Big Data Analytics and Intelligence

This page intentionally left blank

Big Data Analytics and Intelligence: A Perspective For Health Care

EDITED BY

POONAM TANWAR

Manav Rachna International Institute of Research and Studies, India

VISHAL JAIN

Bharati Vidyapeeth's Institute of Computer Applications and Management, India

CHUAN-MING LIU

National Taipei University of Technology (NTUT), Taiwan

VISHAL GOYAL

Punjabi University, India

emerald
PUBLISHING

Reprints and permissions service
Contact: permissions@emeraldinsight.com

Contents

About the Editors

Dr Poonam Tanwar is an Associate Professor, Department of Computer Science & Engineering, Manav Rachna International Institute of Research & Studies (A Deemed to be University) Faridabad, with 17 years of Teaching Experience. Her research areas are artificial intelligence, image processing, knowledge mining, and machine learning. She has filled four patents and one copyright. Beside this she has published 45 research papers in various International Journals and Conferences and published five Book Chapters also. She has been awarded as women researcher award (2019) by VDGOODS Academy, Chennai and best women researcher in Rural Development by GISR Foundation in 2019. She was the Guest Editor for Special issue of "Advancement in Machine Learning (ML) and Knowledge Mining (KM)" for *International Journal of Recent Patents in Engineering* (2018). She received Certificate of RECOGNITION (Silver Category) for Outstanding Contribution to Campus Connect Program for all academic years from 2011 to 2017 by INFOSYS, Chandigarh. She is a Technical Program Committee Member for various international conferences like ICIC 2018, ICFNN, Rome(Italy), European Conference on Natural Language Processing and Information Retrieval, Berlin (Europe), etc.

Dr Vishal Jain is an Associate Professor with Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi, India (affiliated with Guru Gobind Singh Indraprastha University, and accredited by the All India Council for Technical Education). He first joined BVICAM as an Assistant Professor. Before that, he has worked for several years at the Guru Premsukh Memorial College of Engineering, Delhi, India. He has more than 350 research citations indices with Google scholar (h-index score 9 and i-10 index 9). He has authored more than 70 research papers in reputed conferences and journals including Web of Science and Scopus. He has authored and edited more than 10 books with various reputed publishers including Springer, Apple Academic Press, Scrivener, Emerald, and IGI-Global. His research areas include information retrieval, semantic web, ontology engineering, data mining, ad hoc networks, and sensor networks. He has recipient a Young Active Member Award for the year 2012–2013 from the Computer Society of India, Best Faculty Award for the year 2017 and Best Researcher Award for the year 2019 from BVICAM, New Delhi.

Dr Chuan-Ming Liu is a Professor in the Department of Computer Science and Information Engineering (CSIE), National Taipei University of Technology (Taipei Tech), Taiwan, where he was the Department Chair from 2013 to 2017.

Currently, he is pointed to be the Head of the Extension Education Center at the same school. He received his Ph.D. in Computer Science from Purdue University in 2002 and joined the CSIE Department in Taipei Tech in the spring of 2003. In 2010 and 2011, he has held visiting appointments with Auburn University, Auburn, AL, USA, and the Beijing Institute of Technology, Beijing, China. He has services in many journals, conferences, and societies as well as published more than 100 papers in many prestigious journals and international conferences. He was the co-recipients of many conference best paper awards, including ICUFN 2015 Excellent Paper Award, ICS 2016 Outstanding Paper Award, and WOCC 2018 Best Paper Award. His current research interests include big data management and processing, uncertain data management, data science, spatial data processing, data streams, ad hoc and sensor networks, location-based services.

Dr Vishal Goyal, Professor, Department of Computer Science, Punjabi University Patiala, Punjab, with 20 years of Teaching and three years of Software Industry Experience. His research areas are natural language processing, technology development for differently abled people and machine learning. He is an Editor-in-Chief for an International journal – "*Research Cell: An International Journal of Engineering Sciences.*" **He has been awarded one** Young Scientist Award two state awards by Punjab Govt. He has into his credit 10 copyright software and one trademark. He has about 80 research publications in various journals of National and International repute, National and International Conferences. He has delivered about 100 keynote, inaugural addresses, and invited talks at various conferences of national and international repute. He is currently guiding 18 Ph.D. students, 10 students has been awarded degrees. He has guided about 120 M.Tech./M.Phil.(CS) students for their major research projects. He has already completed projects above one crore rupees funded by various ministries. He has into his credit three books published with well-known international publishers. He is the Coordinator of NPTEL Local Chapter of Punjabi University Patiala; Co-Coordinator for "Center for Artificial Intelligence and Data Science"; and the Coordinator for "Research Centre for technology Development for Differently Abled Persons" recently established at his University.

About the Authors

Rashid Ali is an Associate Professor in Computer Engineering Department (ZHCET), Aligarh Muslim University, Aligarh, India. He is a Senior IEEE Member in addition of partaking lifetime affiliation of Computer Society of India. His interests are artificial intelligence, web mining, and soft computing.

Dr Madhulika Bhatia is working as an Associate Professor in the Department of Computer Science and Engineering at Amity School of Engineering and Technology, Amity University, Noida. She holds Diploma in Computer Science & Engineering, BE in Computer Science & Engineering, MBA in Information Technology, MTech in Computer Science, and PhD from Amity University, Noida. She has total 13 years of teaching experience. She published almost 35 research papers in national, international conferences and journals as well as magazine articles. She is also an Author of two books. She filed two Provisional Patents. She attended and organized many workshops, guest lectures, and seminars. She is also the Member of many Technical societies like IET, ACM, and UACEE. She reviewed for Elsevier-Heliyon, IGI, *Indian Journal of Science and Technology*, and completed editorial for Springer Nature, Switzerland Titled "Data Visualization and Knowledge Engineering." She is recently nominated as Brand Ambassador representing India for Bentham Science Publishers (Scopus). Her Innovative Project Startup Idea "Abodsy private limited" get Selected at Startup Event Connecting East & West | Startup Istanbul 2019 among 155.279 startup applications and is selected among Top 100 for a Demo Day at Istanbul inoctober, 2019.

Ms. Gulpreet Kaur Chadha is pursuing her B.Tech in Information Technology from Amity University Uttar Pradesh. Her areas of interest are artificial intelligence, blockchain, Data mining and big data analytics. She has published various research papers in international journals and conferences, and has worked on many technical projects. She is the leader of clubs and committees at college level and a member of clubs like CSI (Computer society of India) and the IOSD (International organization of software developers). She has attended various workshops and seminars, and won many technical competitions. She earned a scholarship for the IUCEE scale student leadership programme as part of the EWB (Engineers without Borders) initiative.

Arindam Chakrabarty is an Assistant Professor at the Department of Management, Rajiv Gandhi University (Central University), India.

Shubham Chaudhary is a Postgraduate Research Scholar at Amity School of Engineering and Technology, Amity University, Noida, India. He attended many workshops and seminars. He has published research papers in conference and Journal.

Uday Sankar Das is a Guest Faculty at the Department of Management & Humanities, National Institute of Technology Arunachal Pradesh, India.

Narayan C. Debnath is currently the Founding Dean of the School of Computing and Information Technology at Eastern International University, Vietnam. He is also serving as the Head of the Department of Software Engineering at Eastern International University, Vietnam. He has been the Director of the International Society for Computers and their Applications (ISCA) since 2014. Formerly, he served as a Full Professor of Computer Science at Winona State University, Minnesota, USA, for 28 years (1989–2017). He was elected as the Chairperson of the Computer Science Department at Winona State University for three consecutive terms and assumed the role of the Chairperson of the Computer Science Department at Winona State University for seven years (2010–2017). He earned a Doctor of Science (DSc) degree in Computer Science and also a PhD degree in Physics. In the past, he served as the elected President for two separate terms, Vice President, and Conference Coordinator of the International Society for Computers and their Applications, and has been a Member of the ISCA Board of Directors since 2001. Before being elected as the Chairperson of the Department of Computer Science in 2010 at Winona State University, he served as the Acting Chairman of the Department. He received numerous Honors and Awards while serving as a Professor of Computer Science during the period 1989–2017. During 1986–1989, he served as an Assistant Professor of the Department of Mathematics and Computer Systems at the University of Wisconsin-River Falls, USA, where he was nominated for the National Science Foundation Presidential Young Investigator Award in 1989. He has made significant contributions in teaching, research, and services across the academic and professional communities. He has made original research contributions in Software Engineering Models, Metrics and Tools, Software Testing, Software Management, and Information Science, Technology and Engineering. He is an Author or Co-author of over 425 publications in numerous refereed journals and conference proceedings in Computer Science, Information Science, Information Technology, System Sciences, Mathematics, and Electrical Engineering. He has made numerous teaching and research presentations at various national and international conferences, industries, and teaching and research institutions in Africa, Asia, Australia, Europe, North America, and South America. He has offered courses and workshops on Software Engineering and Software Testing at universities in Asia, Africa, Middle East, and South America. He has been a Visiting Professor at universities in Argentina, China, India, Sudan, and Taiwan. He has been maintaining an active research and professional collaborations with many universities, faculty, scholars, professionals, and practitioners across the globe. He has been an Active Member of the *ACM*, *IEEE Computer Society*, *Arab Computer Society*, and a Senior Member of the *ISCA*.

B. G. Deepa is an Assistant Professor in School of Computer Science and Applications, REVA University. Holding MCA in Computer Applications from VTU and BSc in Computer Science from Kuvempu University. He is having 9.3 years of teaching experience and four years of research experiance. Presented papers in National level conferences, published technical papers in international journals. Pursuing PhD in REVA University in Data Mining (Medical diagnosis).

Prof. M. Israrul Haque is a Senior Professor at the Department of Business Administration, Aligarh Muslim University, Aligarh. His field of specialization is Management of Human Resources. He has been working on HRM in general and Training in particular. He has edited a book on SHRM and a monograph on Quality in Health Care Sector in India.

Madhurima Hooda is an Assistant Professor at Amity University, Noida. She has strong background in the field of artificial intelligence technologies, computer networks, JAVA programing, video object tracking, software testing, internet of things, and AJAX Applications, etc. She is the Author of various published books in the field of computer networks as well as JAVA programing. She is having various patents to her credit. She has a hands-on experience in developing similar projects using the Arduino UNO, etc.

K. Kalaiselvi has received her MSc, Computer Science degree from Periyar University, MPhil degree from Bharathidasan University, TamilNadu, India, and PhD degree in Computer Science from Anna University, Chennai, TamilNadu, India. She is currently working as Professor and Head in the Department of Computer Science, School of Computing Sciences, and Vels Institute of Science, Technology and Advanced Studies, Chennai, India, which is well-known university. She has more than 16 years of teaching experience in both UG and PG level. Her research interests include knowledge management, data mining, embedded systems, big data analytics, and knowledge mining. She has produced four MPhil scholars. Currently, she is guiding PhD scholars and MPhil scholars in VISTAs. She has published more than 50 research papers in various international journals and presented more than 25 papers in international conferences. She has published a book titled *Protocol to learn C Programming Language* and *Tryambaka {C, C++ & JAVA} Programming Codes*. She has received the Best Researcher Award from DK International Research Foundation, 2018 and received Young Educator and Scholar award – 6th women's day Awards 2019 from the National Foundation for Entrepreneurship Development. She is a Professional Member of Computer Society of India. She serves as Editorial Board Member/Reviewer of various reputed journals. She has been invited as a resource person for various international/national conferences and seminars organized by various institutions. She has completed the mini project worth of Rs. 1,50,000 funded by VISTAS.

Hemanpreet Singh Kalsi is based in the Department of Computer Science & Engineering.

Deepali Kamthania received B.S.c (Hons.) and MCA degrees from Aligarh Muslim University (AMU) Aligarh and Ph. D degree from the Indian Institute of Technology (IIT) Delhi. She has more than 19 years of experience in academics and IT industry. Presently she is working as Professor in the School of Information Technology, VIPS, Delhi. Her areas of research interest include machine learning and hybrid PV systems. She has published over 75 research papers in reputed national and international conferences and journals including SCI, Web of Science and Scopus along with various magazine articles. She has attended and organized many workshops, guest lectures, and seminars. She has delivered many technical talks and chaired sessions at various academic forums. She serves as Editorial board Member/Reviewer of various refereed journals. She is guiding Ph. D scholars and is actively involved in various research projects. She has received Merit scholarship in B. Sc. and Bharat Seva Scholarship in MCA. She is the recipient of Academic Excellence Awards, Best Paper Award in 2019, Best Teacher Award in 2018 and 2019, Active Participation Award (Woman) in 2016-17 from Computer Society of India and Institution of Engineering Technology Awards in 2013 and 2014 for significant contribution to the IT field. Dr. Kamthania is a life time member of IEEE, CSI and ISTE.

Hera Khan is pursuing her Bachelors of Technology in Computer Science and Engineering from DIT University, Dehradun, India. Her areas of interest include data science, machine learning, deep learning, and natural language processing.

Latika Kharb is based at Department of IT, Jagan Institute of Management Studies, Rohini, New Delhi, India.

Dr. Praveen Kumar received his PhD and M.Tech in Computer Science and Engineering. Currently he is working as an Associate professor at Amity University Tashkent, Uzbekistan. He has more than 13+ years of experience in teaching and research. He has been awarded as Best PhD thesis and awarded as Fellow Member of Indian Institute of Machine Learning for outstanding contribution in Artificial Intelligence & Machine Learning, recognized by Govt. of West Bengal, India. His areas of interests include Big data Analytics, Data mining, Machine learning etc. He has to his credit 09 Patents/Copyright and published more then 100+ research paper in International Journals and Conferences (Scopus Indexed). He is guiding 4 PhD Research Scholars in area of Big Data Analytics and Data mining. He has delivered invited talk and guest lecture in Jamia Millia Islamia University, Maharaja Agersen college of Delhi University, Duy Tan University Vietnam, ECE Paris , France etc. He has been associated with many conferences throughout India as TPC member and Session chair etc. He is a lifetime member of IETE, member of ACM, and member of IET (UK) and other renowned technical societies .He is associated with Corporates and he is Technical Adviser of DeetyaSoft Pvt. Ltd. Noida, MyDigital360 etc.

Kumar Saurabh is a Research Scholar at the School of Management Studies, Punjabi University, Patiala, India.

Dr. Amit Kumar Mishra is an Associate Professor in the Department of Computer Science and Engineering at DIT University, Dehradun, India. His areas of expertise include sentiment analysis, machine learning, data mining, and natural language processing.

Shivinder Nijjer is an Assistant Professor at Chitkara Business School, Chitkara University, Punjab, India. She holds a Doctorate in HR Analytics and has authored a book on Predictive HR Analytics, in press for publication by Routledge India, Taylor and Francis Group. She has contributed to international books on *Business Analytics* by Sahil Raj, Cengage Publishers; *Management Information Systems* by Sahil Raj, Pearson Publications; and international bestseller *Management Information Systems* by Laudon and Laudon, Pearson Publications. She has contributed research articles to leading (ABDC ranked and Scopus indexed) journals in the areas of business analytics, social media, academia, Information system design, and so on. She has more than three years of teaching experience and is currently working in other areas of predictive analytics in HRM.

Sahil Raj is an Assistant Professor of Management Information Systems at the School of Management Studies, Punjabi University, Patiala, State of Punjab, India. His main areas of research interest are the application of information systems in business organizations and big data analytics. He has authored four textbooks including *Management Information System and Business Analytics* published by Pearson and Cengage. He is also Adapter and Contributor of global editions of numerous textbooks published by Pearson in Management Information System. He is currently serving AGBA as its Vice-President for Global Publications.

A. Thirumurthi Raja has received his MCA degree from Karunya University, Coimbatore, TamilNadu and PhD degree in computer Science from Bharathiar University, Coimbatore TamilNadu, and India. He is currently working as Assistant Professor in Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies, Chennai, India, which is well-known university. He has more than 10 years of teaching experience in both UG and PG level. His research interests include computer vision, image processing, machine learning, and big data analytics. He produced one MPhil scholar and many PG scholars. Currently, he got guide ship to guide PhD scholars and guided M. Phil. scholars in VISTAS. He has published more than 15 research papers in various International journals and presented papers in international conferences. He has published a book titled "Software Engineering Fundamental Concepts."

Rohit Rastogi received his BE from C.C.S. University, Meerut, 2003. Master's degree in CS from NITTTR-Chandigarh from Punjab University. Currently, he is pursuing his doctoral degree from the Dayalbagh Educational Institute in Agra, India. He is an Associate Professor in the CSE department of ABES Engineering, University, Ghaziabad, India. He has won awards in a variety of areas, including improved education, significant contributions, human value promotion, and long-term service.

He keeps himself engaged in various competition events, activities, webinars, seminars, workshops, projects, and various other educational learning forums.

Dr. Seema Rawat is an Associate Professor, Amity School of Engineering and Technology, Amity University Uttar Pradesh with around 15 years of teaching experience. Her research areas are cloud computing, data mining, artificial intelligence, and big data. She has a lot of research experience and has published more than 100 research papers in international journals and conferences (Scopus indexed) and book chapters, and holds 12 patents. She is the placement head of the Department of Information Technology, Head of the entrepreneurship cell, Amity IEEE student chapter- AICSC and manages the funding and consultancy. She is the members of many international journals and conferences like Elsevier, IGI, Wiley, Inderscience and has reviewed many Technical articles. She is a member of IETE, ACM, WEC (Women entrepreneurship cell), and other renowned technical societies and committees. She is associated with corporates and is a technical advisor at Deetya-Soft Pvt. Ltd. Noida, EnnobleIP Pvt. Ltd., MyDigital360, MSME Technology incubators etc. She received government funding from the Ministry of skill and development for her project .She was awarded the "Faculty Innovation Excellence award 2019" on the World Intellectual day by the secretary of Department of science and technology, Government of India.

S. Senthil, Professor and Director, School of Computer Application, REVA Universty, Bangalore, India, was awarded with doctoral degree by Bharathiar University for his dissertation on Lossless Preprocessing Algorithms for effective text compression. He has completed his BSc (Applied Sciences – Computer Technology) from P.S.G College of Technology, MCA from Bharathidasan University, MPhil in Computer Science from Manonmaniam Sundaranar University, and PhD in Computer Science from Bharathiar University. He has been qualified in State Eligibility Test conducted by Bharathiar University. At present he is guiding eight PhD scholars in the fields of data mining and networks. He has 20 years of experience in teaching. His areas of interest include RDBMS, data mining, data compression, computer networks, and data structures. He has published 30 research papers in various reputed national and international journals. He has presented a paper entitled *Lossless Preprocessing Algorithms for better Compression* in an IEEE International Conference at Zhangjiajie, China. He was also the recipient of the best paper awards, at an International Conference on *Wisdom Based Computing* at Thiruvananthapuram and at a national conference on *Transforming India through Digital Innovations* at Guru Shree Shantivijai Jain College for Women, Chennai.

Tawseef Shaikh is a PhD research scholar in computer engineering department (ZHCET), Aligarh Muslim University, Aligarh, India. His inclinations are artificial intelligence, machine learning, data analytics, and is currently working in biomedical data analytics field. He has published two reputed book chapters: one in Taylor & Francis, and the other in Springer besides publishing in reputed journals.

Prerna Sharma is a web and embedded systems developer and an aspiring IoT Engineer, pursuing her master's in computer application from Jagan Institute of Management Studies(JIMS), New Delhi, India. She was a former research and development intern at Tagglabs Technologies. She also served as the student instructor(IoT) for the academic year (2018-2019) in the skill enhancement division of Vivekananda Institute of Professional Studies(VIPS), New Delhi, India. Her areas of interest are Internet of Things, Embedded Sytems, Microcontroller interfacing, Arduinos, Optimization of embedded systems, and advanced algorithms. She has received the Best Paper Award in an International Conference on Transforming IDEAS (Inter-Disciplinary Exchanges, Analysis, and Search) into Viable Solutions (IDEAS), 2019. Her major publications delineate Optimization & Automation via IoT and scaffolding IoT projects.

Shubham Sharma is pursuing BTech in Computer Science & Engineering with specialization in cloud computing with IBM at Manav Rachna International Institute of Research & Studies, Faridabad. He has participated in many conferences and workshops and published research papers in international conferences. He is placed with TCS, India.

Prateek Singh is based at Department of IT, Jagan Institute of Management Studies, Rohini, New Delhi, India.

Rashbir Singh is based at Department of IT, Amity University, Uttar Pradesh-Noida, India.

Parul Singhal received her BTech degree from AKTU University. Presently, she is MTech. Second year student of Computer Science & Engineering (CSE) in ABESEC, Ghaziabad. She is a Teaching Assistant in the CSE department of ABES Engineering. Ghaziabad, India. She is presently working on data mining and machine learning. She is also working on tension-type headache and stress. She has keen interest in Google surfing. Her hobbies are playing badminton and reading books. She is young, talented, and dynamic.

Ayush Srivastav is pursuing his Bachelors of Technology in Computer Science and Engineering from DIT University, Dehradun, India. His areas of interest include machine learning, artificial intelligence, deep learning, and natural language processing.

Bhawna Suri is based in the Department of Computer Science & Engineering, Bhagwan Parshuram Institute of Technology, India.

Dr Shweta Taneja is based in the Department of Computer Science & Engineering.

Dr. Suryakanthi Tangirala teaches Information Systems in Faculty of Business at University of Botswana. She holds a Ph.D. in computer science from Lingaya's University, India and Masters of Computer Applications from Andhra

University, India. Her current research interests are in areas of Information Systems, Data Science and Big Data Analytics. She has published a number of papers in various refereed journals and conference proceedings in the areas of Machine learning and Big Data analytics.

Bhuvanesh Unhelkar (BE, MDBA, MSc, PhD; FACS; PSM-I, CBAP®) is an accomplished IT professional and Professor of IT at the University of South Florida, Sarasota-Manatee (Lead Faculty). He is also Founding Consultant at MethodScience and a Co-founder/Director at PlatiFi. He has mastery in Business Analysis & Requirements Modeling, Software Engineering, Big Data Strategies, Agile Processes, Mobile Business, and Green IT. His domain experience is banking, financial, insurance, government, and telecommunications. He is a thought-leader and a prolific author of 20 books – including *Big Data Strategies for Agile Business and the Art of Agile Practice* (Taylor and Francis/CRC Press, USA). Recent Cutter executive reports (Boston, USA) include Psychology of Agile, Business Transformation, Collaborative Business & Enterprise Agility and Agile in Practice – a composite approach. He is a winner of the Computerworld Object Developer Award (1995), Consensus IT Professional Award (2006), and IT Writer Award (2010). He has a Doctorate in the area of "Object Orientation" from the University of Technology, Sydney, in 1997. He is the Fellow of the Australian Computer Society, IEEE Senior Member, Professional Scrum Master, Life member of Computer Society of India and Baroda Management Association, Member of SDPS, Past President of Rotary Sarasota Sunrise (Florida) & St. Ives (Sydney), Paul Harris Fellow (+6), Discovery volunteer at NSW parks and wildlife, and a previous TiE Mentor.

Anam is a PhD Research Scholar at the Department of Business Administration, FMSR, Aligarh Muslim University, Aligarh. Her research area is Human Resource Analytics.

Preface

The main focus of this book is to provide the scientific & engineering research of technologies and applications in field of Big Data and to cover the experimental work of chapters concerning the progress towards the Big data technologies, data classification, data analysis and mining, information retrieval, That is emphasizing the problematic issue of knowledge engineering & Management of such systems. The book would be organized as per Background, History, techniques used and various applications of Big data Analytics. Each part will consist of a set of chapters focusing the respective topic outline in order to provide readers an in-depth understanding of concept and technology related to that area. This book includes new developments, Future Perspective and Case studies in the field of Big Data Analytics as well as the overview/ Background knowledge of the Big Data Analytics that the book focuses on

This book can be useful for Under Graduate and Post Graduate students who have taken Big Data Analytics in their Academic Curriculum. This book will be useful to students to develop and deploy their skills to achieve a thorough knowledge in Analytics and in-depth education experience to focus as it pertains to their unique interests. Acknowledging the huge demands of this, Professionals in the industry today, this book has been designed to create trained Computer Science graduates to fulfil the requirements of the industry.

This book comprises of sixteen chapters. Chapter 1 is a general introduction to the applications of Bigdata Analytics in Health care and Challenges. The 2nd Chapter focused on Bigdata Data Analytics Techniques, Tools, and Technologies in Health sector in detail.. Chapter 3 covers the classification Algorithms used in the Healthcare Industry & Training also discussed various Technologies in E-health. In Chapter 4 Morbidity and Hospitalization in India with respect to Bigdata Analytics , Importance of data in Modern Society: Data as Fuel of Modern Economy , Treatment protocol perceptive analytics. & Advance awareness campaigns based on predictive analysis have been discussed. Chapter 5 explained the Predictive Big Data analytics in Healthcare using clinical data, Lab test, Medical and Prescription claims, also presented Challenges of Predictive Big Data Analytics in Healthcare. Chapter 6 is a Big data analytics for Smart Nursery with Health Monitoring system Through Internet of Things. Proposed an IoT and Machine Learning based smart nursery that helps in healthy growing and monitoring of the seed. The system can learn from the internal sensor readings and manage the environment for the best growth of the plant. Chapter 7 is used to discussed the progress of Bigdata Analytics in Health Care., Providing

patient-centric services and how the treatment methods can be improved using Big data analytics & Human genomics/ Genome-Wide Association Studies (GWAS) and Expression Quantitative Trait Loci (EQTLs) have also been discussed. Chapter 8 is how Intrusion Detection and security will be helpful in Healthcare. Chapter 9 is Business Intelligence in the healthcare domain for the growth of any organization. Chapter 10 is case study of facial palsy . Facial palsy is a term used for disruption of facial muscles and could result in temporary or permanent damage of the facial nerve. This chapter explained how big data analytics can be helpful for preventing Facial Palsy. Chapter 11 discussed how Support vector Machine and Naïve Bayes supervised classifiers can be used for selection of best optimized features and prediction of Breast Cancer accuracy & The performance evaluation of the proposed model is estimated by using classification accuracy, confusion matrix, mean, standard deviation, variance and Root mean squared error. The objective of chapter 12 is to present the the Tension-type headaches (TTH) which were associated with diabetes using SPSS, Pearson correlation and ANOVA tests. The purpose of this chapter is to correlate diabetes analysis from TTH and other diseases using the latest technologies to analyze the Internet of Things and Big Data and Stress Correlation (TTH) on human health. In Chapter 13 new Machine Learning approach for meal classification and assessment of nutrients values based on weather conditions along with new and innovative ideas re discussed, Chapter 14 is advancement in the field of Telemedicine and how often the general public are using the services that are provide by the Telehealth and Telemedicine market . Also discussed the models that are being used in today's world, and how these models as implemented and how Telemedicine services are implemented. Chapter 15 illustrate the role of predictive Modeling in health care , as Increasing mobile connectivity and the popularity of wearable devices and advancements in health care IOT devices can help physicians understand the physiological variability in individuals and populations to diagnose and better plan preventive measures.

I hope this book is widely read.

Chapter 1

Big Data Analytics and Intelligence: A Perspective for Health Care

K. Kalaiselvi and A. Thirumurthi Raja

Abstract

Big Data is one of the most promising area where it can be applied to make a change is health care. Healthcare analytics have the potential to reduce the treatment costs, forecast outbreaks of epidemics, avoid preventable diseases, and improve the quality of life. In general, the lifetime of human is increasing along world population, which poses new experiments to today's treatment delivery methods. Health professionals are skillful of gathering enormous volumes of data and look for best approaches to use these numbers. Big data analytics has helped the healthcare area by providing personalized medicine and prescriptive analytics, medical risk interference and predictive analytics, computerized external and internal reporting of patient data, homogeneous medical terms and patient registries, and fragmented point solutions. The data generated level within healthcare systems is significant. This includes electronic health record data, imaging data, patient-generated data, etc. While widespread information in health care is now mostly electronic and fits under the big data as most is unstructured and difficult to use. The use of big data in health care has raised substantial ethical challenges ranging from risks for specific rights, privacy and autonomy, to transparency and trust.

Keywords: Patient predictions; electronic healthcare records; telemedicine; medical imaging; patient engagement; predictive analysis; enhance security

1. Introduction

The concept of enormous data generated from various sources has been accepted and implemented by a lot of information technology-based organizations nowadays. It helps various organizations to understand that capturing and storing all the

data that is being generated within the organization can be beneficial in the future and can get useful insights. Few of the most important advantages that can be gained with the help of using these generated data after analyzing are that it helps in increasing the speed and quality of work in the working environment. Since big data enables the working environment faster and agile it gives the particular organizations a competitive edge and uniqueness from other business organizations.

The implementation of big data analytics in any organizations lets them take a step forward in securing their important data. Hence the data can be used in later stages to identify new opportunities. This helps in building a better organization and takes an appropriate business decisions which will help in attaining more profits and making the customers much happier. The various importance of big data include as follows. Another important feature of big data is that it not only helps in understanding enormous data but also helps in reducing the cost incurred in various ways. By implementing big data technologies and various data manipulation tools it helps in achieving cost advantages. It also ensures that a large amount of collected data can be stored securely in order to ensure privacy protection. A large amount of data can be securely stored and it can be used to identify more efficient ways of doing business. Secondly, it helps in taking faster and better quality decision. With the help of rapid speed analytical tools, various resources and data can be combined to analyze new sources of data. Most of the modern businesses are able to analyze the contents and format of information as and when it is generated and make decisions by analyzing it (Panagiota, Korina, & Sameer, 2019). It also helps in identifying new products and services that can be manufactured to gain more profit and attract more customers. By understanding customer needs and satisfaction through analytics it makes a huge percentage of customers more satisfied.

Since big data and its analytical tools were introduced it has positively helped in healthcare sector in order to save lives more and more. A vast quantity of information collected from various sources over the internet are collected and stored so later on it can be analyzed various analytical tools. The analyzed information can be later on applied to healthcare sector. By using data sets collected from various sources to analyze healthcare situations it has helped to prevent and cure diseases (Ahmed, Fathima-Zahra, & Ayoub, 2018). This also helps the doctors to understand the medical history of the patients. These generated reports can be used to understand if there is any possibility for serious illness. Treatment at an early stage consists of less procedure and can help in reducing the cost incurred by the patients. This also helps the insurance-based business companies or organizations to understand a better picture of patient's medical history in order to give tailored custom insurance packages.

This has also helped the healthcare area by providing personalized medicines and medical risk interface to generate external and internal reports of the particular patient data. The data generated at various levels within healthcare systems are significant. The sources of these data are mainly from electronic health records (EHR), imaging data, patient-generated data, etc. While widespread information in health care is mostly electronic and fits under the big data since most of the information are in unstructured format hence making it time consuming to understand

Fig. 1   Various Applications of Big Data in Health Care.

and make useful information out of it (Uthayasankar, Muhammad, & Zahir, 2017). The uses of big data in various sections of the society have increased at an unexpected rate. This has also increased the various challenges and risk involved. These challenges are faced mainly with regards to rights, privacy, and trust.

The application mainly dependent on the healthcare section takes to help technologies that solve the issues based on computer diagnosis systems. The important task in this section is to upgrade the performance of the system to execute the user required computing. Fig. 1 represents the various applications of big data in healthcare industry. Most widely used areas in which these enormous data can be implemented are mostly EHRs, to improvise security and privacy of the patients, patient predication, medical imaging, patient engagement, and telemedicine. These are the few areas where big data and analytics are used.

2. Big Data Overview

The term big data describes a huge amount of information which can be in any raw format are extracted from various sources in its raw format without making any changes. The users require a computing system that can be powerful enough to organize it and manipulate these data according to the needs of the user. Few examples of these sources are mobile, internet, social media, etc. Later stages of these raw data are used for further processing and can be used to make strategic decisions. With the help of big data and its analytical tools it enables by providing useful decisions that can be taken for future references. It also helps in understand data that were collected decades ago find solutions accordingly. Usually any problems related to data are solved by understanding its scope and impact.

The data collected from big data can be mainly of three types or format. The three types are Volume, Velocity, and Variety. Volume mainly extracts the information from various channels, which includes websites that helps in establishing mass communication and interaction and information generated based on machine-to-machine processing. In the early stages of big data, the main issue was related to storage. Since the advancement in time and technology this burden has been able to be reduced. Velocity refers to the process of collecting data, whereas the third model refers to variety. In this process data can be in various formats of structured, numerical information, and financial transactions. The main difference between big data and data analytics is that big data analytics is the mechanism of collecting large information for a particular task whereas big data is objective for the progression of collecting data that is in raw format and needs further changes to be made to understand and make meaningful information's out of it. The tools that are used for analyzing are referred to as analytical tools.

Few fields where big data mainly works are as follows: The information's stored in various flights are stored in black box. The data generated from this source are huge and mainly stored in its original format as it is. The information in the black box is regarding the communications made within and with the technical staff. Various sites such as Face book and Twitter contain the information and the views posted by people from all around the world. These can be either text, photos, audios, or documents. Another source of big data is from Stock Exchange. The data produced from stock exchanges are stored in servers so that in later stages it can be used for various organizations to understand the market situations. It holds information mainly regarding the price of public shares, financial transactions that take in various business organizations that help the investors to understand how profitable a business is. Power Grid Data is also a source from where data can be extracted. Base station is like a data base storage unit since it consists of information regarding the power grid. Search Engine Data is one of the main sources of big data that are widely used by enormous number of researchers. The data generated from various search engines can be incorporated with existing problems to solve it.

3. Big Data Applications in Health Care

3.1. Various Sources of Data, Methods, and the Challenges Faced

The data retrieved from hospitals and other healthcare industry are difficult to be controlled. It requires a lot of effort and modern techniques to conduct experiments on these data. To conduct experiments, it is important to understand the data, its contents, its source, and format. It is also important to organize it according to various needs. The various difficulties faced can be solved if observational designs are optimized as much as possible to understand the data. The main aim of conducting experiments is to understand the collected data (Etta & Leah, 2019). These data are compared to understand if they are linked and co-related with each other in nature. The process of staffing at different levels of the organization is analyzed to understand the outcome expected from the patients to see if there exists some relation between the data.

Correlational designs are mostly limited from conducting experiments and from determining the relation between outcomes of two levels. Nurse staffing at different levels is the most important factor that help in predicting outcomes for correlational designs. These factors consist of information regarding the environments like nursing care or other services. The data collected from statistical methods can be controlled by various factors that are associated with staffing levels. These factors include the size of the hospital, academic affiliation, or location of the hospital. By carefully selecting the variables and data for processing will help in getting maximum correlation outcome (Etta & Leah, 2019; Uthayasankar et al., 2017).

Reviewing the variables will help in understanding the factors that influences various stages of staffing. The host factors influence elements like important decisions that are to be taken by the organization, quality of nursing care and clinical outcomes.

3.1.1. Levels of Staffing.   Staffing levels are set by administrators of the particular organization and these factors are influenced by various forces such as budgetary considerations and features of local nurse labor markets. The administrative department helps in forming the departmental, work hours, shifts, and other incentives not only to the one level but also to the sub-levels (Antonio, Luis, Maribel, Guilherme, 2019). The practice of the nurse is influenced by the workforce design used in assigning work for a particular project. Few of the other factors that influence the working environment are environment, methods of communication, and the support services available.

There are various variables that contribute toward improving the care and needs of the patient. These factors are inclusive of how serious the patient's health condition is, if any previous medical conditions and family medical history. The health situation of the patient can get worse or better during the stay in the hospital.

- The quality of care provided can results in appropriate execution of assessments and also to improve patient's health situation to get expected outcome and prevent unexpected events. For example, the care provided by the nurses, the examinations done by the doctors to understand the patient's health condition, the medicines or drugs prescribed by the doctors are factors that help in measuring the quality of the care.
- One main factor that is used to measure the quality standards is by giving more importance to safety issues. For example, it is very important to measure the accuracy of medical administration. If the doctors identify the patients' health problems at an early stage it will be much more beneficial and result in rapid improvement of patient's health condition.

3.1.2. Outcomes.   Capturing and analyzing the patient information helps in generating a summarized report so that it can be used in later stages for better understanding. Even though it has resulted in great success still this method is very challenging because it requires a lot of practical understanding and financial considerations. Medical records of patients are used widely in this area as secondary source of data. To understand the outcomes most of the researchers usually use summarized versions of patient records maintained by hospital. These data contain useful healthcare records that explain mainly about how diseases are treated and

also regarding the procedures undergone by the patients before date of discharge. The quality and reliability of these documentations can be different depending on various organizations and the way they maintain it. The form of maintaining electronic medical record helps in keeping information regarding assessment conducted. Analyzing these documents or records will help in improving the performance of the healthcare organization. Wider application of information technology in these types of organizations will be helpful in various ways. This also leads to making users search for data sources that can be trusted to improve performance.

If the healthcare settings are compared accurately it will help in understanding the various risks the patients may face in the future. Eventually these reports are taken for better understanding so that efforts can be increased with the help of risk adjustment methods. The methods of understanding risk are classified into two stages. The first section understands to identify the population of patients that are risk. These can be either categorized by age, gender, or various other related factors. The next stage aims at collecting information that are valid and can be analyzed to understand the population. With the help of risk adjustment, it is easier to find the mechanism for staffing, and also improve the outcome expected. From various studies conducted previously, it helps to understand the importance of staffing and safety outcomes. By providing better quality care by the hospitals it helps in understanding how various serious health issues can be avoided but it also helps in classifying the nursing care. Whereas the positive outcomes cannot be expected if the staffing level is too low. Few factors like psychosocial methods to cure problems and improve care, and the level of self-care capability can be used in later stages for improving the results.

3.1.3. Conclusion. A difference can be seen in healthcare sections where staffing is less when compared to institutions where staffing is more. Most of the researches that were conducted suggest that if nurses appointed are less than required it creates unwanted dangerous situations for both patients and nurses. Therefore, it is necessary to appoint necessary staffs as required to provide quality health care. Studies also suggest that increasing the staffing quantity alone will not be enough to improve the outcome generated.

3.2. Electronic Health Records. Needs and Advantages

3.2.1. Introduction. The most important task of EHR is to help in understanding the medical background of patients with the help of electronic mechanism rather than using traditional techniques of maintaining papers or folders. This helps in reducing time consumed to get the information. It also helps to make the record easily available whenever required to access it. With the introduction of maintaining EHRs has saved money and also helped in accessing it at multiple locations (Cano et al., 2017; Wang & Hajli, 2017). While the receptionist is trying to register the appointment for the patient meanwhile the billing clerk can access the same file using the electronic chart.

By maintaining various pre-defined templates of coding, it can help to easily identify the history or details of the physical exam. A research conducted to understand EHR shows that it has improved the method of maintaining documents. EHRs also help in providing a mechanism for taking decisions and to set alerts.

3.2.2. Importance for Improving Efficiency and Productivity. One of the main aims of maintaining EHR is that it helps to retrieve information's regarding the patients whenever required. Lab results can be gathered from decades ago with less amount of time, thus saving time spend in visiting the labs and also reducing the money spent. It also helps in reducing duplicated tests (Wang & Hajli, 2017). A study conducted shows that using computerized method of entering records have resulted in reduction of duplication.

EHRs are more efficient because it helps in reducing paperwork and have the capability of submitting health claim automatically to the respective insurance organization. EHR supports the doctors and other medical staffs in taking decision. Although EHRs appear to improve overall office productivity but still it may increase the work load in data entry.

3.2.3. Application. The application of EHRs ranges from government sectors to financial sections of various industries. Few of the applications and the expected outcomes from the particular industries are as follows.

The application of EHR's in Government institutions (Sutherland et al., 2016) are considered transforming in the section. It is the goal of the Government to have a dependent and reliable EHR for future references. The introduction of EHR has created an awareness of the potential benefits to help coordinate and improve disease management in older patients. A survey conducted in the year of 2007 showed that with the introduction of EHR has helped in avoiding the malpractices.

3.2.4. Aggregated Data. Since decisions are to be taken based on the past experiences, most of the organizations collect high quality data in raw format. These data are mainly procured from the data collected from inpatient and outpatient data and details regarding the populations at risk. The healthcare data after it is extracted need to be analyzed to gain meaningful information. In near future large healthcare organizations will have to adopt electronic health recording systems and big data analytical tools to organize the data.

3.2.5. Integrated Data. The main disadvantage of maintaining paper-based health records in that it can be used to combine other paper health records and store as the same. Since this mechanism lacks the ability to integrate with other paper forms of information EHRs base mechanisms are introduced. This information can also be used with various other sections of the healthcare sector to take useful decisions. These applications include ability to combine and integrate data.

3.2.6. Conclusion. In order to modernize the infrastructure in healthcare sector it is required to adopt and implement EHRs-based systems. A survey conducted to understand the importance of EHR shows that it helps to identify patients with serious health condition and use various tools that are available to save their life and get a better understanding of the population.

EHR system also helps to decrease the number of office visits. The introduction of EHRs has helped in improving in setting a standard of care to be provided by nurses and understand population health. This has also helped in improving and advancing EHRs. In addition, it has helped in the process of collecting genomic information for future linking to their electronic records.

3.3. Enhancing Patient Engagement

3.3.1. Introduction.   The healthcare industry like any other sector of the society works mainly to gain profit and survive in the business field. Since patients are the most important factor in the healthcare institutions it is important to ensure they are satisfied. Few of the methods to improve the business are by improving the quality of care provided for the patients and the environment of the hospital. This helps in engaging with more and more patience and understanding the rate of staff required take care of the patient's health.

One of the most important ways to increase this engagement rate is by combining the various outcome reports with hospital environment settings. Later stages these data can be used for analyzing the outcomes and increase the care to provide for the patients. The main challenge is to ensure that the privacy of the patients is securely preserved (Patient-Centered Outcomes Research Institute, 2013). This mechanism ensures that while collecting the data and the generated report can only be provide directly to the patient to secure the privacy. Most of these types of reports mainly consist of information regarding the patient's health condition, information describing the symptoms, etc. Finally, the future challenge is to maximize the implementation of these reports in clinical settings and staffing.

A previously conducted research suggests it is the patients that take decisions regarding their health. Hence the systems that is being used in the healthcare institutes needs to be developed in such a way that it provides the required freedom to the patients and enables to take appropriate decisions. At the same time, it is important to ensure that these records can be accessed by the patients whenever required. It also suggests that more steps need to be taken to improve data that are provided.

With the improvement of the new technologies that is available and introducing appropriate changes in what the customers expects will increase the profit day by day. Even though the time consumed for this is increasing when rapid progress and advances are only limited. One of the most widely used mechanism that supports making patients more engaged in healthcare delivery is by using the reports and the outcomes to measure the impact and improvement. These reports and their outcomes are developed individually by using traditional techniques. The reports generated after analyzing the records are used in later stages to take useful measures. This helps in giving a much better insight into how the patient can improve their health. Analyzing these reports it helps to understand the patient's health condition and if there is any improvement. It enables to identify and judge if the person is ready to go back to his daily activities. It also helps in establishing a communication between the patient and the doctor to take appropriate decisions.

3.3.2. Patient-reported Outcomes.   Most of the people who invest in a healthcare sector are mainly interested in improving and expanding the existing business. There are few techniques that are followed to increase the profit. One of the techniques is by collecting patient data and generating report-based outcomes for improving the care provided for patients. It also helps in getting feedback, introducing electronic data collection, and understanding the needs of the population to help in taking decisions. Thus, from this it is clear that with the adoption of patient-reported outcomes it has resulted in a lot of advantages.