digital technologies will become a new source of inaccuracy and data breaches. The deployment of CNN architectures in medicine requires rigorous validation, regulatory compliance, and ethical considerations. Medical