

Lecture Notes on Data Engineering
and Communications Technologies 95

Mohammad Shamsul Arefin

M. Shamim Kaiser

Anirban Bandyopadhyay

Md. Atiqur Rahman Ahad

Kanad Ray *Editors*

Proceedings of the International Conference on Big Data, IoT, and Machine Learning

BIM 2021

Lecture Notes on Data Engineering and Communications Technologies

Volume 95

Series Editor

Fatos Xhafa, Technical University of Catalonia, Barcelona, Spain

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It will publish latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series will have a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

Indexed by SCOPUS, INSPEC, EI Compendex.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at <https://link.springer.com/bookseries/15362>

Mohammad Shamsul Arefin · M. Shamim Kaiser ·
Anirban Bandyopadhyay ·
Md. Atiqur Rahman Ahad · Kanad Ray
Editors

Proceedings of the International Conference on Big Data, IoT, and Machine Learning

BIM 2021

Editors

Mohammad Shamsul Arefin
Chittagong University of Engineering
and Technology (CUET)
Chittagong, Bangladesh

Anirban Bandyopadhyay
National Institute for Materials Science
Tsukuba, Japan

Kanad Ray
Amity University
Jaipur, India

M. Shamim Kaiser
Jahangirnagar University
Dhaka, Bangladesh

Md. Atiqur Rahman Ahad
University of Dhaka
Dhaka, Bangladesh

ISSN 2367-4512

ISSN 2367-4520 (electronic)

Lecture Notes on Data Engineering and Communications Technologies

ISBN 978-981-16-6635-3

ISBN 978-981-16-6636-0 (eBook)

<https://doi.org/10.1007/978-981-16-6636-0>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

This work is subject to copyright. All rights are solely and exclusively licensed 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, expressed 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 Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Organization

General Chairs

V. R. Singh, National Physical Laboratory, New Delhi, India
Yasuhiko Morimoto, Hiroshima University, Japan
Nazmul Siddique, University of Ulster, UK

General Co-chairs

M. Ibrahim Khan, CUET, Bangladesh
K. Hasan Talukder, KU, Bangladesh
Rashed Mustafa, University of Chittagong, Bangladesh

International Advisors

Muhammad H. Rashid, University of West Florida, USA
Manzur Murshed, Federation University Australia
M. Tariqul Islam, Universiti Kebangsaan Malaysia
Tarek M. Sobh, Lawrence Technological University, USA
Fakhrul Alam, Massey University, New Zealand
Issam W. Damaj, Beirut Arab University, Lebanon
Saidur Rahman, Sunway University, Malaysia
Santanu Behera, NIT Rourkela, India
M. Julius Hossain, Research Scientist, EMBL, Heidelberg, Germany
Stefano Vassanelli, University of Padova, Italy
Amir Hussain, Edinburgh Napier University, UK
Ning Zhong, Maebashi Institute of Technology, Japan

Mufti Mahmud, Nottingham Trent University, UK
Yang Yang, Maebashi Institute of Technology, Japan
Wolfgang Maas, Technische Universität Graz, Austria
S. M. Riazul Islam, Sejong University, South Korea

National Advisors

M. Kaykobad, Brac University, Bangladesh
B. Chandra Ghosh, RUET, Bangladesh
S. Prasad Majumder, BUET, Bangladesh
M. Rafiqul Alam, CUET, Bangladesh
R. Alam Seikh, RUET, Bangladesh
M. H. Azad Khan, EWU, Bangladesh
H. M. Hasan Babu, DU, Bangladesh
M. Shorif Uddin, JU, Bangladesh
A. S. M. Latiful Hoque, BUET, Bangladesh
M. Saidur Rahman, BUET, Bangladesh
M. Mahfuzul Islam, BUET, Bangladesh
M. Rafiqul Islam, KU, Bangladesh
M. Shahadat Hossain, CU, Bangladesh
M. Mahbubur Rahman, MIST, Bangladesh
Nasim Akhtar, DUET, Bangladesh
M. Abdur Razzaque, DU, Bangladesh
Kaushik Deb, CUET, Bangladesh
M. Quamruzzaman, CUET, Bangladesh
M. Sohel Rahman, BUET, Bangladesh

Organizing Chair

M. Shamsul Arefin, CUET, Bangladesh

Organizing Co-chairs

M. R. Tanvir Hossain, CUET, Bangladesh
M. Mokammel Haque, CUET, Bangladesh
M. Obaidur Rahman, DUET, Bangladesh

Organizing Secretary

Abdur Rouf, DUET, Bangladesh

Organizing Committee Members

K. M. Rokibul Alam, KUET, Bangladesh
M. Obaidur Rahman, CUET, Bangladesh
M. Kamal Hossain, CUET, Bangladesh
M. Hoq Chowdhury, CUET, Bangladesh
G. M. Atiqur Rahaman, KU, Bangladesh
M. A. Rahman Khan, NSTU, Bangladesh
M. M. Kamal Bhuiya, CUET, Bangladesh
S. Md. Galib, JSTU, Bangladesh
Juel Sikder, RMSTU, Bangladesh
S. M Taohidul Islam, PSTU, Bangladesh
Rezaul Karim, CU, Bangladesh
Md. Ahsan Habib, MVSTU, Bangladesh
Sajeeb Saha, JnU, Bangladesh
Tahmina Khanam, CUET, Bangladesh
Lamia Alam, CUET, Bangladesh
Shayla Sharmin, CUET, Bangladesh
A. Chandra Roy, CUET, Bangladesh
M. Sabir Hossain, CUET, Bangladesh
M. S. Alam Forhad, CUET, Bangladesh
S. Chandra Tista, CUET, Bangladesh
M. Mynul Hasan, CUET, Bangladesh
Annesha Das, CUET, Bangladesh
Ashim Dey, CUET, Bangladesh
Omar Sharif, CUET, Bangladesh
Billal Hossain, CUET, Bangladesh
M. A. Islam Rizvi, CUET, Bangladesh
Sabiha Anan, CUET, Bangladesh
M. Rashadur Rahman, CUET, Bangladesh
M. Fazlur Rahman, CUET, Bangladesh
M. Rashedul Islam, CUET, Bangladesh
M. Kamrul Hossain, CUET, Bangladesh
M. K. I. Zinnah Apu, CUET, Bangladesh
M. Akber Hossain, CUET, Bangladesh
M. Moyeen Uddin, CUET, Bangladesh
K. Hassan Shakib, CUET, Bangladesh
M. Shafiqul Islam, CUET, Bangladesh

M. M. M. Hussain, UCTC, Bangladesh
Nishu Chowdhury, UCTC, Bangladesh

TPC Chairs

K. M. Azharul Hasan, KUET, Bangladesh
M. Al Mamun, RUET, Bangladesh
A. S. M. Kayes, La Trobe University, Australia

TPC Co-chairs

M. Zahidul Islam, Islamic University, Bangladesh
A. Wasif Reza, EWU, Bangladesh
M. A. Akber Dewan, Athabasca University, Canada

Track Chairs

M. Shamim Kaiser, JU, Bangladesh
P. K. Dhar, CUET, Bangladesh
M. S. Chowdhury, University of Chittagong, Bangladesh
M. Chowdhury, CUET, Bangladesh
A. Kumar Bairagi, KU, Bangladesh
W. Rahman Miah, DUET, Bangladesh
Kamruddin Nur, AIUB, Bangladesh
Firoz Mridha, BUBT, Bangladesh
M. J. Alam Patwary, Shenzhen University, China
Syed Galib, JUST, Bangladesh
G. M. Atiqur Rahaman, KU, Bangladesh
S. Islam Khan, IIUC, Bangladesh
A. K. M. Masum, IIUC, Bangladesh

TPC Members

John H. L. Hansen, The University of Texas at Dallas, USA
C. K. Roy, University of Saskatchewan, Canada
Paul Watters, La Trobe University, Australia
Zubair Fadlullah, Lakehead University, Canada

M. M. Hassan, King Saud University, KSA
Alex Ng, La Trobe University, Melbourne, Australia
Indika Kumara, Jheronimus Academy of Data Science, Netherlands
Tarique Anwar, Macquarie University, Australia
Omaru Maruatona, Aiculus Pty Ltd, Melbourne, Australia
A. R. Bin Shahid, Concord University, USA
Shahriar Badsha, University of Nevada, Reno, USA
Kanad Ray, Amity University, Jaipur, India
Abdur Rouf, DUET, Bangladesh
M. Kamal Hossain, CUET, Bangladesh
Mohammad Hammoudeh, Manchester Metropolitan University, UK
Tauhidul Alam, LSU Shreveport University, USA
M. J. Alam Patwary, Shenzhen University, China
P. K. Dhar, CUET, Bangladesh
Nursadul Mamun, The University of Texas at Dallas, USA
M. Ahmed, Edith Cowan University, Perth, Australia
M. Azad Hossain, CUET, Bangladesh
Adnan Anwar, Deakin University, Australia
A. Bandyopadhyay, NIMS, Japan
M. Moshuiul Hoque, CUET, Bangladesh
A. K. M. Masum, IIUC, Bangladesh
M. Hanif Seddiqui, CU, Bangladesh
Asaduzzaman, CUET, Bangladesh
M. Osiur Rahman, CU, Bangladesh
M. S. Chowdhury, CU, Bangladesh
Rashed Mustafa, CU, Bangladesh
Nur Mohammad, CUET, Bangladesh
J. Chakrabartty, CUET, Bangladesh
G. M. Sadiqul Islam, CUET, Bangladesh
N. Karim Chowdhury, CU, Bangladesh
R. Uddin Faruqui, CU, Bangladesh
Ahmed Imteaj, Florida International University, USA
S. M. Rafizul Haque, Canadian Food Inspection Agency, Canada
M. Fazlul Kader, CU, Bangladesh
S. Chandra Banik, CUET, Bangladesh
S. M. Galib, JSTU, Bangladesh
M. H. Chowdhury, City University of Hong Kong, Hong Kong
S. Majumder, Texas A&M University, USA
Kafi Rahman, Truman State University, USA
M. Mizanur Rahman, CUET, Bangladesh
M. Sanaul Rabbi, CUET, Bangladesh
Monjurul Islam, Canberra Institute of Technology, Australia
Rahma Mukta, University of New South Wales, Australia
A. S. Muhammad Sayem, CUET, Bangladesh
M. A. Rahman Ahad, Osaka University, Japan

M. Golam Hafez, CUET, Bangladesh
M. Kamruddin Nur, AIUB, Bangladesh
M. Nurul Huda, UIU, Bangladesh
Aseef Iqbal, CIU, Bangladesh
E. Hoque Prince, York University, Canada
M. Saiful Islam, Griffith University, Australia
M. Firoz Mridha, BUBT, Bangladesh
S. Kibria, SUST, Bangladesh
A. Ahmad, SUST, Bangladesh
S. I. Khan, IIUC, Bangladesh
M. Manirul Islam, AIUB, Bangladesh
A. Das Antar, University of Michigan, USA
I. H. Sarkar, CUET, Bangladesh
M. H. Chowdhury, CUET, Bangladesh
Nilanjan Dey, JIS University, Kolkata, India
A. A. Mamun, JU, Bangladesh
A. K. M. Mahbubur Rahman, IUB, Bangladesh
A. B. M. Aowlad Hossain, KUET, Bangladesh
M. Saiful Islam, CUET, Bangladesh
A. S. M. Sanwar Hosen, JNU, South Korea
Antesar Shabut, Leeds Trinity University, UK
Antony Lam, Mercari Inc., Japan
Banani Roy, University of Saskatchewan, Canada
Atik Mahabub, Concordia University, Canada
Aye Su Phy, Computer University Kalay, Myanmar
Belayat Hossain, Loughborough University, UK
Manohar Das, Oakland University, USA
M. Hoque-Tania, Oxford University, UK
Md. Abu Yousuf, JU, Bangladesh
M. B. Alam Miah, UPM, Malaysia
M. I. Miah, CUET, Bangladesh
M. Sanaul Haque, University of Oulu, Finland
M. Abu Layek, Jagannath University, Bangladesh
M. Ahsan Habib, MBSTU, Bangladesh
M. Golam Rashed, Rajshahi University, Bangladesh
M. Habibur Rahman, MBSTU, Bangladesh
Farah Deeba, DUET, Bangladesh
G. D. Bashar, Boise State University, USA
H. Liu, Wayne State University, USA
Hishato Fukuda, Saitama University, Japan
Imtiaz Mahmud, Kyungpook National University, Korea
Khoo Bee Ee, Universiti Sains Malaysia, Malaysia
Linta Islam, Jagannath University, Bangladesh
Lu Cao, Saitama University, Japan
M. Rashidul Hasan, Rajshahi University, Bangladesh

Ryote Suzuki, Saitama University, Japan

Saiful Azad, Universiti Malaysia Pahang, Malaysia

Sajjad Waheed, MBSTU, Bangladesh

Surapong Uttama, Mae Fah Luang University, Thailand

Tabin Hassan, AIUB, Bangladesh

Tomonori Hashiyama, The University of Electro-Communications, Japan

Wladyslaw Homenda, Warsaw University of Technology, Poland

Tushar Kanti Shaha, JKKNIU, Bangladesh

Preface

We are in the era of the Fourth Industrial Revolution that is a new chapter in human development enabled by extraordinary technological advances and making a fundamental change in the way we live, work, and relate to one another. It is an opportunity to help everyone, including leaders, policy-makers, and people from all income groups and nations, to harness converging technologies in order to create an inclusive, human-centered future. Big data, IoT, and machine learning are three important components of 4.0 Industrial Revolution. To do this, we must prepare our graduates and researchers to conduct their research using Industry 4.0-related technologies such as big data, machine learning, Internet of things, robotics, augmented reality, virtual reality, 3D printing, and so on. As part of our efforts to achieve sustainable development, we must develop and put into effect policies that are focused on the components of 4.0 Industrial Revolution. Considering this fact, we organized the International Conference on Big Data, Internet of Things (IoT) and Machine Learning (BIM 2021) on September 23–25, 2021. Initially, we planned to organize BIM 2021 at Cox’s Bazar, Bangladesh. However, due to the COVID-19 pandemic situation, BIM 2021 took place in full virtual mode. Although we had to arrange BIM 2021 virtually, the research community reacted amazingly well at this challenging time. The support partners of BIM 2021 were CUET Intelligent Computing Lab, IEEE Computer Society Bangladesh Chapter, and the Center for Natural Science and Engineering Research (CNSER).

There were three main tracks at BIM 2021. These are data science and big data, Internet of things, and machine learning. There were a total of 263 submissions from fourteen different countries at BIM 2021. The submitted papers underwent a double-blind review process soliciting expert opinion from at least three experts: at least two independent reviewers and the respective track chair. After the rigorous review reports from the reviewers and the track chairs, the technical committee has selected 59 high-quality papers for presentation in the conference and possible inclusion in Lecture Notes on Data Engineering and Communications Technologies. We hope that the papers published in this volume will help researchers, professionals, and students to enrich their knowledge to continue their research with cutting-edge technologies.

We are thankful to authors who have made a significant contribution to the conference and have developed relevant research and literature in data science, IoT, and machine learning. We would like to express our gratitude to the members of international and national advisory committees, general chairs, general co-chairs, organizing committee members, and technical committee members for their unconditional support to make BIM 2021 a grand success. We are highly grateful to the faculty members of the Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, for their wholehearted support for BIM 2021. We are grateful to Mr. Aninda Bose, Mr. Nareshkumar Mani, and other members of Springer Nature for their continuous support in coordinating this volume publication. Last but not least, we thank all of our volunteers for their tremendous support during this challenging time to make BIM 2021 a successful one.

Chittagong, Bangladesh
Dhaka, Bangladesh
Tsukuba, Japan
Dhaka, Bangladesh
Jaipur, India
July 2021

Mohammad Shamsul Arefin
M. Shamim Kaiser
Anirban Bandyopadhyay
Md. Atiqur Rahman Ahad
Kanad Ray

Contents

Machine Learning for Disease Detection

Performance Analysis of Classifier for Chronic Kidney Disease Prediction Using SVM, DNN and KNN	3
Md. Omaer Faruq Goni, Abdul Matin, Tonmoy Hasan, and Md. Rafidul Islam Sarker	
Comparative Analysis of Machine Learning Techniques in Classification Cervical Cancer Using Isolation Forest with ADASYN	15
Fariha Iffath, Sabrina Jahan Maisha, and Maliha Rashida	
Computer-Aided Cataract Detection Using Random Forest Classifier	27
Tasmina Tasin and Mohammad Ashfak Habib	
COV-Doctor: A Machine Learning Based Scheme for Early Identification of COVID-19 in Patients	39
Ferdib-Al-Islam and Mounita Ghosh	
Ovarian Cancer Prediction from Ovarian Cysts Based on TVUS Using Machine Learning Algorithms	51
Laboni Akter and Nasrin Akhter	
A Comprehensive Analysis of Most Relevant Features Causes Heart Disease Using Machine Learning Algorithms	63
Faria Rahman and Md. Ashiq Mahmood	
Early Stage Detection of Heart Failure Using Machine Learning Techniques	75
Zulfikar Alom, Mohammad Abdul Azim, Zeyar Aung, Matloob Khushi, Josip Car, and Mohammad Ali Moni	

Artificial Intelligence for Imaging Applications

Automatic License Plate Recognition System for Bangladeshi Vehicles Using Deep Neural Network 91

Syed Nahin Hossain, Md. Zahim Hassan, and Md. Masum Al Masba

Vulnerability Analysis and Robust Training with Additive Noise for FGSM Attack on Transfer Learning-Based Brain Tumor Detection from MRI 103

Debashis Gupta and Biprodip Pal

Performance Evaluation of Convolution Neural Network Based Object Detection Model for Bangladeshi Traffic Vehicle Detection 115

S. M. Sadakatul Bari, Rafiul Islam, and Syeda Radiatum Mardia

Hyperspectral Image Classification Using Factor Analysis and Convolutional Neural Networks 129

A. F. M. Minhazur Rahman and Boshir Ahmed

A Convolutional Neural Network Model for Screening COVID-19 Patients Based on CT Scan Images 141

Md. Fazle Rabbi, S. M. Mahedy Hasan, Arifa Islam Champa, Md. Rifat Hossain, and Md. Asif Zaman

Real-time Face Recognition System for Remote Employee Tracking 153

Mohammad Sabik Irbaz, MD Abdullah Al Nasim, and Refat E Ferdous

Data Science and Big Data

Large Scale Image Registration Utilizing Data-Tunneling in the MapReduce Cluster 167

Amit Kumar Mondal, Banani Roy, Chanchal K. Roy, and Kevin A. Schneider

Incorporation of Kernel Support Vector Machine for Effective Prediction of Lysine Formylation from Class Imbalance Samples 181

Md. Sohrawordi and Md. Ali Hossain

Extreme Gradient Boost with CNN: A Deep Learning-Based Approach for Predicting Protein Subcellular Localization 195

Md. Ismail and Md. Nazrul Islam Mondal

Enhancing the Performance of 3D Rotation Perturbation in Privacy Preserving Data Mining Using Correlation Based Feature Selection 205

Mahit Kumar Paul and Md. Rabiul Islam

Developing a Text Mining Framework to Analyze Cricket Match Commentary to Select Best Players	217
Ratul Roy, Md. Rashadur Rahman, M. Shamim Kaiser, and Mohammad Shamsul Arefin	
Indexed Top-k Dominating Queries on Highly Incomplete Data	231
H. M. Abdul Fattah, K. M. Azharul Hasan, and Tatsuo Tsuji	
Development of an Efficient ETL Technique for Data Warehouses	243
Md Badiuzzaman Biplob and Md. Mokammel Haque	
Informatics for Emerging Applications	
Downlink Performance Enhancement of High-Velocity Users in 5G Networks by Configuring Antenna System	259
Mariea Sharaf Anzum, Moontasir Rafique, Md. Asif Ishrak Sarder, Fehima Tajrian, and Abdullah Bin Shams	
Artificial Bee Colony and Genetic Algorithm for Optimization of Non-smooth Economic Load Dispatch with Transmission Loss	271
Mohammad Hanif and Nur Mohammad	
Forecasting Closing Price of Stock Market Using LSTM Network: An Analysis from the Perspective of Dhaka Stock Exchange	289
Md. Mohsin Kabir, Aklima Akter Lima, M. F. Mridha, Md. Abdul Hamid, and Muhammad Mostafa Monowar	
A Dimensionality Reduction Based Efficient Multiple Voice Disease Recognition Scheme Using Mel-Frequency Cepstral Coefficients and K-Nearest Neighbors Algorithm	301
Shovan Bhowmik, Mahedi Hasan, and Muhammad Ataul Hakim	
Dynamic Topology Reconstruction on Next Generation WLAN Using Spatial Reuse Gain by DBSCAN Clustering Algorithm	315
Maleeha Sheikh, Syeda Myesha Mashuda, Redwan Abedin, and Md. Obaidur Rahman	
Aggressive Fault Tolerance in Cloud Computing Using Smart Decision Agent	329
Md. Mostafijur Rahman and Mohammad Abdur Rouf	
Classification of Functional Grasps Using Hybrid CNN/LSTM Network	345
C. Millar, N. Siddique, and E. Kerr	

Internet of Things for Smart Applications

Fire Safety and Supervision System: Fire Hazard Monitoring Based on IoT 367

Ahsan Habib, Srejon Sharma, Mohammad Riduanur Rahman,
Md. Neamul Haque, and Mohammad Ariful Islam Bhuyan

Development of an Optimal Design and Subsequent Fabrication of an Electricity-Generating Ground Platform from Footstep 379

Sudipta Mondal, Md. Tazul Islam, Arnab Das,
and Mayeen Uddin Khandaker

An Automated Planning Approach for Scheduling Air Conditioning Operation Using PDDL+ 391

Amina Shaikh Miah, Fazlul Hasan Siddiqui,
and Md. Waliur Rahman Miah

Nuclear Power Plant Burst Parameters Prediction During a Loss-of-Coolant Accident Using an Artificial Neural Network 407

Priyanti Paul Tumpa, Md. Saiful Islam, Zazilah May,
and Md. Khorshed Alam

IoT Controlled Six Degree Freedom Robotic Arm Model for Repetitive Task 419

Aditi Barua, Tazul Islam, Aidid Alam, and Suvrangshu Barua

An ECC Based Secure Communication Protocol for Resource Constraints IoT Devices in Smart Home 431

Towhidul Islam, Ravina Akter Youki, Bushra Rafia Chowdhury,
and A. S. M. Touhidul Hasan

Internet of Things for Wellbeing

IoT-Based Smart Blind Stick 447

Asraful Islam Apu, Al-Akhir Nayan, Jannatul Ferdous,
and Muhammad Golam Kibria

Deep Learning Techniques in Cyclone Detection with Cyclone Eye Localization Based on Satellite Images 461

Md. Nazmul Haque, A. A. M. Ashfaqul Adel, and Kazi Saeed Alam

Detection of Autism Spectrum Disorder by Discriminant Analysis Algorithm 473

Mirza Muntasir Nishat, Fahim Faisal, Tasnimul Hasan,
Sarker Md. Nasrullah, Afsana Hossain Bristy,
Md. Minhajul Islam Shawon, and Md. Ashraful Hoque

A BRBES to Support Diagnosis of COVID-19 Using Clinical and CT Scan Data	483
S. M. Shafkat Raihan, Raihan Ul Islam, Mohammad Shahadat Hossain, and Karl Andersson	
Performance Analysis of Particle Swarm Optimization and Genetic Algorithm in Energy-Saving Elevator Group Control System	497
Mohammad Hanif and Nur Mohammad	
An Automated and Online-Based Medicine Reminder and Dispenser	513
Shayla Sharmin, Md. Ibrahim Khulil Ullah Ratan, and Ashraful Haque Piash	
Pattern Recognition and Classification	
Power Transformer Fault Diagnosis with Intrinsic Time-Scale Decomposition and XGBoost Classifier	527
Shoaib Meraj Sami and Mohammed Imamul Hassan Bhuiyan	
Human Fall Classification from Indoor Videos Using Modified Transfer Learning Model	539
Arifa Sultana and Kaushik Deb	
Road Sign Detection Using Variants of YOLO and R-CNN: An Analysis from the Perspective of Bangladesh	555
Aklima Akter Lima, Md. Mohsin Kabir, Sujoy Chandra Das, Md. Nahid Hasan, and M. F. Mridha	
Densely-Populated Traffic Detection Using YOLOv5 and Non-maximum Suppression Ensembling	567
Raian Rahman, Zadid Bin Azad, and Md. Bakhtiar Hasan	
Real-time Pothole Detection and Localization Using Convolutional Neural Network	579
Atikur Rahman, Rashed Mustafa, and Mohammad Shahadat Hossain	
Analysis of EEG Signal Classification for Application in SSVEP-Based BCI Using Convolutional Neural Network	593
Md. Saiful Islam Leon, Jarina Akter, Nazmus Sakib, and Md. Kafiul Islam	
Security Detection and Countermeasures	
A Blockchain-Based Approach to Detect Counterfeit Drugs in Medical Supply Chain	609
Shabnam Sabah, A. S. M. Touhidul Hasan, and Apubra Daria	

Enhanced Steganography Technique via Visual Cryptography and Deep Learning 623
Tasfia Seuti, Md. Al Mamun, and A. H. M. Sarowar Sattar

Developing a Framework for Credit Card Fraud Detection 637
Yeasin Arafath, Animesh Chandra Roy, M. Shamim Kaiser, and Mohammad Shamsul Arefin

Automatic Malware Categorization Based on K-Means Clustering Technique 653
Nazifa Mosharrat, Iqbal H. Sarker, Md Musfique Anwar, Muhammad Nazrul Islam, Paul Watters, and Mohammad Hammoudeh

Improved Spam Email Filtering Architecture Using Several Feature Extraction Techniques 665
Priyo Ranjan Kundu Prosun, Kazi Saeed Alam, and Shovan Bhowmik

Detecting Smishing Attacks Using Feature Extraction and Classification Techniques 677
Rubaiath E. Ulfath, Iqbal H. Sarker, Mohammad Javed Morshed Chowdhury, and Mohammad Hammoudeh

InterPlanetary File System-Based Decentralized and Secured Electronic Health Record System Using Lightweight Algorithm 691
Sanjida Sharmin, Iqbal H. Sarker, M. Shamim Kaiser, and Mohammad Shamsul Arefin

Text Mining and Education 4.0

Multi-label Emotion Classification of Tweets Using Machine Learning 705
Simon Islam, Animesh Chandra Roy, Mohammad Shamsul Arefin, and Sonia Afroz

Bangla News Classification Using GloVe Vectorization, LSTM, and CNN 723
Pallab Chowdhury, Ettilla Mohiuddin Eumi, Ovi Sarkar, and Md. Faysal Ahamed

An Ensemble Method-Based Machine Learning Approach Using Text Mining to Identify Semantic Fake News 733
Fahima Hossain, Mohammed Nasir Uddin, and Rajib Kumar Halder

Fuzzy Logic-Based Assessment of Students Learning Outcome in Implementing Outcome-Based Education 745
Abdul Aziz and M. M. A. Hashem

Performance Comparisons in Association Rule Mining Over Public Datasets	761
Jaher Hassan Chowdhury, Md. Billal Hossain, M. Shamim Kaiser, and Mohammad Shamsul Arefin	
Students' Satisfaction with Virtual Interaction Mediated Online Learning: An Empirical Investigation	777
Md. Hafiz Iqbal, Md. Masumur Rahaman, Md. Shakil Mahamud, Serajum Munira, Md. Armanul Haque, Md. Amirul Islam, Md. Abdul Mazid, and Md. Elias Hossain	
Identification of the Resting Position Based on EGG, ECG, Respiration Rate and SpO₂ Using Stacked Ensemble Learning	789
Md. Mohsin Sarker Raihan, Muhammad Muinul Islam, Fariha Fairoz, and Abdullah Bin Shams	
Author Index	799

About the Editors

Professor Dr. Mohammad Shamsul Arefin is affiliated with the Department of Computer Science and Engineering (CSE), Chittagong University of Engineering and Technology, Bangladesh. Earlier he was the Head of the Department. Prof. Arefin received his Doctor of Engineering Degree in Information Engineering from Hiroshima University, Japan with support of the scholarship of MEXT, Japan. As a part of his doctoral research, Dr. Arefin was with IBM Yamato Software Laboratory, Japan. His research includes privacy preserving data publishing and mining, distributed and cloud computing, big data management, multilingual data management, semantic web, object oriented system development and IT for agriculture and environment. Dr. Arefin has more than 110 referred publications in international journals, book series and conference proceedings. He is a senior member of IEEE, Member of ACM, Fellow of IEB and BCS. Dr. Arefin is the Organizing Chair of BIM 2021; TPC Chair, ECCE 2017; Organizing Co-Chair, ECCE 2019; and Organizing Chair, BDML 2020. Dr. Arefin visited Japan, Indonesia, Malaysia, Bhutan, Singapore, South Korea, Egypt, India, Saudi Arabia and China for different professional and social activities.

Dr. M. Shamim Kaiser is currently working as Professor at the Institute of Information Technology of Jahangirnagar University, Savar, Dhaka-1342, Bangladesh. He received his Bachelor's and Master's degrees in Applied Physics Electronics and Communication Engineering from the University of Dhaka, Bangladesh, in 2002 and 2004, respectively, and the Ph.D. degree in Telecommunication Engineering from the Asian Institute of Technology, Thailand, in 2010. His current research interests include data analytics, machine learning, wireless network and signal processing, cognitive radio network, big data and cyber security, renewable energy. He has authored more than 100 papers in different peer-reviewed journals and conferences. He is Associate Editor of the *IEEE Access Journal*, Guest Editor of *Brain Informatics Journal* and *Cognitive Computation Journal*. He is Life Member of Bangladesh Electronic Society; Bangladesh Physical Society. He is also a senior member of IEEE, USA, and IEICE, Japan, and active volunteer of the IEEE Bangladesh Section. He is the founding Chapter Chair of the IEEE Bangladesh Section Computer Society

Chapter. He organized various international conferences such as ICEEICT 2015–2018, IEEE HTC 2017, IEEE ICREST 2018 and BI2020.

Anirban Bandyopadhyay is Senior Scientist in the National Institute for Materials Science (NIMS), Tsukuba, Japan. He received Ph.D. from Indian Association for the Cultivation of Science (IACS), Kolkata, 2005, December, on Supramolecular Electronics. From 2005 to 2007, he was ICYS Research Fellow NIMS, Japan, and, 2007, is now a permanent Scientist in NIMS, Japan. He has ten patents on building artificial organic brain, big data, molecular bot, cancer and Alzheimer drug, fourth circuit element, etc. From 2013 to 2014, he was a visiting scientist in MIT, USA, on biorhythms. He worked in World Technology Network, as WTN Fellow, (2009 continued); he received Hitachi Science and Technology Award 2010, Inamori Foundation Award 2011–2012, Kurata Foundation Award, Inamori Foundation Fellow (2011), Sewa Society International SSS Fellow (2012), Japan; SSI Gold Medal (2017).

Md. Atiqur Rahman Ahad, SMIEEE, SMOSA; Professor, University of Dhaka (DU); Specially Appointed Associate Professor, Osaka University. He studied at the University of Dhaka, University of New South Wales, and Kyushu Institute of Technology. His authored/edited 10 books in Springer, e.g., *IoT-sensor based Activity Recognition; Motion History Images for Action Recognition and Understanding; Computer Vision and Action Recognition*. He published 180+ journal/conference papers, chapters, 120+ keynote/invited talks, 35+ Awards/Recognitions. He is an Editorial Board Member of Scientific Reports, Nature; Associate Editor of *Frontiers in Computer Science*; Editor of *International Journal of Affective Engineering*; Editor-in-Chief: *IJCVSP*; Guest-Editor: *PRL*, Elsevier; *JMUI*, Springer; *JHE*, Hindawi; *IJICIC*; Member: ACM, IAPR.

Kanad Ray (Senior Member, IEEE) received the M.Sc. degree in physics from Calcutta University and the Ph.D. degree in physics from Jadavpur University, West Bengal, India. He has been Professor of Physics and Electronics and Communication and is presently working as Head of the Department of Physics, Amity School of Applied Sciences, Amity University Rajasthan (AUR), Jaipur, India. His current research areas of interest include cognition, communication, electromagnetic field theory, antenna and wave propagation, microwave, computational biology and applied physics. He has been serving as Editor for various Springer book series. He was Associate Editor of the *Journal of Integrative Neuroscience* (The Netherlands: IOS Press). He has been visiting Professor to UTM and UTeM, Malaysia, and visiting Scientist to NIMS, Japan. He has established MOU with UTeM Malaysia, NIMS Japan and University of Montreal, Canada. He has visited several countries such as Netherlands, Turkey, China, Czechoslovakia, Russia, Portugal, Finland, Belgium, South Africa, Japan, Singapore, Thailand and Malaysia for various academic missions. He has organized various conferences such as SoCPROS, SoCTA, ICOEVCI and TCCE as General Chair and Steering Committee Member.

Machine Learning for Disease Detection

Performance Analysis of Classifier for Chronic Kidney Disease Prediction Using SVM, DNN and KNN



Md. Omaer Faruq Goni, Abdul Matin, Tonmoy Hasan,
and Md. Rafidul Islam Sarker

Abstract Disruption of the regular operation of the kidney is named chronic kidney disease (CKD). CKD is widespread, and the death rate due to this increases rapidly. To reduce the amount of death, early detection of CKD is necessary. This paper aims to help medical practitioners to diagnose CKD patients by applying machine learning (ML) techniques. We have applied several ML algorithms to the chronic kidney disease dataset which is archived at the machine learning repository of the University of California Irvine (UCI). The classification approaches have been analyzed in this study including deep neural network(DNN), support vector machine (SVM) and K-nearest Neighbor (KNN). To fulfill this study, the missing values have been imputed with different techniques according to the characteristics of the features and relations among them. Hyperparameters of each algorithm have been tuned through experiments. The proposed approach has been evaluated with the best-tuned parameter. The assessment has done based on different performance metrics such as train–test sensitivity, accuracy, f-measure, specificity and Matthews correlation coefficient (MCC). The empirical result shows that SVM and KNN have enhanced accuracy, and DNN shows the most optimistic result with 100% accuracy compared to the existing.

Keywords Chronic kidney disease · Support vector machine · Deep neural network · K-nearest neighbors

1 Introduction

It is called chronic kidney disease when it fails to perform its regular operation. It is also called kidney failure. It refines our blood by removing wastes and spare fluid as urine. It has some very critical effects such as damage to the nerve and immune system that adversely reduce living standards and the chance of living. The number

Md. O. F. Goni (✉) · A. Matin · T. Hasan · Md. R. I. Sarker

Department of Electrical and Computer Engineering, Rajshahi University of Engineering and Technology (RUET), Rajshahi, Bangladesh

of CKD patients increases due to diabetes, high blood pressure and unsound habit [1]. Gradual failure of the kidney leads to death. The number of CKD patients increases globally [2]. Hence, an early screening system is required.

In the recent world, there is a large number of available data that requires proper usage, otherwise it will be useless. Data mining is the technique to make the appropriate use of large data. Using statistical methods and machine learning (ML) algorithms, it discovers exciting patterns and interactions in data which helps in making useful predictions [3]. In this paper, we use different ML methods that include SVM, KNN and DNN for early diagnosis of CKD in CKD dataset recorded at the machine learning repository of the UCI.

In real-world data, there is a probability of missing and erroneous values that can fail the actual purpose. Adequate data cleaning and missing value imputation techniques are required. It has a crucial impact on model accuracy. CKD dataset is imputed using different techniques. Again, the best feature subset is selected to make the model more efficient. Performance of all the models is assessed with some performance metrics including train–test precision, accuracy, recall, specificity, sensitivity, f -measure and MCC.

This study is structured as follows. In Sect. 2, existing methods on CKD patients classification are discussed. Section 3 represents the detailed explanation of CKD dataset as well as the proposed methods. Section 4 illustrates the result and discussion of this study. Section 5 implies the conclusion.

2 Literature Review

Charleonnann et al. [4] have performed a comparative study of SVM, KNN, decision tree (DT) and logistic regression (LR) classifiers for predicting CKD. SVM has achieved the highest accuracy of 98.3%.

Salekin and Stankovic [5] have analyzed SVM, KNN and ANN with and without imputing the missing values for classification. Feature selection techniques have also been employed. Among them, KNN has achieved the highest prediction accuracy of 0.993 in terms of $F1$ -measure with missing values imputation.

Polat et al. [6] have analyzed the effect of feature selection on the performance of a ML classifier for CKD patients classification. To conduct the analysis, SVM has been examined with different feature subset searching methods and evaluators. Among all the combinations, SVM obtained a great accuracy of 98.5% where the filter method and the best first search have been considered as feature subset evaluator and the feature selector, respectively.

Sara and Kalaiselvi [7] have introduced a hybrid feature selection (FS) technique HWFFS, which combines the warper and the filter method of FS. With and without HWFFS, a performance comparison of NB, SVM and ANN has been conducted. SVM-HWFFS has achieved the highest accuracy of 90%.

Almansour et al. [8] have performed a comparative study of not only the most prominent but also strong classifiers, ANN and SVM in the diagnosis of CKD

patients. All the missing values have been filled with the mean value of associated features. ANN has achieved the highest accuracy of 99.75% while SVM has also attained a great accuracy of 97.75%.

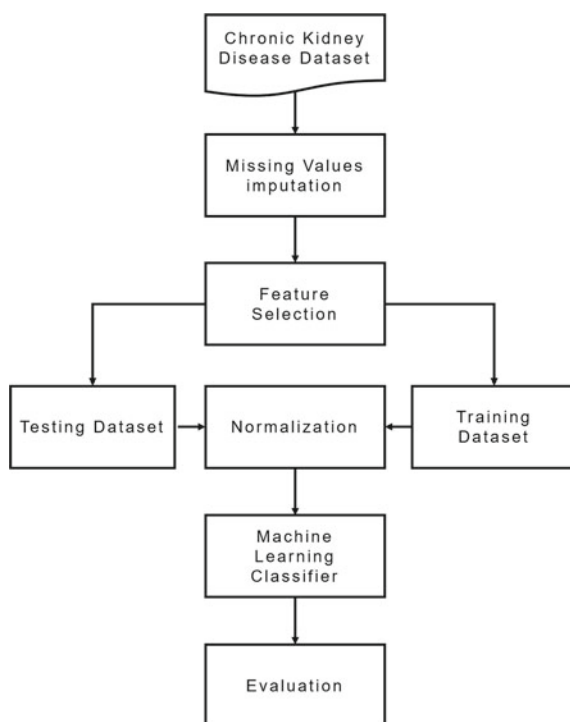
3 Materials and Methods

This study has been conducted into four stages, data imputation, feature subset selection, training & testing of ML model and assessment using performance metrics on CKD dataset recorded at the machine learning repository of the UCI. All the stages are discussed in brief in the next sections. Figure 1 represents the steps of this study.

3.1 Dataset Description

The chronic kidney disease dataset used in this research is openly accessible from the machine learning repository of UCI. It has 400 instances, 24 features, 11 numerical and 13 nominal features. The nominal features are albumin, anemia, appetite,

Fig. 1 Proposed methodology



bacteria, coronary artery disease, diabetes mellitus, hypertension, pus cell, pus cell clumps, pedal edema, red blood cells, specific gravity and sugar. The numerical features are age, blood glucoses, blood pressure, blood urea, hemoglobin, packed cell volume, potassium, red blood cell count, serum creatinine, sodium and white blood cell count. These 400 samples contain 250 (62.5%) CKD patients and 150 (37.5%) non-CKD patients. The dataset has 1012 (10.1%) missing values.

3.2 Dataset Cleaning

It is very important to prepare data before implementing the data for a classification model. CKD dataset has only 160 instances that do not have any missing attributes. That means more than 50% of instances have missing values. There are three things to do with missing values.

- (a) By using a strong and rapid nonlinear classifier that can manage missing values as well as noisy data together. But data cleaning should be such that it can be applied to any model.
- (b) Eliminate the missing values, but there is a problem of losing a large amount of data that may contain important data patterns.
- (c) Imputation of missing value . But due to poor imputation, a classifier may be biased to the imputed value. If missing values are imputed properly, it can overcome all the problems.

Missing values can be imputed in different ways.

Imputation using mean value

All the missing values can be imputed with the mean value of the corresponding feature. In this experiment, age, Bp, Sg, Bgr, Bu, Sc, sod and pot are imputed using this method.

Imputation using maximum occurrence

This method can be used for missing nominal data types. In this study, this technique is used for all the missing data of the nominal type using the most frequently occurred value of the corresponding feature.

Imputation using regression model

To apply this technique, it is required to study the interaction between features and find relation among the features if there is any. Then fit a regression model using the related features and predict the missing values. In Fig. 2, it shows the interaction of bu, hemo and pcv with rc. We can fit a regression model to impute the missing values.

In this study, at first, some of the missing values of rc are imputed with the predicted value of the linear regression model fitted by hemo and rc. Then the second part of

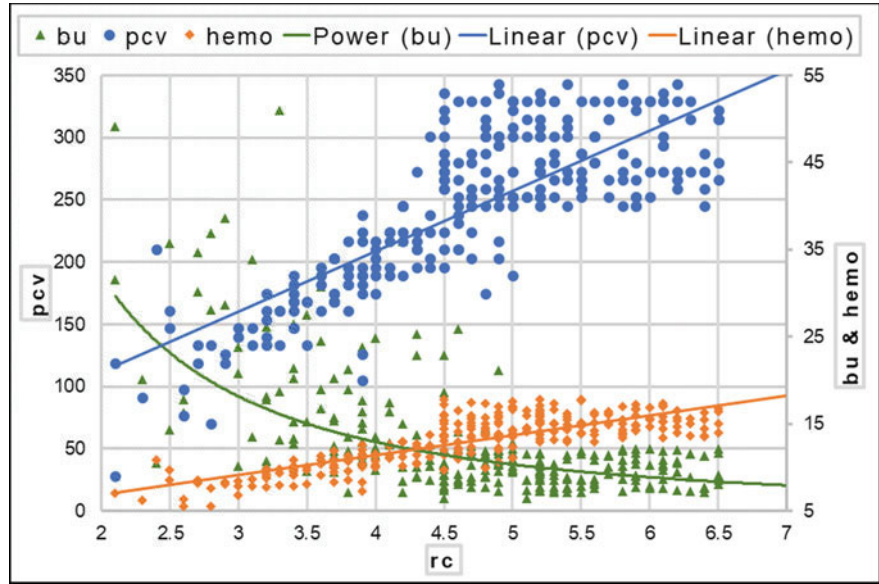


Fig. 2 Interaction of bu, hemo and pcv with rc

the missing values is imputed in the same way, and the model is fitted by rc and pcv. The rest of the missing values of rc are fitted with the mean of rc. And finally, missing values of pcv, hemo and bu are imputed using a regression model fitted using rc and corresponding features.

3.3 Feature Selection

All the dataset features do not have enough impact on decision making. It is better to remove less important and irrelevant features, and this technique is called feature selection [9]. The efficiency of a classifier along with its effectiveness is greatly impacted by feature selection. It reduces the time consumption of a model and makes it faster and efficient [10].

In this study, to select the best features, different feature subsets are created according to the feature importance and tested with the models. The best feature subset consists of 11 features. Feature importance can be calculated in many ways. In this paper, a random forest model is used to do this job, and the best feature subset is selected using the feature importance shown in Fig. 3.

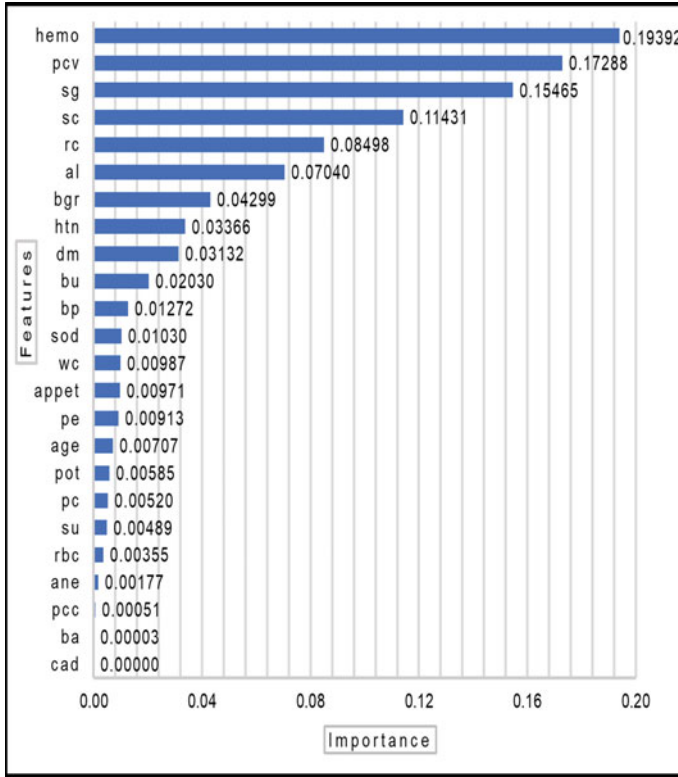


Fig. 3 Feature importance

3.4 Normalization

As part of data preparation for machine learning, technique of normalization is used sometimes to adjust the numeric column values to a standard scale in the dataset, without altering the differences in the value ranges [11]. Each dataset does not need normalization for machine learning. It is applied only when dataset features have different ranges. There are several approaches to normalization that include z -score, logistic, min–max, tanh, lognormal, etc. To conduct this study, min–max normalization has been used. Mathematically, it can be expressed as follows:

$$\text{reshaped: } Z = \frac{Z - \min(Z)}{\max(Z) - \min(Z)} \quad (1)$$

where Z is the column required to normalize.