

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/331302149>

# Data Science for Healthcare: Methodologies and Applications

Book · January 2019

DOI: 10.1007/978-3-030-05249-2

---

CITATIONS

2

---

READS

6,248

3 authors, including:



[Sergio Consoli](#)

European Commission

87 PUBLICATIONS 564 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Affective knowledge extraction [View project](#)



Refugium Peccatorum [View project](#)

# Data Science for Healthcare

Sergio Consoli • Diego Reforgiato Recupero •  
Milan Petković  
Editors

# Data Science for Healthcare

Methodologies and Applications



Springer

*Editors*

Sergio Consoli  
Philips Research  
Eindhoven, The Netherlands

Diego Reforgiato Recupero  
Dept of Mathematics and Computer Science  
University of Cagliari  
Cagliari, Italy

Milan Petković  
Data Science Department  
Philips Research  
Eindhoven, The Netherlands

ISBN 978-3-030-05248-5      ISBN 978-3-030-05249-2 (eBook)

<https://doi.org/10.1007/978-3-030-05249-2>

Library of Congress Control Number: 2018966867

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Foreword

It is becoming obvious that only by fundamentally rethinking our healthcare systems we can successfully address the serious challenges we are facing globally.

One of the most significant challenges is the aging of populations, which comes with a high percentage of chronically ill people, often with multiple conditions. In addition, there is a rising incidence of preventable lifestyle-related diseases caused by risk factors such as obesity, smoking, and alcohol consumption. Today, chronic diseases in EU already result in the loss of 3.4 million potential productive life years, which amounts to an annual loss of €115 billion for the EU economy. At the same time, we are being faced with a shortage of qualified healthcare professionals, and with quality and efficiency issues in the way healthcare is delivered. Finally, public spending on healthcare is steadily rising. The EU spends around 10% of its GDP on healthcare. In 2015, US healthcare spending increased 5.8% to \$3.2 trillion. The costs are expected to continue rising—to unaffordable levels.

We need to transition to new care delivery models, addressing the quadruple aim of (1) improving the health of populations, (2) reducing the per capita cost of healthcare, (3) improving the patient experience including quality and satisfaction, and (4) improving the work life of healthcare providers by providing necessary support.

The good news is that digital technologies are by now so powerful, affordable, and pervasive, that they help to make these goals achievable. The Internet of Medical Things and artificial intelligence (AI) in particular are key enablers of the digital transformation in healthcare. Connected medical devices will soon be everywhere, from hospital to home, providing a rich variety of data. AI will be instrumental in turning these data into actionable insights across the continuum of care.

But technology by itself will not be the answer. In the end, healthcare is all about people. Meaningful innovation occurs when technology enables professionals to deliver better care and when it empowers consumers and patients to better manage their own health. This means that applying AI and data science to healthcare requires a deep understanding of the personal, clinical, or operational context in which they are used. That is why, at Philips, we believe in the power of *adaptive intelligence*.

Adaptive intelligence combines AI with human domain knowledge to create solutions that adapt to people's needs and environments—supporting them in their daily work and lives. *Adaptive intelligence augments people, rather than replacing them.* It acts like a personal assistant that can learn and adapt to the skills and preferences of the person that uses it, and to the situation he or she is in. The technology does not call attention to itself, but runs in the background—deeply integrated into the interfaces and workflows of hospitals, and almost invisibly embedded into solutions for the consumer environment.

This is not merely a future vision—it is becoming a reality today. This book includes examples that show how data science and AI-enabled solutions are already supporting clinical care and prevention of disease or health incidents. It is very encouraging that advances in AI methods such as machine learning, natural language processing, and computer vision can all improve people's lives, when they are employed wisely.

As we continue to make strides in the digital transformation of healthcare systems, it is important to be aware of the possibilities of AI and data science—and how they can be used in an effective and responsible way to help achieve the quadruple aim. This book will help the reader to learn how to (1) extract new knowledge from health data to improve healthcare delivery, (2) enable healthcare systems to deliver better outcomes at lower costs, and (3) support the transition from an acute, episodic care model to proactive chronic disease management.

Enjoy the read, and join this exciting journey!

Chief Technology Officer, Philips  
Eindhoven, The Netherlands

Henk van Houten

# Preface

Healthcare systems around the world are facing vast challenges in responding to trends of aging population, the rise of chronic diseases, resources constraints, and the growing focus of citizens on healthy living and prevention. Consequently, there is an increasing focus on answering important questions such as: (1) How do we improve the rate of fast, accurate first-time-right diagnoses? (2) How can we reduce the huge variance in costs and outcomes in health systems? (3) How do we get people to take more accountability for their own health? (4) How can we provide better health care at lower cost?

On the other hand, digitization and rapid advances in ICT technology are enabling the capture of more data than ever before, including medical health records, people's vital signs and their lifestyle, data about health systems, and data about population health in general. This tsunami of data per se does not immediately result in better healthcare insights, but, on the contrary, if not used properly, it can be a burden to people and result in clinicians spending more time with computers than face to face with patients, or citizens being lost in data they are getting from health trackers and many different sensors, or, again, patients reluctant to accept assistive technologies. This is exactly the point where unlocking the power of data science and artificial intelligence can help by making sense of the large amounts of data, turning them into actionable insights providing mutual benefits to both patient and medical professionals, also helping in answering the abovementioned questions.

## Aim

The goal of this book is to boost the adoption of data science and artificial intelligence solutions for healthcare by raising awareness of existing proof points of these applications and underlying world-class innovations on data science and artificial intelligence in healthcare. The book builds on several interconnected disciplines, including advanced machine learning, big data analytics, data mining, statistics, probabilistic modeling, pattern recognition, computer vision, and seman-

tic reasoning, with direct application to modern HealthTech. Consequently, it shows how the advances in the aforementioned scientific disciplines, as well as digital data platforms, can create value within the healthcare domain and help in reaching the quadruple aim of improving healthcare outcomes, lowering the cost of care, enhancing the patient experience, and improving the work life of care providers.

In particular, the focus of this book is threefold. Firstly, the book aims at demystifying data science and artificial intelligence methods that can be used to extract new knowledge from health data and to improve healthcare delivery. The application of digital technologies for healthcare is seeing a gradual transition to integrated care delivery networks with the consumer at the center. The incoming trends include increased self-management and individualized treatment paths. Thus, secondly, the focus is on applications that enable health systems to deliver better outcomes at lower cost, by boosting the digitization of the healthcare system. This is the starting point for the application of data science and artificial intelligence technologies supporting the move from reactive acute care to pro-active chronic disease management, which is the third focus point of this book. By unlocking the power of big data, connected health systems will be able to deliver personalized and industrialized care models that will lead to a new era of outcome-based healthcare.

## **Organization**

The book starts with three solid tutorial chapters on data science in healthcare, to help readers understand the opportunities and challenges; become familiar with the latest methodological findings in machine learning, in particular deep learning, for healthcare; and help them understand how to use and evaluate the performance of novel data science and artificial intelligence tools and frameworks. These chapters are followed by 11 other chapters showing successful stories on the application of the specific data science technologies in healthcare. The discussed data science technologies and their applications in healthcare focus on, among others, supervised learning, unsupervised learning, deep learning, natural language processing, information retrieval, knowledge management and reasoning, data-to-text, cognitive computation, process mining, smart networking, computational optimization, visual analytics, and robotics.

## **Audience**

This book is primarily intended for data scientists involved in the healthcare domain. There is a clear need for healthcare data analysts to make sense of clinical and personally generated health data more systematically. By reading this book, on one hand computer scientists involved in the medical sector will be able to learn the modern effective data science technologies to create innovation for HealthTech



businesses; on the other, experts involved in the healthcare sector will become more familiar with the advances in ICT and will be able to analyze and process (big) data in order to apply these technologies holistically for patient care. Prior knowledge in data science with real-world applications to the healthcare sector is recommended to interested readers in order to have a clear understanding of this book.

## **Final Words**

We are quite convinced that artificial intelligence and data science will further advance, creating a great potential to industrialize the healthcare sector and to improve the quality of healthcare while managing the costs. In the long run, these technologies might be so impactful that they could result in a giant leap of humanity, changing also the healthcare beyond our current expectations and bringing it closer to maintenance of robotic technology. Let's see which future we will create. Enjoy the reading!

Eindhoven, The Netherlands  
Cagliari, Italy  
Eindhoven, The Netherlands

Sergio Consoli  
Diego Reforgiato Recupero  
Milan Petković

# Contents

## Part I Challenges and Basic Technologies

<b>Data Science in Healthcare: Benefits, Challenges and Opportunities . . . . .</b>	<b>3</b>
Ziawasch Abedjan, Nozha Boujemaa, Stuart Campbell, Patricia Casla, Supriyo Chatterjea, Sergio Consoli, Cristobal Costa-Soria, Paul Czech, Marija Despenic, Chiara Garattini, Dirk Hamelinck, Adrienne Heinrich, Wessel Kraaij, Jacek Kustra, Aizea Lojo, Marga Martin Sanchez, Miguel A. Mayer, Matteo Melideo, Ernestina Menasalvas, Frank Moller Aarestrup, Elvira Narro Artigot, Milan Petković, Diego Reforgiato Recupero, Alejandro Rodriguez Gonzalez, Gisele Roesems Kerremans, Roland Roller, Mario Romao, Stefan Ruping, Felix Sasaki, Wouter Spek, Nenad Stojanovic, Jack Thoms, Andrejs Vasiljevs, Wilfried Verachtert, and Roel Wuyts	
<b>Introduction to Classification Algorithms and Their Performance</b>	
<b>Analysis Using Medical Examples . . . . .</b>	<b>39</b>
Jan Korst, Verus Pronk, Mauro Barbieri, and Sergio Consoli	
<b>The Role of Deep Learning in Improving Healthcare . . . . .</b>	<b>75</b>
Stefan Thaler and Vlado Menkovski	

## Part II Specific Technologies and Applications

<b>Making Effective Use of Healthcare Data Using Data-to-Text Technology . . . . .</b>	<b>119</b>
Steffen Pauws, Albert Gatt, Emiel Krahmer, and Ehud Reiter	
<b>Clinical Natural Language Processing with Deep Learning . . . . .</b>	<b>147</b>
Sadid A. Hasan and Oladimeji Farri	

**Ontology-Based Knowledge Management for Comprehensive Geriatric Assessment and Reminiscence Therapy on Social Robots** ..... 173  
Luigi Asprino, Aldo Gangemi, Andrea Giovanni Nuzzolese, Valentina Presutti, Diego Reforgiato Recupero, and Alessandro Russo

**Assistive Robots for the Elderly: Innovative Tools to Gather Health Relevant Data** ..... 195  
Alessandra Vitanza, Grazia D’Onofrio, Francesco Ricciardi, Daniele Sancarlo, Antonio Greco, and Francesco Giuliani

**Overview of Data Linkage Methods for Integrating Separate Health Data Sources** ..... 217  
Ana Kostadinovska, Muhammad Asim, Daniel Pletea, and Steffen Pauws

**A Flexible Knowledge-Based Architecture for Supporting the Adoption of Healthy Lifestyles with Persuasive Dialogs** ..... 239  
Mauro Dragoni, Tania Bailoni, Rosa Maimone, Michele Marchesoni, and Claudio Eccher

**Visual Analytics for Classifier Construction and Evaluation for Medical Data** ..... 267  
Jacek Kustra and Alexandru Telea

**Data Visualization in Clinical Practice** ..... 289  
Monique Hendriks, Charalampos Xanthopoulos, Pieter Vos, Sergio Consoli, and Jacek Kustra

**Using Process Analytics to Improve Healthcare Processes** ..... 305  
Bart Hompes, Prabhakar Dixit, and Joos Buijs

**A Multi-Scale Computational Approach to Understanding Cancer Metabolism**..... 327  
Angelo Lucia and Peter A. DiMaggio

**Leveraging Financial Analytics for Healthcare Organizations in Value-Based Care Environments** ..... 347  
Dieter Van de Craen, Daniele De Massari, Tobias Wirth, Jason Gwizdala, and Steffen Pauws