

Bioinformatics and Medical Applications

Big Data Using Deep Learning Algorithms

Edited by

A. Suresh

S. Vimal

Y. Harold Robinson

Dhinesh Kumar Ramaswami

and

R. Udendhran



WILEY

This edition first published 2022 by John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA and Scrivener Publishing LLC, 100 Cummings Center, Suite 541J, Beverly, MA 01915, USA

© 2022 Scrivener Publishing LLC

For more information about Scrivener publications please visit www.scrivenerpublishing.com.

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>.

Wiley Global Headquarters

111 River Street, Hoboken, NJ 07030, USA

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

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. 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. 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. 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. 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.

Library of Congress Cataloging-in-Publication Data

ISBN 978-1-119-79183-6

Cover image: Pixabay.Com

Cover design by Russell Richardson

Set in size of 11pt and Minion Pro by Manila Typesetting Company, Makati, Philippines

Printed in the USA

10 9 8 7 6 5 4 3 2 1

Contents

Preface	xv
1 Probabilistic Optimization of Machine Learning Algorithms for Heart Disease Prediction	1
<i>Jaspreet Kaur, Bharti Joshi and Rajashree Shedge</i>	
1.1 Introduction	2
1.1.1 Scope and Motivation	3
1.2 Literature Review	4
1.2.1 Comparative Analysis	5
1.2.2 Survey Analysis	5
1.3 Tools and Techniques	10
1.3.1 Description of Dataset	11
1.3.2 Machine Learning Algorithm	12
1.3.3 Decision Tree	14
1.3.4 Random Forest	15
1.3.5 Naive Bayes Algorithm	16
1.3.6 K Means Algorithm	18
1.3.7 Ensemble Method	18
1.3.7.1 Bagging	19
1.3.7.2 Boosting	19
1.3.7.3 Stacking	19
1.3.7.4 Majority Vote	19
1.4 Proposed Method	20
1.4.1 Experiment and Analysis	20
1.4.2 Method	22
1.5 Conclusion	25
References	26

2	Cancerous Cells Detection in Lung Organs of Human Body: IoT-Based Healthcare 4.0 Approach	29
	<i>Rohit Rastogi, D.K. Chaturvedi, Sheelu Sagar, Neeti Tandon and Mukund Rastogi</i>	
2.1	Introduction	30
2.1.1	Motivation to the Study	30
2.1.1.1	Problem Statements	31
2.1.1.2	Authors' Contributions	31
2.1.1.3	Research Manuscript Organization	31
2.1.1.4	Definitions	32
2.1.2	Computer-Aided Diagnosis System (CADe or CADx)	32
2.1.3	Sensors for the Internet of Things	32
2.1.4	Wireless and Wearable Sensors for Health Informatics	33
2.1.5	Remote Human's Health and Activity Monitoring	33
2.1.6	Decision-Making Systems for Sensor Data	33
2.1.7	Artificial Intelligence and Machine Learning for Health Informatics	34
2.1.8	Health Sensor Data Management	34
2.1.9	Multimodal Data Fusion for Healthcare	35
2.1.10	Heterogeneous Data Fusion and Context-Aware Systems: A Context-Aware Data Fusion Approach for Health-IoT	35
2.2	Literature Review	35
2.3	Proposed Systems	37
2.3.1	Framework or Architecture of the Work	38
2.3.2	Model Steps and Parameters	38
2.3.3	Discussions	39
2.4	Experimental Results and Analysis	39
2.4.1	Tissue Characterization and Risk Stratification	39
2.4.2	Samples of Cancer Data and Analysis	40
2.5	Novelties	42
2.6	Future Scope, Limitations, and Possible Applications	42
2.7	Recommendations and Consideration	43
2.8	Conclusions	43
	References	43
3	Computational Predictors of the Predominant Protein Function: SARS-CoV-2 Case	47
	<i>Carlos Polanco, Manlio F. Márquez and Gilberto Vargas-Alarcón</i>	
3.1	Introduction	48
3.2	Human Coronavirus Types	49

3.3	The SARS-CoV-2 Pandemic Impact	50
3.3.1	RNA Virus vs DNA Virus	51
3.3.2	The <i>Coronaviridae</i> Family	51
3.3.3	The SARS-CoV-2 Structural Proteins	52
3.3.4	Protein Representations	52
3.4	Computational Predictors	53
3.4.1	Supervised Algorithms	53
3.4.2	Non-Supervised Algorithms	54
3.5	Polarity Index Method [®]	54
3.5.1	The PIM [®] Profile	54
3.5.2	Advantages	55
3.5.3	Disadvantages	55
3.5.4	SARS-CoV-2 Recognition Using PIM [®] Profile	55
3.6	Future Implications	59
3.7	Acknowledgments	60
	References	60
4	Deep Learning in Gait Abnormality Detection: Principles and Illustrations	63
	<i>Saikat Chakraborty, Sruti Sambhavi and Anup Nandy</i>	
4.1	Introduction	63
4.2	Background	65
4.2.1	LSTM	65
	4.2.1.1 Vanilla LSTM	65
	4.2.1.2 Bidirectional LSTM	66
4.3	Related Works	67
4.4	Methods	68
	4.4.1 Data Collection and Analysis	68
	4.4.2 Results and Discussion	69
4.5	Conclusion and Future Work	71
4.6	Acknowledgments	71
	References	71
5	Broad Applications of Network Embeddings in Computational Biology, Genomics, Medicine, and Health	73
	<i>Akanksha Jaiswar, Devender Arora, Manisha Malhotra, Abhimati Shukla and Nivedita Rai</i>	
5.1	Introduction	74
5.2	Types of Biological Networks	76
5.3	Methodologies in Network Embedding	76
5.4	Attributed and Non-Attributed Network Embedding	82

5.5	Applications of Network Embedding in Computational Biology	83
5.5.1	Understanding Genomic and Protein Interaction via Network Alignment	83
5.5.2	Pharmacogenomics	84
5.5.2.1	Drug-Target Interaction Prediction	84
5.5.2.2	Drug-Drug Interaction	84
5.5.2.3	Drug-Disease Interaction Prediction	85
5.5.2.4	Analysis of Adverse Drug Reaction	85
5.5.3	Function Prediction	86
5.5.4	Community Detection	86
5.5.5	Network Denoising	87
5.5.6	Analysis of Multi-Omics Data	87
5.6	Limitations of Network Embedding in Biology	87
5.7	Conclusion and Outlook	89
	References	89
6	Heart Disease Classification Using Regional Wall Thickness by Ensemble Classifier	99
	<i>Prakash J., Vinoth Kumar B. and Sandhya R.</i>	
6.1	Introduction	100
6.2	Related Study	101
6.3	Methodology	103
6.3.1	Pre-Processing	103
6.3.2	Region of Interest Extraction	104
6.3.3	Segmentation	105
6.3.4	Feature Extraction	106
6.3.5	Disease Classification	107
6.4	Implementation and Result Analysis	108
6.4.1	Dataset Description	108
6.4.2	Testbed	108
6.4.3	Discussion	108
6.4.3.1	K-Fold Cross-Validation	110
6.4.3.2	Confusion Matrix	110
6.5	Conclusion	115
	References	115
7	Deep Learning for Medical Informatics and Public Health	117
	<i>K. Aditya Shastry, Sanjay H. A., Lakshmi M. and Preetham N.</i>	
7.1	Introduction	118
7.2	Deep Learning Techniques in Medical Informatics and Public Health	121

7.2.1	Autoencoders	122
7.2.2	Recurrent Neural Network	123
7.2.3	Convolutional Neural Network (CNN)	124
7.2.4	Deep Boltzmann Machine	126
7.2.5	Deep Belief Network	127
7.3	Applications of Deep Learning in Medical Informatics and Public Health	128
7.3.1	The Use of DL for Cancer Diagnosis	128
7.3.2	DL in Disease Prediction and Treatment	129
7.3.3	Future Applications	133
7.4	Open Issues Concerning DL in Medical Informatics and Public Health	135
7.5	Conclusion	139
	References	140
8	An Insight Into Human Pose Estimation and Its Applications	147
	<i>Shambhavi Mishra, Janamejaya Channegowda and Kasina Jyothi Swaroop</i>	
8.1	Foundations of Human Pose Estimation	147
8.2	Challenges to Human Pose Estimation	149
8.2.1	Motion Blur	150
8.2.2	Indistinct Background	151
8.2.3	Occlusion or Self-Occlusion	151
8.2.4	Lighting Conditions	151
8.3	Analyzing the Dimensions	152
8.3.1	2D Human Pose Estimation	152
	8.3.1.1 Single-Person Pose Estimation	153
	8.3.1.2 Multi-Person Pose Estimation	153
8.3.2	3D Human Pose Estimation	153
8.4	Standard Datasets for Human Pose Estimation	154
8.4.1	Pascal VOC (Visual Object Classes) Dataset	156
8.4.2	KTH Multi-View Football Dataset I	156
8.4.3	KTH Multi-View Football Dataset II	156
8.4.4	MPII Human Pose Dataset	157
8.4.5	BBC Pose	157
8.4.6	COCO Dataset	157
8.4.7	J-HMDB Dataset	158
8.4.8	Human3.6M Dataset	158
8.4.9	DensePose	158
8.4.10	AMASS Dataset	159
8.5	Deep Learning Revolutionizing Pose Estimation	159

8.5.1	Approaches in 2D Human Pose Estimation	159
8.5.2	Approaches in 3D Human Pose Estimation	163
8.6	Application of Human Pose Estimation in Medical Domains	165
8.7	Conclusion	166
	References	167
9	Brain Tumor Analysis Using Deep Learning: Sensor and IoT-Based Approach for Futuristic Healthcare	171
	<i>Rohit Rastogi, D.K. Chaturvedi, Sheelu Sagar, Neeti Tandon and Akshit Rajan Rastogi</i>	
9.1	Introduction	172
9.1.1	Brain Tumor	172
9.1.2	Big Data Analytics in Health Informatics	172
9.1.3	Machine Learning in Healthcare	173
9.1.4	Sensors for Internet of Things	173
9.1.5	Challenges and Critical Issues of IoT in Healthcare	174
9.1.6	Machine Learning and Artificial Intelligence for Health Informatics	174
9.1.7	Health Sensor Data Management	175
9.1.8	Multimodal Data Fusion for Healthcare	175
9.1.9	Heterogeneous Data Fusion and Context-Aware Systems a Context-Aware Data Fusion Approach for Health-IoT	176
9.1.10	Role of Technology in Addressing the Problem of Integration of Healthcare System	176
9.2	Literature Survey	177
9.3	System Design and Methodology	179
9.3.1	System Design	179
9.3.2	CNN Architecture	180
9.3.3	Block Diagram	181
9.3.4	Algorithm(s)	181
9.3.5	Our Experimental Results, Interpretation, and Discussion	183
9.3.6	Implementation Details	183
9.3.7	Snapshots of Interfaces	184
9.3.8	Performance Evaluation	186
9.3.9	Comparison with Other Algorithms	186
9.4	Novelty in Our Work	186
9.5	Future Scope, Possible Applications, and Limitations	188
9.6	Recommendations and Consideration	188
9.7	Conclusions	188
	References	189

10 Study of Emission From Medicinal Woods to Curb Threats of Pollution and Diseases: Global Healthcare Paradigm Shift in 21st Century	191
<i>Rohit Rastogi, Mamta Saxena, Devendra Kr. Chaturvedi, Sheelu Sagar, Neha Gupta, Harshit Gupta, Akshit Rajan Rastogi, Divya Sharma, Manu Bhardwaj and Pranav Sharma</i>	
10.1 Introduction	192
10.1.1 Scenario of Pollution and Need to Connect with Indian Culture	192
10.1.2 Global Pollution Scenario	192
10.1.3 Indian Crisis on Pollution and Worrying Stats	193
10.1.4 Efforts Made to Curb Pollution World Wide	194
10.1.5 Indian Ancient Vedic Sciences to Curb Pollution and Related Disease	196
10.1.6 The Yajna Science: A Boon to Human Race From Rishi-Muni	196
10.1.7 The Science of Mantra Associated With Yajna and Its Scientific Effects	197
10.1.8 Effect of Different Woods and Cow Dung Used in Yajna	197
10.1.9 Use of Sensors and IoT to Record Experimental Data	198
10.1.10 Analysis and Pattern Recognition by ML and AI	199
10.2 Literature Survey	200
10.3 The Methodology and Protocols Followed	201
10.4 Experimental Setup of an Experiment	202
10.5 Results and Discussions	202
10.5.1 Mango	202
10.5.2 Bargad	203
10.6 Applications of Yagya and Mantra Therapy in Pollution Control and Its Significance	207
10.7 Future Research Perspectives	207
10.8 Novelty of Our Research	208
10.9 Recommendations	208
10.10 Conclusions	209
References	209
11 An Economical Machine Learning Approach for Anomaly Detection in IoT Environment	215
<i>Ambika N.</i>	
11.1 Introduction	215
11.2 Literature Survey	218

11.3	Proposed Work	229
11.4	Analysis of the Work	230
11.5	Conclusion	231
	References	231
12	Indian Science of Yajna and Mantra to Cure Different Diseases: An Analysis Amidst Pandemic With a Simulated Approach	235
	<i>Rohit Rastogi, Mamta Saxena, Devendra Kumar Chaturvedi, Mayank Gupta, Puru Jain, Rishabh Jain, Mohit Jain, Vishal Sharma, Utkarsh Sangam, Parul Singhal and Priyanshi Garg</i>	
12.1	Introduction	236
12.1.1	Different Types of Diseases	236
12.1.1.1	Diabetes (Madhumeha) and Its Types	236
12.1.1.2	TTH and Stress	237
12.1.1.3	Anxiety	237
12.1.1.4	Hypertension	237
12.1.2	Machine Vision	237
12.1.2.1	Medical Images and Analysis	238
12.1.2.2	Machine Learning in Healthcare	238
12.1.2.3	Artificial Intelligence in Healthcare	239
12.1.3	Big Data and Internet of Things (IoT)	239
12.1.4	Machine Learning in Association with Data Science and Analytics	239
12.1.5	Yajna Science	240
12.1.6	Mantra Science	240
12.1.6.1	Positive Impact of Recital of Gayatri Mantra and OM Chanting	241
12.1.6.2	Significance of Mantra on Indian Culture and Mythology	241
12.1.7	Usefulness and Positive Aspect of Yoga Asanas and Pranayama	241
12.1.8	Effects of Yajna and Mantra on Human Health	242
12.1.9	Impact of Yajna in Reducing the Atmospheric Solution	242
12.1.10	Scientific Study on Impact of Yajna on Air Purification	243
12.1.11	Scientific Meaning of Religious and Manglik Signs	244
12.2	Literature Survey	244
12.3	Methodology	246

12.4	Results and Discussion	249
12.5	Interpretations and Analysis	250
12.6	Novelty in Our Work	258
12.7	Recommendations	259
12.8	Future Scope and Possible Applications	260
12.9	Limitations	261
12.10	Conclusions	261
12.11	Acknowledgments	262
	References	262
13	Collection and Analysis of Big Data From Emerging Technologies in Healthcare	269
	<i>Nagashri K., Jayalakshmi D. S. and Geetha J.</i>	
13.1	Introduction	269
13.2	Data Collection	271
13.2.1	Emerging Technologies in Healthcare and Its Applications	271
13.2.1.1	RFID	272
13.2.1.2	WSN	273
13.2.1.3	IoT	274
13.2.2	Issues and Challenges in Data Collection	277
13.2.2.1	Data Quality	277
13.2.2.2	Data Quantity	277
13.2.2.3	Data Access	278
13.2.2.4	Data Provenance	278
13.2.2.5	Security	278
13.2.2.6	Other Challenges	279
13.3	Data Analysis	280
13.3.1	Data Analysis Approaches	280
13.3.1.1	Machine Learning	280
13.3.1.2	Deep Learning	281
13.3.1.3	Natural Language Processing	281
13.3.1.4	High-Performance Computing	281
13.3.1.5	Edge-Fog Computing	282
13.3.1.6	Real-Time Analytics	282
13.3.1.7	End-User Driven Analytics	282
13.3.1.8	Knowledge-Based Analytics	283
13.3.2	Issues and Challenges in Data Analysis	283
13.3.2.1	Multi-Modal Data	283
13.3.2.2	Complex Domain Knowledge	283
13.3.2.3	Highly Competent End-Users	283

	13.3.2.4	Supporting Complex Decisions	283
	13.3.2.5	Privacy	284
	13.3.2.6	Other Challenges	284
13.4		Research Trends	284
13.5		Conclusion	286
		References	286
14		A Complete Overview of Sign Language Recognition and Translation Systems	289
		<i>Kasina Jyothi Swaroop, Janamejaya Channegowda and Shambhavi Mishra</i>	
14.1		Introduction	289
14.2		Sign Language Recognition	290
	14.2.1	Fundamentals of Sign Language Recognition	290
	14.2.2	Requirements for the Sign Language Recognition	292
14.3		Dataset Creation	293
	14.3.1	American Sign Language	293
	14.3.2	German Sign Language	296
	14.3.3	Arabic Sign Language	297
	14.3.4	Indian Sign Language	298
14.4		Hardware Employed for Sign Language Recognition	299
	14.4.1	Glove/Sensor-Based Systems	299
	14.4.2	Microsoft Kinect–Based Systems	300
14.5		Computer Vision–Based Sign Language Recognition and Translation Systems	302
	14.5.1	Image Processing Techniques for Sign Language Recognition	302
	14.5.2	Deep Learning Methods for Sign Language Recognition	304
	14.5.3	Pose Estimation Application to Sign Language Recognition	305
	14.5.4	Temporal Information in Sign Language Recognition and Translation	306
14.6		Sign Language Translation System—A Brief Overview	307
14.7		Conclusion	309
		References	310
Index			315