

Naresh Babu Muppalaneni  
Maode Ma  
Sasikumar Gurumoorthy

# Soft Computing and Medical Bioinformatics

 Springer

Naresh Babu Muppalaneni  
Department of Computer Science  
and Systems Engineering  
Sree Vidyanikethan Engineering College  
(Autonomous)  
Tirupati, Andhra Pradesh  
India

Sasikumar Gurumoorthy  
Department of Computer Science  
and Systems Engineering  
Sree Vidyanikethan Engineering College  
(Autonomous)  
Tirupati, Andhra Pradesh  
India

Maode Ma  
School of Electrical and Electronic  
Engineering  
Nanyang Technological University  
Singapore  
Singapore

ISSN 2191-530X ISSN 2191-5318 (electronic)  
SpringerBriefs in Applied Sciences and Technology  
ISSN 2196-8845 ISSN 2196-8853 (electronic)  
SpringerBriefs in Forensic and Medical Bioinformatics  
ISBN 978-981-13-0058-5 ISBN 978-981-13-0059-2 (eBook)  
<https://doi.org/10.1007/978-981-13-0059-2>

Library of Congress Control Number: 2018939136

© The Author(s) 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.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. part of Springer Nature  
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Contents

<b>1</b>	<b>A Novel Evolutionary Automatic Clustering Technique by Unifying Initial Seed Selection Algorithms into Teaching–Learning-Based Optimization</b>	<b>1</b>
1.1	Introduction	2
1.2	Algorithm Constituents	2
1.3	Automatic Clustering Using TLBO Algorithm (AutoSPSSTLBO)	4
1.3.1	AutoSpssTLBO Algorithm	4
1.4	Results and Discussion	5
1.5	Conclusion	7
	References	8
<b>2</b>	<b>Construction of Breast Cancer-Based Protein–Protein Interaction Network Using Multiple Sources of Datasets</b>	<b>11</b>
2.1	Introduction	11
2.2	Materials and Methods Utilized	12
2.2.1	Selection of Target Protein Associated Datasets	12
2.2.2	Data Concatenation	13
2.2.3	Reconstruction of Protein–Protein Interaction Network	13
2.3	Conclusion	15
	References	15
<b>3</b>	<b>Design and Implementation of Intelligent System to Detect Malicious Facebook Posts Using Support Vector Machine (SVM)</b>	<b>17</b>
3.1	Introduction	17
3.1.1	Face Book Apps	18
3.1.2	Malicious	18

3.1.3	Online Social Network . . . . .	18
3.1.4	Spam . . . . .	18
3.2	Literature Survey . . . . .	18
3.3	Methodology . . . . .	20
3.4	Modules and Discussion . . . . .	21
3.4.1	User . . . . .	21
3.4.2	System Server . . . . .	21
3.4.3	Application Server . . . . .	21
3.4.4	FRAppE . . . . .	21
3.4.5	Similarity of Features . . . . .	22
3.4.6	Tips to Facebook . . . . .	22
3.5	Implementation and Results . . . . .	22
3.5.1	Result . . . . .	22
3.6	Conclusion . . . . .	24
3.7	Future Enhancement . . . . .	24
	References . . . . .	24
<b>4</b>	<b>Classification of Abnormal Blood Vessels in Diabetic Retinopathy Using Neural Network . . . . .</b>	<b>25</b>
4.1	Introduction . . . . .	26
4.2	Literature Reviews . . . . .	26
4.3	Data Description . . . . .	27
4.3.1	Image Acquisition . . . . .	27
4.3.2	Image Annotation . . . . .	27
4.4	Implementation . . . . .	28
4.4.1	Preprocessing . . . . .	28
4.4.2	Neovascularization Region Detection . . . . .	29
4.4.3	Feature Extraction . . . . .	31
4.4.4	Classification . . . . .	31
4.5	Results and Analysis . . . . .	32
4.5.1	Feature Selection . . . . .	32
4.5.2	Neural Network Classification . . . . .	33
4.5.3	Performance in Real-Time Images . . . . .	35
4.6	Conclusions . . . . .	36
	References . . . . .	37
<b>5</b>	<b>Hybridizing Spectral Clustering with Shadow Clustering . . . . .</b>	<b>39</b>
5.1	Introduction . . . . .	40
5.2	Related Work . . . . .	41
5.3	Proposed Methodology . . . . .	41
5.3.1	Hybrid Clustering Architecture . . . . .	43
5.3.2	Algorithm for Hybrid Clustering . . . . .	43
5.4	Conclusion . . . . .	45
	References . . . . .	46

- 6 Dimension Reduction and Storage Optimization Techniques for Distributed and Big Data Cluster Environment . . . . . 47**
  - 6.1 Introduction . . . . . 48
  - 6.2 Dimension Reduction Methods . . . . . 49
    - 6.2.1 MVR, LVF, HCF, and Random Forests . . . . . 49
    - 6.2.2 PCA and Multiclass PCA . . . . . 50
    - 6.2.3 BFE and FFC . . . . . 50
  - 6.3 Data Compression Methods . . . . . 51
  - 6.4 Conclusion . . . . . 53
  - References . . . . . 54
  
- 7 A Comparative Analysis on Ensemble Classifiers for Concept Drifting Data Streams . . . . . 55**
  - 7.1 Introduction . . . . . 55
  - 7.2 Ensemble Classifier on Data Stream . . . . . 56
    - 7.2.1 Online Ensembles . . . . . 57
    - 7.2.2 Block-Based Ensembles . . . . . 57
    - 7.2.3 Hybrid Approaches . . . . . 58
  - 7.3 Results and Discussions . . . . . 59
    - 7.3.1 Comparison of Ensemble Classifiers . . . . . 59
    - 7.3.2 Experimental Results . . . . . 59
  - 7.4 Conclusions . . . . . 61
  - References . . . . . 61
  
- 8 Machine Learning Algorithms with ROC Curve for Predicting and Diagnosing the Heart Disease . . . . . 63**
  - 8.1 Introduction . . . . . 63
    - 8.1.1 Heart Diseases . . . . . 64
    - 8.1.2 Machine Learning (ML) . . . . . 64
    - 8.1.3 Supervised Machine Learning Algorithms . . . . . 65
    - 8.1.4 Unsupervised Machine Learning Algorithms . . . . . 65
    - 8.1.5 Reinforcement Machine Learning Algorithms . . . . . 66
  - 8.2 Materials and Methods . . . . . 66
    - 8.2.1 The Cleveland Dataset . . . . . 66
    - 8.2.2 Random Forests . . . . . 67
    - 8.2.3 Logistic Regression . . . . . 68
    - 8.2.4 Gradient Boosting . . . . . 68
    - 8.2.5 Support Vector Machine . . . . . 69
    - 8.2.6 ROC Curve . . . . . 69
    - 8.2.7 Software Tools . . . . . 70
  - 8.3 Experimental Evaluation . . . . . 70
  - 8.4 Results and Conclusion . . . . . 71
  - 8.5 Future Work . . . . . 72
  - References . . . . . 72

<b>9</b>	<b>Latent Variate Factorial Principal Component Analysis of Microelectrode Recording of Subthalamic Nuclei Neural Signals with Deep Brain Stimulator in Parkinson Disease</b> . . . . .	73
9.1	Introduction . . . . .	74
9.2	Objective . . . . .	75
9.3	Methods . . . . .	75
9.3.1	Latent Variate Factorial (LVF)—PCA . . . . .	75
9.4	Results . . . . .	82
9.5	Conclusions . . . . .	82
	References . . . . .	82
<b>10</b>	<b>CNN Data Mining Algorithm for Detecting Credit Card Fraud</b> . . . . .	85
10.1	Introduction . . . . .	85
10.2	Literature Review . . . . .	88
10.2.1	Data Mining . . . . .	88
10.3	Existing Concept . . . . .	90
10.4	Proposed Concept . . . . .	91
10.5	Conclusion and Future Work . . . . .	92
	References . . . . .	93
<b>11</b>	<b>Study on Computational Intelligence Approaches and Big Data Analytics in Smart Transportation System</b> . . . . .	95
11.1	Introduction . . . . .	95
11.1.1	Operational Big Data . . . . .	97
11.1.2	Analytical Big Data . . . . .	98
11.1.3	Importance of Hadoop . . . . .	98
11.2	Related Work . . . . .	99
11.3	Algorithms . . . . .	99
11.3.1	Swarm Intelligence . . . . .	100
11.3.2	<i>K</i> -Means Clustering . . . . .	100
11.4	Discussion . . . . .	101
11.5	Conclusion . . . . .	101
	References . . . . .	101
<b>12</b>	<b>A Survey on Efficient Data Deduplication in Data Analytics</b> . . . . .	103
12.1	Introduction . . . . .	103
12.2	Deduplication Tool . . . . .	104
12.3	Data Deduplication . . . . .	105
12.4	Deduplication Storage System . . . . .	107
12.5	Encryption in Deduplication . . . . .	107
12.6	Hashing Mechanism for Deduplication . . . . .	110
12.7	Routing . . . . .	110

- 12.8 Data Deduplication: Advantages . . . . . 111
- 12.9 Conclusion . . . . . 112
- References . . . . . 112
- 13 Classification of Alzheimer’s Disease by Using FTD Tree . . . . . 115**
  - 13.1 Introduction . . . . . 115
  - 13.2 Related Work . . . . . 117
  - 13.3 Sample Data . . . . . 117
  - 13.4 Data Mining Methods . . . . . 118
  - 13.5 FTD Tree Algorithm for Classification  
of Alzheimer’s Data . . . . . 119
  - 13.6 Conclusion . . . . . 128
  - References . . . . . 128
- 14 Experimental Investigation of Cognitive Impact of Yoga  
Meditation on Physical and Mental Health Parameters  
Using Electro Encephalogram . . . . . 129**
  - 14.1 Introduction . . . . . 130
  - 14.2 Related Work . . . . . 131
  - 14.3 Proposed Work . . . . . 132
    - 14.3.1 Implementation Methodology . . . . . 133
  - 14.4 Results and Discussion . . . . . 135
    - 14.4.1 Performance Measures . . . . . 136
    - 14.4.2 Statistical Analysis . . . . . 137
  - 14.5 Conclusion . . . . . 138
  - References . . . . . 138