All India Seminar on Machine Learning and Soft Computing Applications in Engineering and Science (MLSC-ES)

April 6–7, 2024

Organized

By

The Institution of Engineers (India)

Durgapur Local Centre

Under the Aegis of

COMPUTER ENGINEERING DIVISION BOARD

Seminar Proceedings

Prepared By

Technical Committee

MLSC-ES'24

All India Seminar on Machine Learning and Soft Computing Applications in Engineering and Science (MLSC-ES)

April 6–7, 2024

Organized
By

The Institution of Engineers (India)

Durgapur Local Centre

Under the Aegis of

COMPUTER ENGINEERING DIVISION BOARD

Seminar Proceedings

Prepared By

Technical Committee

MLSC-ES'24

Copyright Statement:

The Institution of Engineers (India)
Durgapur Local Centre
Nehru Avenue, B-Zone, Durgapur, West Bengal 713205

Preface

Welcome to the proceedings of the All India Seminar on Machine Learning and Soft Computing Applications in Engineering and Science (MLSC-ES 2024), focused on the growing intersection of machine learning and soft computing applications in engineering and science. In recent years, the rapid advancement of computational techniques has revolutionized our approach to problem-solving across various domains. Machine learning algorithms, in particular, have emerged as powerful tools for extracting insights from complex datasets, while soft computing techniques offer flexible frameworks for handling uncertainty and imprecision in real-world scenarios.

This seminar brings together researchers, practitioners, and experts from academia and industry to exchange ideas, share insights, and explore the latest developments in this exciting field. The papers included in this seminar proceedings represent a diverse array of applications spanning engineering and science disciplines, including but not limited to medical imaging, image analysis, control systems, optimization, power engineering, civil and structural engineering, chemical engineering, biotechnology, railway engineering and safety, public healthcare, data analysis, etc.

As editors of this volume, we are thrilled to present a collection of cutting-edge research contributions that showcase the transformative potential of machine learning and soft computing techniques in addressing real-world challenges. We believe that the insights presented herein will not only advance the state-of-the-art in these fields but also inspire further innovation and collaboration in the pursuit of novel solutions to complex problems.

We extend our sincere gratitude to all the authors for their valuable contributions, as well as to the reviewers for their diligent efforts in ensuring the quality and rigor of the manuscripts. We also express our appreciation to the seminar organizers, sponsors, and attendees for their support and participation, without which this event would not have been possible.

We hope that the papers contained in these proceedings will serve as a valuable resource for researchers, practitioners, students, and enthusiasts alike, fostering continued exploration and advancement at the exciting nexus of machine learning, soft computing, engineering, and science.

Convener

MLSC-ES 2024

The Institution of Engineers (India)

AN ISO 9001 : 2015 CERTIFIED ORGANISATION
(ESTABLISHED 1920, INCORPORATED BY ROYAL CHARTER 1935)

Dr G Ranganath, FIE
President

Headquarters : 8 Gokhale Road, Kolkata 70002
Contact : 91- 33-40106202 / 40106204
E-mail : president@ieindia.org

A Century of Service to the Nation

MESSAGE

It is my pleasure to note that Durgapur Local Centre of The Institution of Engineers (India) is organising the All India Seminar on the theme "Machine Learning and Soft Computing Applications in Engineering and Science (MLSC-ES) " on 06-07 April 2024 under the aegis of Computer Engineering Division of the Institution in Durgapur.

Machine Learning and Soft Computing applications in Engineering and Science (MLSC-ES) encompass the development of algorithms that enable computers to learn from data, enhancing performance across diverse domains. Techniques such as fuzzy logic, neural networks, evolutionary computing, and probabilistic reasoning play pivotal roles in advancing these applications, driving innovation and solving complex problems in fields like robotics, healthcare, and environmental science. For policy makers and stakeholders, it is crucial

I am sure that the speakers and experts will provide newer insights on the topic. The deliberations at the Seminar would surely explore and recommend to prioritise investments by the policy makers and stakeholders in research and development, foster collaborations between academia and industry, and promote the integration of MLSC-ES in educational curricula to harness its full potential for societal benefit.

I convey my best wishes to the organisers and wish the Seminar a grand success.

(Dr G Ranganath)

CSIR-CMERI, Durgapur

Message

I am pleased to learn that the Institution of Engineers (India), Durgapur Local Centre is hosting an All India Seminar on "*Machine Learning and Soft Computing Applications in Engineering and Science (MLSC-ES)"* on April 6-7, 2024.

The world of Artificial Intelligence (AI) and Machine Learning (ML) is undergoing a revolution that is transforming industries and reshaping our existence. AI and ML are not just buzzwords, they are catalysts for unprecedented change across all fields.

I extend my best wishes to the organizers and participants of the Seminar and hope for the event's success.

(Naresh Chandra Murmu)

राष्ट्रीय प्रौद्योगिकी संस्थान दुर्गापुर
(शिक्षा मंत्रालय, भारत सरकार के अधीन राष्ट्रीय महत्व का संस्थान)
NATIONAL INSTITUTE OF TECHNOLOGY DURGAPUR
(An Institute of National Importance under Ministry of Education, Govt. of India)

प्रो. अरविन्द चौबे / Prof. Arvind Choubey, निदेशक / Director

Ref. No. : NITD/DS/2024/ //

Date : March 22, 2024

MESSAGE

It is a great pleasure and excitement to know that The Institution of Engineers (India) Durgapur Local Centre is going to organize an All India Seminar on Machine Learning and Soft Computing Applications in Engineering and Science (MLSC-ES'24) on 6th and 7th April, 2024 and as part of the seminar a souvenir is being published on this occasion.

Machine learning and soft computing have revolutionized the way we approach complex problems in engineering and science. Through ingenious algorithms, novel methodologies, and interdisciplinary collaboration, we continue to witness remarkable breakthroughs that surpass traditional boundaries. This seminar serves as a testimony to our commitment to harnessing the power of these technologies to address the most pressing challenges of our time.

I wish to express my sincere gratitude to the organizing committee, keynote and invited speakers, sponsors, and volunteers for their unwavering dedication and tireless efforts in making this event possible.

I am confident that this seminar will not only foster meaningful connections and collaborations but also inspire new avenues of research. Let us seize this opportunity to push the boundaries of knowledge and innovation, and together, pave the way for a future where machine learning and soft computing continue to drive transformative change in engineering and science.

(Arvind Choubey) 22/3/24

E-mail : director@nitdgp.ac.in, Website : www.nitdgp.ac.in
महात्मा गांधी एवेन्यु, दुर्गापुर- 713209, भारत / Mahatma Gandhi Avenue, Durgapur - 713209, India
टेल/Tel. No.- +91-343 254 6397, फैक्स/Fax No.- +91-343 254 7375

The Institution of Engineers (India)

[Established 1920, Incorporated by Royal Charter 1935]

DURGAPUR LOCAL CENTRE

"A Century of Service to the Nation"

Nehru Avenue, B-Zone, Durgapur - 713205

9475912594 • E-mail : durgapurlc@ieindia.org •Web: www.ieidurgapurlc.org

Prof (Dr) Chandan Koner, FIE
Chairman
Mob. : 9382425362 / 9434535556

Dr. Bimal Das, MIE
Honorary Secretary
Mob.: 9434789023

Date: 01/04/2024

MESSAGE

I extend my warmest wishes for a successful and inspiring All India Seminar on "Machine Learning and Soft Computing Applications in Engineering and Science (MLSC-ES)" during April 6–7, 2024 organized by The Institution of Engineers (India), Durgapur Local Centre under the aegis of Computer Engineering Division Board.

The conference will serve as a platform for the exchange of knowledge, ideas, and innovations in the vital fields of machine learning and soft computing in Engineering and Science.

May conference be filled with fruitful discussions, enlightening presentations, and meaningful connections. May participates leave this conference with renewed inspiration, strengthened partnerships, and actionable insights in machine learning and soft computing.

I applaud the efforts of the organizers and participants who have contributed to making this event possible. Your commitment to advancing environmental stewardship is truly commendable and serves as a beacon of hope for generations to come.

Wishing you all a memorable and impactful conference experience.

Warm regards

Dr. Bimal Das
Honorary Secretary
IEI Durgapur Local Centre

ALL INDIA SEMINAR ON 'MACHINE LEARNING AND SOFT COMPUTING APPLICATIONS IN ENGINEERING AND SCIENCE (MLSC-ES)'

Organized by

The Institution of Engineers (India), Durgapur Local Centre

April 6-7, 2024

NATIONAL ADVISORY COMMITTEE

Chairman : Dr. G. Ranganath, FIE, President, IEI
Co-Chairman : Dr. Amit Kumar Das, FIE, Chairman, CPDB, IEI

Members:

Prof (Dr.) Chandrashekara S. N. FIE, Member, IEI
Prof (Dr.) Deepu R. FIE, Member, IEI
Dr. Labh Singh Bhari FIE, Member, IEI
Prof (Dr.) Ravimaran S. FIE, Member, IEI
Prof (Dr.) Sanjay Kumar FIE, Member, IEI
Dr. Thangaraj K. FIE, Member, IEI
Prof. Phalguni Gupta, GLA University
Prof. Jamuna Kanta Sing, Jadavpur University
Prof. Arnab Bhattacharyya, IIT Kanpur
Prof. Kartik Chandra Ghanta, NIT Durgapur
Prof. Niladri Chatterjee, IIT Delhi
Prof. Pabitra Mitra, IIT Kharagpur
Prof. Amitava Chatterjee, Jadavpur University
Prof. Rajesh M. Hegde, IIT Kanpur
Prof. Tanmay De, NIT Durgapur

ORGANIZING COMMITTEE

Org. Chairman : Prof. (Dr.) Chandan Koner, FIE, Chairman, IEI, Durgapur Local Centre
Org. Secretary : Dr. Bimal Das, MIE, Hony. Secretary, IEI, Durgapur Local Centre
Convener : Dr. Dakshina Ranjan Kisku, FIE, Executive Committee Member, Computer Engg. Division, IEI, Durgapur Local Centre

Members:

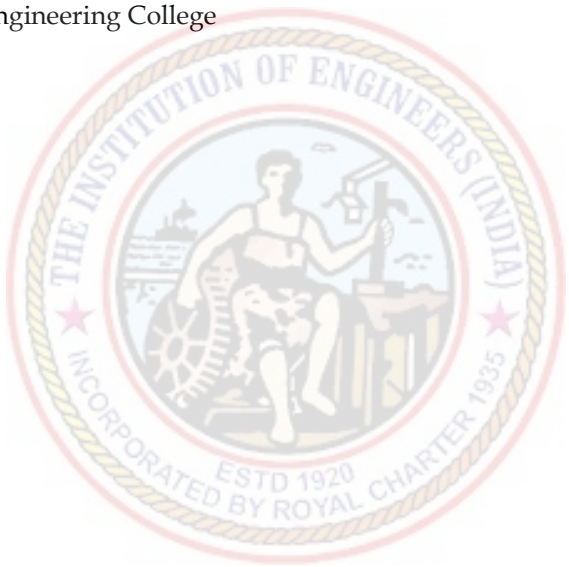
Dr. Tushar Kanti Bera, FIE
Mr. Safikul Islam, FIE
Dr. Rajib Ghosh Chaudhuri, AMIE
Mr. Debi Prasad Das, MIE
Mr. Koushik Chatterjee, AMIE
Dr. Subrata Kumar Mandal, AMIE
Mr. Uttam Kumar Biswas, AMIE
Mr. Sudipta Sikdar, AMIE
Mr. Nozrul Hussain Mozumder, FIE
Mr. Manik Nath Bandyopadhyay, FIE
Mr. Parasuram Shaw, FIE
Mr. Debashis Guha, FIE
Dr. Arup Kumar Mandal, MIE
Mr. Manas Majumder, FIE

TECHNICAL COMMITTEE

Chairman : Prof. Monish Chatterjee, Asansol Engineering College

Members:

Dr. Kamlesh Tiwari, BITS Pilani
Dr. Aditya Nigam, IIT Mandi
Prof. Nibaran Das, Jadavpur University
Dr. D. N. Ray, CSIR-CMERI Durgapur
Dr. Nanda Dulal Jana, NIT Durgapur
Dr. Parag Kumar Guhathakurta, NIT Durgapur
Prof .Ashwani Kush, Kurukshetra University
Dr. Sandesh Gupta, Chhatrapati Shahu Ji Maharaj University
Er. Anand Kumar, SAIL-DSP
Er. Rakesh Ojha, SAIL-DSP
Dr. Supriya Bera, NIT Durgapur
Dr. Abhijit Sanyal, Novartis Healthcare Pvt. Ltd.
Dr. Sujit Kumar Mandal, NIT Durgapur
Dr. Debasish Sarkar, Asansol Engineering College
Dr. Mamata Dalui, NIT Durgapur
Dr. Sujan Kumar Saha, NIT Durgapur
Dr. Sandip Halder, Asansol Engineering College



About the Institution of Engineers (India)

The Institution of Engineers (India) or IEI is the largest multidisciplinary professional body that encompasses 15 engineering disciplines and gives engineers a global platform for sharing professional interest. IEI has membership strength of above 0.8 million. Established in 1920, with its headquarter at 8 Gokhale Road, Kolkata-700020, IEI has served the engineering fraternity close to a century. During this period of time, IEI has been inextricably linked with the history of modern-day engineering. In 1935, IEI was incorporated by Royal Charter and remains the only professional body in India which has been accorded this honour. Today, its quest for professional excellence has given it a place of pride in almost every prestigious and relevant organization across the globe. It provides a vast array of technical, professional and supporting services to the Government, Industries, Academia and the Engineering fraternity, operating through its 125 Centres located across the country and 6 overseas chapters. Besides, IEI has bilateral agreements with about 31 international bodies and membership of another 8 international bodies of the developed nations across the globe. Being recognized as a Scientific and Research Organization (SIRO) by the Department of Scientific and Industrial Research, Ministry of Science and Technology, Government of India, IEI promotes the cause of research and development by providing Grant-in-Aid support to undergraduate, postgraduate students and Ph.D. Research Scholars of Engineering Institutions and Universities. IEI conducts Section A & B Examinations in various engineering disciplines (popularly known as AMIE examination), the successful completion of which is recognized as equivalent to a Degree in appropriate field of Engineering of recognized Universities of India by the Ministry of Human Resources Development, Government of India. Every year, a large number of candidates appear for these exams. IEI in collaboration with Springer regularly publishes a peer-reviewed international journal in five series, namely, Series A, Series B, Series C, Series D and Series E covering fifteen engineering disciplines.

About IEI Durgapur Local Centre

The Institution of Engineers (India), Durgapur Local Centre established in 1961 in National Institute of Technology, (formerly known as R. E. College) Durgapur. In 1986 it was shifted from NIT Durgapur to Nehru Avenue, B-Zone, Durgapur - 713205. Durgapur Local Centre achieved Best Local Centre for the session 2003-2004, 2015-16 and again 2018-2019. Durgapur chapter also achieved 2nd best technicians' chapter for the sessions 2015-2016, 2016-2017, 2017-2018, 2018-2019 consecutively. Present Corporate Membership Strength is 2867 and Non-corporate (i.e. MT, AT, SM, Technician and Sr Technicians') Strength is 12094 as on 30th September 2023.

About Durgapur City

Durgapur is located in the district of Paschim Burdwan, in the state of West Bengal, India. It is a well-planned industrial city on the banks of river Damodar. It was founded in the late fifties by the then Chief Minister of West Bengal Dr. Bidhan Chandra Roy. It is about 175 Kms away from the international Airport Kolkata, capital city of West Bengal. A large number of manufacturing industries are there in Durgapur, which includes Durgapur Steel Plant (DSP), Alloy Steel Plant (ASP), General Electric Manufacturing Unit, Graphite India Ltd, Durgapur Project Ltd., Durgapur Thermal Power Corporation, Philips Carbon Black Ltd, Graphite India Ltd. It is surrounded by many Coal Mines also. It is also having a software technology park of India consisting of various software firms like Pinnacle, Webel etc. It has an ambient environment conducive for studies. The well-built, well lit roads add to the strong infrastructure of Durgapur. At the present day, Durgapur city is evolving as one of the major attractions (after Kolkata) in West Bengal, not only because of its excellent infrastructure, but availability of all modern entertainment and health avenues, like shopping malls, multiplexes, five star hotels, specialty hospitals, etc. Well connectivity with Road, Rail & Air. The connectivity with Kazi Nazrul Islam Airport (Andal) to Delhi, Mumbai, Hyderabad, Chennai, Bangalore etc. is very good.

Meet Our Chief Guest

Dr. Naresh Chandra Murmu
Director, CSRI-CMERI Durgapur

Dr. Naresh Chandra Murmu is currently serve as the Director, CSIR-Central Mechanical Engineering Research Institute, Durgapur since February 06, 2023. He is a Fellow of Indian National Academy of Engineering (INAE) and National Academy of Sciences, India (NASI). Dr. Murmu received his Bachelor of Engineering in Mechanical Engineering from erstwhile Bengal Engineering College in 1992 (currently IIEST, Shibpur), Master of Engineering in Mechanical Engineering from Indian Institute of Science, Bangalore in 1994 and Ph.D in Mechanical Engineering from Indian Institute of Technology (BHU), Varanasi in 2010.

Dr. Murmu was Chief Scientist at CSIR-Central Mechanical Engineering Research Institute, Durgapur. Prior to it, he has worked as a Scientist in CSIR-National Aerospace Laboratories, Bangalore. He also worked as a Visiting Scientist in the University of Erlangen-Nuremberg, Germany during 2001-2003 and Research Scholar in North Western University, USA during 2011-2012. He is also Dean of Faculty of Engineering Sciences, Academy of Scientific and Innovative Research (AcSIR), An Institute of National Importance - Set up by an act of Parliament.

He is a recipient of VASVIK Award-2015, National Design Award (2012), CSIR-Raman Research Fellowship (2011), DAAD Fellowship (German Academic Exchange Service)-2000, CSIR@70 Recognition for Developing Five Axis Micro Milling Machine (2012) and Co-Author of MSEB Best Paper Award, 2014 (Elsevier).

Dr. Murmu has published over 100 research papers in SCI journals, 4 book chapters and has filed 7 patents & 8 copyrights/design registrations. He is currently serving as Associate Editor, Journal of the Institution of Engineers (India) Series–C and Co-Guest Edited the Special Theme(2017): India's Reusable Launch Vehicle Technology Demonstrator: The Future of Space Transportation System with Shri Sivan, K., Chairman, ISRO. He has supervised/co-supervised 9 PhD students and several Master students.

His current research interest includes Additive and Smart Manufacturing, Graphene Composite, Ink & Lubricants. Some of his notable contributions include development and co-development of 5 Axis Micro-Milling Machine, AM Machine - Direct Energy Deposition System, Foil Bearing, EHD-Ink Jet Printing, Graphene Lubricant for hot forging and Miniature Turbine.

Meet Our Guest of Honour

Dr. Mou Sen
Joint Director, MSME&T, Govt. of West Bengal

Mou Sen is presently working at Joint Director at Department of MSME&T, Government of West Bengal since 2017. She has more than twenty eight years of working experience in which 11 years she served as a General Manager at District Industries Centre at different districts of West Bengal under Department of MSME&T, Govt. of West Bengal. She is a PhD in Chemical Engineering from National Institute of Technology, Durgapur, on ground water pollution and effect of arsenic. She spent 9 years in R& D lab, headed a chemical engineering department of a private engineering institute for 1 year, and published good number of papers in premier journals of science and technology on different environmental issues.

However, she found her lifelong research interest on rural entrepreneurship and livelihood, its technological development, local economic development through development of MSME, handloom and handicraft artisanship, sustainable business and life practices, women entrepreneurship, etc after joining in West Bengal Government by topping in Entrance Examination of Public Service Commission. In her area of Sustainable Entrepreneurship, She talked and presented papers in Calcutta Business School, Indian Institute of Management Calcutta, Xavier Institute of Management, Bhubaneswar, National Institute of Fashion Technology, and various colleges and universities. Recently, she got an acceptance of publishing one of her papers in Springer on 'opportunities of Moringa Powder for Sustainable Livelihood'.

Apart from her administrative role she has interest in writing poetries, practicing & promoting sustainable lifestyle, ethical consumption, feminism and humanism. Two of her poetry books and one short story book got published by renowned publishing house. Her writings get published in popular newspapers like Anandabazar Patrika and many Bengali magazines.

Meet Our Guest of Honour

Mr. Ponnan Murugesan
Executive Director (Projects), SAIL-DSP

Biography: Sri P Murugesan, A BE(Mech.) from Madras University, having 34 years of working experience in three plants of SAIL namely Salem Steel Plant, Bhilai Steel Plant and Durgapur Steel Plant. He has acquired expertise in managing Cold Rolling Mills and Hot Rolling Mils at Salem. During his tenure as in-charge of Medium Structural Mill at Durgapur, the plant production increased from 17% to 37% of rated capacity. During his tenure as head of Prestigious Universal Rail Mill at Bhilai, the plant production ramped up from 1200Tonne per day to 2200 T per day. As Executive Director (Projects), Shri Murugesan steered implementation of large capital projects for technological upgradation, Digital Project Monitoring System, e-Tendering platform and Electronic Design and Drawing Management System etc. He steered brownfield expansion of 2.4 to 3.4 MTPA and Green Field expansion of 2.5MTPA at Durgapur Steel Plant. He visited countries like Italy, Belgium, Netherlands, Switzerland and Saudi Arabia for his profession assignment and development.

Meet Our Keynote Speaker

Prof. Jamuna Kanta Sing
Professor,
Department of Computer Science & Engineering
Jadavpur University, Kolkata

Biography: Jamuna Kanta Sing has received his B.E. (Computer Science & Engineering) degree from Jadavpur University in 1992, M.Tech. (Computer & Information Technology) degree from Indian Institute of Technology (IIT) Kharagpur in 1994 and Ph.D. (Engineering) degree from Jadavpur University in 2006. Dr. Sing has joined the Department of Computer Science & Engineering, Jadavpur University in March 1997 and presently serving as a Professor since 2010. He is a recipient of the BOYSCAST Fellowship of the Department of Science & Technology, Govt. of India for doing advanced research at the University of Pennsylvania and the University of Iowa, USA in 2006 and the UGC Research Award in 2014. He is a senior member of the IEEE, USA. He has published more than 45 research papers in SCI/SCOPUS and other reputed refereed International Journals and more than 65 papers in international conferences. He has supervised 12 PhD scholars and completed 5 R&D projects from the AICTE, UGC and DST of worth around ₹65 Lakhs as principal investigator (PI). His research interest includes face recognition and detection, video analytics, medical image processing, computational intelligence and pattern recognition.

Title of the Talk: Advances in Fuzzy Clustering Algorithms for Medical Image Segmentation

Abstract: This talk focuses on the recent advances of fuzzy clustering algorithms for medical image segmentation, especially for the brain magnetic resonance (MR) images. It also briefly includes the different modalities for acquiring medical images. Since the brain MR images are corrupted by noise and intensity inhomogeneity (IIH), the segmentation task becomes very difficult and challenging, resulting erroneous results. Starting from the well-known clustering algorithms, it covers some of the advanced fuzzy clustering algorithms for segmentation of brain MR images.

Meet Our Invited Speaker

Prof. Sunil Karforma
Professor & Dean (Faculty Council for PG Studies)
Department of Computer Science
The University of Burdwan, Burdwan

Biography: Professor Sunil Karforma teaches courses in the discipline area of computer and information sciences. Develops and designs curriculum plans to foster student learning, stimulate class discussions, ensure student engagement to the student, and give knowledge on the basic components of a computer, as well as their uses and applications. In addition to teaching, undertake research proposals, supervise research scholars and publish papers, geared to enhance the learning experience. Professor Sunil Karforma has completed a Bachelor of Computer Science and Engineering and a Master of Computer Science and Engineering from Jadavpur University. He has a Ph.D. in cryptography under his belt. He presently holds the positions of Dean of, the Faculty Council for PG Studies in Science at the University of Burdwan and Professor for the Department of Computer Science at The University of Burdwan. He has 24 years of teaching experience. His research interests include Machine Learning, Artificial Intelligence, Network Security, Electronic Commerce, Cryptography, Block Chains, the Internet of Things and Elliptic Curve Cryptography. He has published more than 150 research publications, of which many are indexed in SCI, SCOPUS, WEB-OF-SCIENCE and other databases. He is also the owner of four patents issued by various nations all over the world. 12 Ph.D. scholars have come from him, and several are still working under his direction. In addition to holding various administrative positions at numerous institutions, serves as an external member of the Board of Studies (BOS) and the Board of Research Studies (BRS) of various reputed Universities Such as Binod Bihari Mahto Koyalanchal University (Dhanbad), Kazi Nazrul University, Bankura University, Rani Rashmoni Green University, University of Gour Banga, Barasat State University. He is associated with the West Bengal College Service Commission (WBCSC) as a member of the selection committee (2019 & 2022). He is also acting as an examiner, moderator and paper setter in different universities like Jadavpur University, Kalyani University, North Bengal University, etc.

Title of the Talk: Application of Machine Learning Techniques in Agriculture: An Overview

Abstract: For many developing nations, agriculture is the fundamental source of the economic engine. Without a significant increase in food production, the rise in global population during the

21st century would not have been conceivable. Artificial Intelligence (AI) is defined as the combination of computers and other associated tools to simulate human decision making and problem solving abilities. In the present era, AI is the most trending technology in the field of Computer Science and Engineering that concerns building intelligent and smart machines capable of performing automated task. Modern technology such as Machine Learning technique of AI has caused about major changes in agriculture, offering solutions to improve profitability, productivity and sustainability. Thus, a key issue in agricultural expansion is the accurate forecast of healthy crop output / crop type utilizing machine learning (ML) algorithms. Prediction of crop yield is influenced by input factors including location, irrigation methods, temperature, pesticide application, etc. Farmers find it difficult to meet the changing needs of the earth. They toil day and night to ensure a plentiful harvest at session's end. The availability of water, soil fertility, the prevention of rodent damage to crops, the timely use of pesticides and other helpful chemicals, and the presence of nature all contribute to a healthy harvest. A farmer can regulate the quantity and frequency of pesticide applications, even though many of these other elements are difficult to manage. The precise prediction of healthy crop output using machine learning algorithms is thus a crucial problem in agricultural growth. To establish an accurate effective model for crop classification, including crop yield estimates based on weather, crop disease, crop classification based on the growth phase, etc., several investigations are advised. Five popular algorithms, such as K-Nearest Neighbor (KNN), Random Forest (RF) Classifier, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Multi-layer Perceptron (MLP), are compared on the crop dataset using the metrics mean accuracy, standard deviation, and boxplot for the purpose of crop yield prediction. Experimental results reveal that XGBoost gives the highest accuracy among the five, while MLP and LightGBM show a delicate difference with XGBoost.

Meet Our Invited Speaker

Dr. Rajesh P. Barnwal
Principal Scientist
CSIR-Central Mechanical Engineering Research Institute,
Durgapur, West Bengal

Biography: Dr. Rajesh P. Barnwal is serving as a Principal Scientist of CSIR-India and is currently leading the AI & IoT Lab at CSIR-Central Mechanical Engineering Research Institute, Durgapur. He has more than 16 years of professional and R&D experience in the field of Computer Science and allied disciplines. He earned his M.Tech. and Ph.D. from the Indian Institute of Technology (IIT), Kharagpur, India. He was awarded the Senior Member award by the Association of Computing Machinery (ACM), USA, in the year 2013 and was elevated to Senior Member grade of IEEE in the year 2022 for his significant contributions in the field of IT/ CSE. Dr. Barnwal has led several R&D and industrial consultancy projects funded by different agencies of Govt. of India in the field of Artificial Intelligence, Internet of Things, Data Informatics, and Cyber-Physical Systems. He has authored several research papers, book chapters, and technical reports in SCI journals and International Conferences. He also has several patents, copyrights, and design registrations to his credit for his authored software and technological products. He is also associated with the Academy of Scientific and Innovative Research (AcSIR), India as its Associate Professor. Further to this, Dr. Barnwal has also served as one of the guest editors of the International Journal of Distributed Sensor Networks and is currently in the reviewers' panel of several high-impact-factor SCI journals. Recently, he has been awarded the prestigious INSA Visiting Scientist Fellowship for the year 2022-23 from Indian National Science Academy. His current areas of research interest include Intelligent Machines, Internet of Things, Cloud Computing, and Cyber-Physical Systems.

Title of the talk: Modelling complex self-organization systems through Modern Machine Learning Approach

Abstract: A self-organization system is a dynamic, decentralized structure where individual components or agents autonomously interact and adapt to achieve collective goals without centralized control. These systems are prevalent in nature, ranging from ant colonies to neural networks, and are increasingly utilized in various fields, including biology, engineering, and

computer science. Understanding self-organization is one of the key tasks for controlling and manipulating the structure of materials at the micro- and nanoscale. Inspired by the strength of neural networks and the dynamics of cellular automata, Neural Cellular Automata (NCA) offer a new paradigm in the field of artificial intelligence and computer systems. Fundamentally, NCA represents the essence of emergent behaviour and decentralized computation. Through the application of local transition rules and the arrangement of interconnected nodes in a grid-like structure, NCA demonstrates exceptional ability in pattern development, self-organization, and dynamic system modelling. We go into the architecture of NCA, explaining how they are evolved through an iterative updating process and emphasizing their parallelism and scalability. The session shall also explore the theoretical foundations of NCA and explain how they may simulate complicated systems using basic rules that control the interactions of networked nodes. The talk would further present how emergent behavior, self-organization, and pattern development are captured in NCA models. Moving from theory to reality, different real applications of NCA will be discussed.

Meet Our Invited Speaker

Dr. Sandip Kumar Lahiri
Associate Professor
Department of Chemical Engineering
National Institute of Technology Durgapur

Biography: Dr. Sandip Kumar Lahiri is a seasoned and accomplished Professional Chemical Engineer, holding a Doctorate degree (PhD) in Chemical Engineering. With a career spanning over 25 years in the petrochemical industry across globe, Dr. Lahiri has left an indelible mark on the field, showcasing expertise in key domains such as production, technical services, and technology advisory.

Dr. Lahiri's journey began with a solid educational foundation, obtaining a Bachelor's degree in Chemical Engineering from Jadavpur University, followed by a Master's degree from the Indian Institute of Technology (IIT) Kharagpur and PhD from NIT Durgapur. This foundation laid the groundwork for an illustrious career that included pivotal roles in various esteemed organizations across the globe.

During this time, Dr. Lahiri became an authority in the production and technical services of petrochemical plants across globe. Furthermore, the extensive international exposure gained through consultancy positions with prominent petrochemical companies in Europe, US, the Middle East, and Asia has honed Dr. Lahiri's understanding of global industry dynamics and challenges.

An avid researcher and thought leader, Dr. Lahiri has contributed significantly to the advancement of chemical engineering. With over 60+ technical publications in esteemed international journals, covering a diverse range of subjects including modeling and simulation, artificial intelligence, process design, optimization, fault diagnosis, and computational fluid dynamics, Dr. Lahiri has proven expertise in driving innovation in the field.

Notably, Dr. Lahiri is an author of three influential books and holds European and US Patents for ground breaking work in online fault diagnosis in chemical plants. Recognized in the World Who's Who for outstanding contributions, Dr. Lahiri remains committed to pushing the boundaries of chemical engineering, particularly in the realms of Industry 4.0 applications,

advanced process control, fault diagnosis systems, AI application in process industry, and energy optimization in process plants globally.

Currently serving as an Associate Professor at NIT Durgapur, Dr. Lahiri continues to impart knowledge and nurture the next generation of engineers, solidifying a legacy of innovation and academic excellence.

Title of the Talk: Transformative Impact of Artificial Intelligence: Trends, Applications, and Workforce Implications Across Industries

Abstract: Artificial Intelligence (AI) stands as a pivotal force in contemporary technological advancements, exerting a profound influence across a myriad of industries and engendering transformative opportunities. This abstract delineates the prevailing trends, prospective applications, and workforce ramifications of AI across diverse industry verticals.

In the manufacturing sector, AI-driven predictive maintenance and quality control are effectuating revolutionary changes. Leveraging machine learning algorithms, real-time sensor data is scrutinized to prognosticate equipment failures, consequently curtailing downtime and augmenting machinery longevity. Complementing this, AI-powered computer vision systems meticulously inspect products for defects, ensuring the production of high-calibre goods.

The healthcare industry is undergoing a metamorphosis catalysed by AI, enhancing diagnostics, treatment optimization, and personalized medicine. Machine learning models facilitate precise diagnosis by parsing through medical images, genomic data, and patient records. Furthermore, AI-fuelled predictive analytics empower healthcare providers to pre-emptively address patient needs and judiciously allocate resources.

AI technologies are profoundly reshaping the retail landscape, bolstering personalized marketing initiatives and amplifying customer experiences. By analysing consumer behaviour and historical purchase data, machine learning algorithms orchestrate targeted marketing campaigns and refine pricing strategies. Additionally, AI-driven chatbots and virtual assistants expedite shopping processes and elevate customer satisfaction levels.

In the finance domain, AI is gaining traction for fraud detection, risk management, and algorithmic trading. Sophisticated AI systems scrutinize transactional patterns to discern anomalous activities and flag potential fraud in real-time. Concurrently, machine learning algorithms evaluate market dynamics, facilitating high-frequency trading characterized by unparalleled speed and precision.

The transportation sector is witnessing a paradigm shift with the advent of AI-powered autonomous vehicles and intelligent logistics solutions. Equipped with advanced sensors, computer vision capabilities, and machine learning algorithms, vehicles are endowed with the ability to perceive and navigate their environment autonomously. Complementing this, AI-driven logistics optimization systems refine route planning and streamline supply chain management.

Looking ahead, AI's prospective applications in the energy sector encompass smart grid management and renewable energy optimization. Machine learning algorithms are adept at analysing energy consumption patterns and forecasting demand fluctuations, thereby optimizing

grid operations and facilitating the strategic placement of renewable energy sources like solar and wind power.

In the agricultural landscape, AI harbours the potential to revolutionize farming practices through precision agriculture and crop monitoring. Machine learning algorithms assimilate soil health data, weather forecasts, and crop growth metrics to fine-tune irrigation, fertilization, and pest control methodologies.

The construction industry is poised for a transformational journey with AI at its helm, automating design optimization and augmenting safety monitoring. Generative design algorithms harness AI to explore innovative design possibilities, optimize building layouts, and curtail material wastage.

Within the educational arena, AI's potential is manifested in personalized learning and adaptive assessment paradigms. Machine learning algorithms analyse student performance metrics and tailor educational content to individual requirements, thereby enhancing engagement and academic outcomes.

In the entertainment sector, AI technologies are recalibrating content delivery mechanisms and virtual reality experiences. By scrutinizing user preferences, machine learning algorithms proffer personalized content recommendations. Concurrently, AI-driven virtual reality platforms engender immersive gaming and entertainment experiences.

The integration of AI into industries engenders profound workforce implications. Notwithstanding the potential for job creation and productivity enhancement, AI precipitates challenges pertaining to job displacement, skill gap, and evolving job roles. Reskilling and upskilling initiatives are imperative to bridge the burgeoning skill gap, with a substantial portion of jobs at risk of automation or undergoing significant metamorphosis.

In conclusion, AI emerges as a transformative force redefining various industry paradigms, catalysing innovation, optimizing efficiency, and engendering novel opportunities. While AI's potential for growth is unequivocal, it concurrently presents challenges necessitating proactive strategies encompassing skill acquisition, continuous learning, ethical considerations, and inclusive innovation to navigate the impending AI revolution adeptly.

Meet Our Session Chair

Dr Abhiram Hens

Biography: Dr Abhiram Hens joined the Department of Chemical Engineering, National Institute of Technology (NIT), Durgapur, as an Assistant Professor in 2019. Before that, he worked as a Scientist and Senior Scientist in CSIR – Central Mechanical Engineering Research Institute, Durgapur, for almost nine years. He completed his Ph.D in Engineering Science from Academy of Scientific and Innovative Research (AcSIR –under CSIR) in 2016. He did his M.Tech from IIT Kanpur in 2010 and B.Tech from Calcutta University in 2008. His areas of expertise and research interest lies in the field of CFD and Heat Transfer, Process Modeling and Simulation, Molecular Dynamics simulations, Microfluidics and Nanotechnology. He has published 31 research papers in reputed SCI indexed international journals like Physical Review E, Chemical Engineering Science, Physics of Fluid, Journal of Colloid and Interface Science, Journal of Chemical Physics, International Journal of Heat and Mass transfer etc. He has also authored several conference papers and a book chapter. He has filed three Indian patents (one granted). He completed two CSIR funded, and another three institute funded projects as PI and another five projects as co-PI. In 2018, he received CSIR Young Scientist Award. He received Thermax Sampark Fellowship in 2020. Currently, he is one of the Associate Deans (Research & Consultancy) of NIT Durgapur.

Meet Our Session Chair

Prof. Monish Chatterjee

Biography: Monish Chatterjee is presently working as a Professor and Head of the Department of Computer Science and Engineering of Asansol Engineering College, Asansol, West Bengal, India. He received the Ph.D. (Engineering) degree in the branch of Computer Science and Engineering from Jadavpur University, Kolkata, India in 2012 and the M.E. (Computer Science and Technology) degree from Indian Institute of Engineering Science and Technology (IIEST), Shibpur, India in 2000 formerly Bengal Engineering and Science University (BESU), Shibpur, India, one of the oldest pioneering engineering institutes in India. He started his career as a Software Engineer in the IT industry, but after a short period he subsequently switched over to academia. He has 21 years of teaching experience and 2 years of IT industry experience. He has so far supervised 3 PhD thesis, two of them are awarded form IIEST Shibpur and the other from Jadavpur University. His research interests are solving problems related to routing, spectrum allocation, quality of transmission, survivability and energy-efficient communication in optical networks and OFDM based optical grids/clouds. He has publications in refereed International Journals and IEEE International Conference Proceedings. He is a reviewer of the following International Journals: IEEE Transactions on Green Communications and Networking, Computer Communications (Elsevier), Optical Switching and Networking (Elsevier) and IEEE Access. He has chaired sessions and has been a member of technical program committee of many International Conferences and Workshops.

Meet Our Session Chair

Dr. Gour Sundar Mitra Thakur

Biography: Gour Sundar Mitra Thakur is an AI enthusiast with 16 years of teaching and research experience. After pursuing B. Tech and M. Tech in Computer Science and Engineering, he completed his PhD, with the thesis titled "Intelligent Systems in Manifold Decision Making" from National Institute of Technology Durgapur, India. He is currently associated with Dr. B. C. Roy Engineering College Durgapur as Associate Professor in the department of Computer Science and Engineering (AIML). He is able to publish his research findings in 22 international journals, national and international conferences and book chapters till date. He is working as a reviewer of many reputed international journals like, Knowledge Based Systems, Applied Soft Computing, International Journal of Fuzzy Systems, Frontiers in Oncology etc. He is passionate about research and curious about applying different AI/ML/DL algorithms to solve various real life problems in biology, agriculture, environment and manufacturing industries. He has the ability to collaborate with other experts from different fields to achieve a common goal. He believes in and is committed to develop AI systems that are fair, transparent, and unbiased. He has developed various modern ML/DL models using Tensorflow and Pytorch and developed some expertise in ideating ML pipelines for complex real-life problems.

Meet Our Session Chair

Dr. Avishek Banerjee

Biography: Avishek Banerjee received the B.Tech. degree in information technology from Vidyasagar University, in 2004. He received the M.Tech. degree in information technology from Calcutta University, in 2007, and the Ph.D. degree in engineering from Jadavpur University, in 2019. Currently, He is holding the position of HOD (CSBS) at Asansol Engineering College. Prior to this, he was an Assistant Professor in the School of Computer Science Engineering and Technology, Bennett University, New Delhi. He has worked in several scientific fields, such as reliability optimization and evolutionary algorithms. He also has deep knowledge in engineering fields, such as wireless sensor networks, power distribution systems, machine and deep learning, and natural language processing. He received the EU sponsored "Erasmus Mundus" scholarship with the University of Gheorghe Asachi, Iasi, Romania, for 23 months of mobility. He has contributed to the writing of more than 40 research papers, out of them 15 are journal articles. He has more than 200 academic citations as per Google scholar. He published more than 10 patents in different engineering fields. He is a Life Member of the Institute of Engineers (IEI) and the Indian Society of Technical Education (ISTE). He is serving as a reviewer for several international and national journals and conferences.

Meet Our Session Chair

Dr. Mamata Dalui

Biography: Mamata Dalui Received the B.Sc.(Hons) degree in Physics from Vidyasagar College, Calcutta University, West Bengal, India, in 2004, the B.Tech. degree in Computer Science and Engineering from Calcutta University, in 2007, M.Tech degree from Bengal Engineering and Science University, Shibpur, West Bengal, India, in 2009 and Ph.D. degree in Engineering from Indian Institute of Engineering Science and Technology, Shibpur, West Bengal, India. Presently, she is an Assistant Professor in the Department of Computer Science and Engineering, National Institute of Technology, Durgapur, West Bengal, India. Her research interests include Intelligent Biomedical Devices and Assistive Technology, Hardware Security, VLSI design and test, Theory and applications of Cellular Automata.

Meet Our Session Chair

Dr. Anirban Sarkar

Biography: Anirban Sarkar is presently a faculty member in the Department of Computer Science & Engineering, National Institute of Technology, Durgapur, India. He received his PhD degree from National Institute of Technology, Durgapur, India in 2010. His areas of research interests include Cloud Computing, Database Systems and Software Engineering. He has more than 125 publications in various international Journals and Conferences. He has handled several research projects funded by various Govt. agencies of India. Dr. Sarkar has contributed as International Programme Committee member and special session chair of several conferences associated with reputed publishers.

Meet Our Session Chair

Dr. Sankar Mukherjee

Biography: Dr. Sankar Mukherjee is working as an Associate professor and Head of the department of Computer science and engineering at Sanaka Educational Trust's group of Institutions, Durgapur. He has almost 23 years of teaching experiences in different engineering college and universities. He has completed Ph.D. and M.Tech in computer science and engineering from IIT(ISM), Dhanbad. His research field is MANET, WSN, IoT and Machine Learning. He has published papers in several SCIE and Scopus indexed journals and International Conferences.

Meet Our Session Chair

Mr. Debi Prasad Das

Biography: Sri Debi Prasad Das was born in 1978 in a suburban area of Hooghly district. He graduated in science from Calcutta University in 1998. For the next seven years, he worked in Rishra Municipality, Irrigation Department of West Bengal Govt. at Hooghly, Simplex Concrete Piles(I) Ltd in Orissa, Serampore-Uttarpara Block under W.B. Govt. In 2005, he joined CSIR-CMERI, Durgapur for research work. He serves in various projects of Indian Government i.e. COR-13, COR-22, COR-12, etc. and assists Non-Destructive Testing, Health Assessment of structure for various industries i.e. IOCL, NSPCL, NTPC, etc. He also excelled further in his education career by completing AMIE from IEI in 2008, MTech in Structural Engineering from NIT Durgapur in 2011, Submitted PhD thesis in March 2023 in NIT Durgapur. He published four research papers in prestigious journals and four papers in International Conferences. He had done a Patent in the field of "composition for concrete comprising building demolition waste" and a Copyright on the "process for development of concrete of grade M15 to M20 using construction debris as coarse aggregate for application in pavement." He is now servicing at CSIR-CMERI as the Sr. Tech. Officer for over 19 years.

Meet Our Session Chair

Mrs. Priyanka Roy

Biography: Priyanka Roy is an accomplished educator and AI enthusiast with a strong background in Information Technology. With an extensive 18-year teaching journey, Mrs. Priyanka Roy currently serves as an Assistant Professor in the Department of Information Technology at Dr. B.C. Roy Engineering College, Durgapur. She has demonstrated expertise in Network Security, Artificial Intelligence (AI), and Machine Learning (ML). She has published three papers in national and international journals and has also contributed as a contributor to a patent titled "Deep Reinforcement Learning for Energy-Efficient Computation Offloading with DVFS for Time-Critical IoT Applications in Edge Computing. She has supervised many B.Tech. and M.Tech. Projects, mentoring students in their research endeavours. Mrs. Priyanka Roy is actively engaged in pioneering research focused on the application of AI and ML to address real-life global challenges. With her dedication to exploring innovative solutions, she exemplifies a commitment to advancing technologies that tackle pressing issues on a global scale.

PROGRAM SCHEDULE

Time	DAY 01, April 6, 2024 (Saturday)	Essential Information
9:30 AM – 10:00 AM	Registration Venue: Main Entrance Desk	
10:00 AM – 10:30 AM	Inauguration Program Venue: Visvesvaraya Auditorium	Chief Guest: Dr. Naresh Chandra Murmu, Director, CSRI-CMERI Guest of Honour: Dr. Mou Sen , Director (MSME-HQ, Govt. of WB) Guest of Honour: Mr. P. Murugesan, Executive Director (Projects), SAIL-DSP
10:30 AM – 10:50 AM	TEA BREAK Venue: Dining Hall	
10:50 AM – 11:40 AM	KEYNOTE TALK Venue: Visvesvaraya Auditorium	Prof. Jamuna Kanta Sing Professor, Jadavpur University, Kolkata
11:40 AM – 12:25 PM	INVITED TALK 1 Venue: Visvesvaraya Auditorium	Dr. Rajesh P. Barnwal Principal Scientist, CSRI-CMERI, Durgapur
12:30 PM – 1:30 PM	ORAL SESSION 1 Venue: Visvesvaraya Auditorium	Chair: Dr. Anirban Sarkar, NIT Durgapur Co-Chair: Dr. Rajib Ghosh Chawdhuri PAPER ID: 11, 12, 26, 42
	ORAL SESSION 2 Venue: Lecture Hall	Chair: Dr. Gour Sundar Mitra Thakur, Dr. B. C. Roy Engineering College Co-Chair: Mr. Dei Prasad Das PAPER ID: 66, 85, 88, 21
	ORAL SESSION 3 Venue: Committee Room	Chair: Ms. Priyanka Roy, Dr. B. C. Roy Engineering College Co-Chair: Mr. Manas Majumder PAPER ID: 24, 30, 34, 62
1:30 PM – 2:30 PM	LUNCH BREAK Venue: Dining Hall	
2:30 PM – 4:00 PM	ORAL SESSION 4 Venue: Visvesvaraya Auditorium	Chair: Dr. Mamata Dalui, NIT Durgapur Co-Chair: Mr. Parasumram Shaw PAPER ID: 06, 08, 14, 22, 23
	ORAL SESSION 5 Venue: Lecture Hall	Chair: Dr. Avishek Banerjee, Asansol Engineering College Co-Chair: Dr. Arup Kumar Mandal PAPER ID: 33, 48, 83, 73, 53
	ORAL SESSION 6 Venue: Committee Room	Chair: Dr. Sankar Mukherjee, SETGOI Durgapur Co-Chair: Mr. Debasis Guha PAPER ID: 76, 77, 78, 80
4:00 PM – 4:30 PM	TEA BREAK Venue: Dining Hall	
4:30 PM – 6:00 PM	ORAL SESSION 7 Venue: Visvesvaraya Auditorium	Chair: Dr. Rajib Ghosh Chaudhuri, NIT Durgapur Co-Chair: Dr. Tushar Kanti Bera PAPER ID: 90, 57, 61, 65, 81, 82, 39
	ORAL SESSION 8 Venue: Lecture Hall	Chair: Dr. Abhiram Hens, NIT Durgapur Co-Chair: Mr. Kausik Chatterjee PAPER ID: 44, 35, 36, 37, 45, 46
	DAY 02, April 7, 2024 (Sunday)	Essential Information
10:00 AM – 10:45 AM	INVITED TALK 2 Venue: Visvesvaraya Auditorium	Dr. Sunil Karforma Professor, The University of Burdwan
10:45 AM – 11:05 AM	TEA BREAK Venue: Dining Hall	
11:05 AM – 11: 50 AM	INVITED TALK 3 Venue: Visvesvaraya Auditorium	Dr. Sandip Kumar Lahiri, Associate Professor, NIT Durgapur
12:00 NOON – 1:30 PM	ORAL SESSION 9 Venue: Visvesvaraya Auditorium	Chair: Dr. Dakshina Ranjan Kisku, NIT Durgapur Co-Chair: Mr. Sudipto Sikder PAPER ID: 20, 25, 41, 56, 60
	ORAL SESSION 10 Venue: Lecture Hall	Chair: Dr. Monish Chatterjee, Asansol Engineering College Co-Chair: Mr. Safikul Islam PAPER ID: 13, 69, 71, 74, 58
	ORAL SESSION 11 Venue: Committee Room	Chair: Mr. Debi Prasad Das, CMERI Durgapur Co-Chair: Dr. Bimal Das PAPER ID: 28, 31, 32, 38, 54, 91
1:30 PM – 2:30 PM	LUNCH BREAK Venue: Dining Hall	
2:30 PM – 2:40 PM	VALEDICTORY SESSION Venue: Visvesvaraya Auditorium	Chair: Dr. Monish Chatterjee, Asansol Engineering College Vote of Thanks by Honorary Secretary, IEI DLC
2:40 PM	MLSC-ES 2024 CLOSES	

INDEX

Enhancing Distillation Column Performance: Soft Sensor Development Using Artificial Neural Networks and Multi-gene genetic programming Models

Saswata Banerjee[1], Srinjay Rakshit[1], and Sandip Kumar Lahiri[1*]

[1]Department of Chemical Engineering, National Institute of Technology Durgapur,
Durgapur, West Bengal, India
sklahiri.che@nitdgp.ac.in

Extended Abstract

Distillation plays a pivotal role in the chemical industry for separating miscible liquids, with product purity being a paramount target. However, distillation column product purity indication is notavailable in real time basis due to non-availability of online analyzers or very high cost of these type of industrial analyzers. The lack of readily available online analyzers poses a challenge, as offline lab analyses are time-consuming (typically 3-8 hrs delay) and lead to product loss during distillation column disturbances as plant operator becomes clueless about the impact of their preventive or corrective actions on product purity. To address this issue, this study proposes the development of a soft sensor for bottom product purity of industrial distillation columns using available real-time process parameter data i.e typically distillation columns intermediate tray teperatures.

Data are collected fromindustrial distillation column separating alcohol fromwater. Input variable consists of 3 tray temperature of distillation column namly tray 12 , tray 27 and tray 42 temperatre and corresponding bottom product purity (water wt%) is kept at output. Two promising machine learning models namely Artificial Neural Networks (ANNs) & Multi-gene genetic programming (MGGP) are employed to predict the bottom product purity from 3 tray temperatures.

Artificial Neural Networks (ANNs) are computational models inspired by the human brain's neural structure, composed of interconnected nodes or "neurons." They work by processing input data through layers of these neurons, each applying weights and activation functions, to produce an output that is optimized through iterative training using algorithms like backpropagation to minimize prediction errors [1].

On the other hand ,Multi-gene genetic programming (MGGP) is an advanced machine learning technique that evolves computer programs using principles inspired by biological evolution. It starts with a population of randomly generated programs, evaluates their fitness based on a predefined objective function, and iteratively evolves the programs through selection, crossover, and mutation to find the most fit solution [2].

Employing artificial neural networks (ANN) and multi-gene genetic programming (MGGP) (R^2 of 0.99 and RMSE of 0.002 for ANN, and R^2 of 0.98 and RMSE of 0.002 for MGGP) modelling techniques, the soft sensor is designed to predict bottom product purity—a critical indicator of distillation column performance. Three tray temperatures, readily available in real-time industrial settings, serve as input parameters for the model.

Both ANN and MGGP models exhibit high accuracy in predicting bottom product purity (refer Figure 2 and Figure 3 respectively), providing valuable insights into column performance without the need for costly online analysers. Notably, the MGGP model offers a distinct advantage over ANN by providing a closed-loop equation deployable within plant control loops (DCS). This feature enhances control and optimization capabilities within industrial settings, contributing to improved operational efficiency and product quality.

Furthermore, the MGGP model's transparency enhances its explainability compared to the black-box nature of ANN models. This transparency is crucial for understanding the underlying mechanisms driving distillation column behaviour, facilitating informed decision-making and process optimization.

By leveraging available real-time process parameter data and offline lab analysis, the developed soft sensor offers a cost-effective solution for monitoring and controlling distillation column performance. Its ability to accurately predict bottom product purity in real-time mitigates the need for frequent offline lab analyses,

reducing downtime and minimizing product loss during disturbances.

By offering immediate feedback on product purity, operators can swiftly identify and address any deviations from desired specifications, reducing the production of off-spec products. With real-time output from the sensor, operators can take proactive corrective and preventive actions without waiting for lab analysis results, minimizing wastage and increasing profitability. The reduction in off-spec product generation leads to more efficient resource utilization and higher production output, resulting in cost savings and improved profitability for the plant. Additionally, leveraging the real-time insights provided by the soft sensor allows for fine-tuning of process parameters and optimization of operations to maintain consistent product quality. Ultimately, the implementation of the soft sensor empowers plant engineers to enhance efficiency, competitiveness, and product quality standards.

In conclusion, this study demonstrates the effectiveness of soft sensors in enhancing distillation column operations by providing real-time insights into product purity. The combination of ANN and MGGP modelling techniques offers a robust approach for developing accurate and transparent soft sensors, thereby contributing to improved efficiency and profitability in the chemical industry.

Keyword(s): Artificial Neural Network; Multi-gene Genetic Programming; Distillation Column; Soft Sensors; Modelling and Optimization.

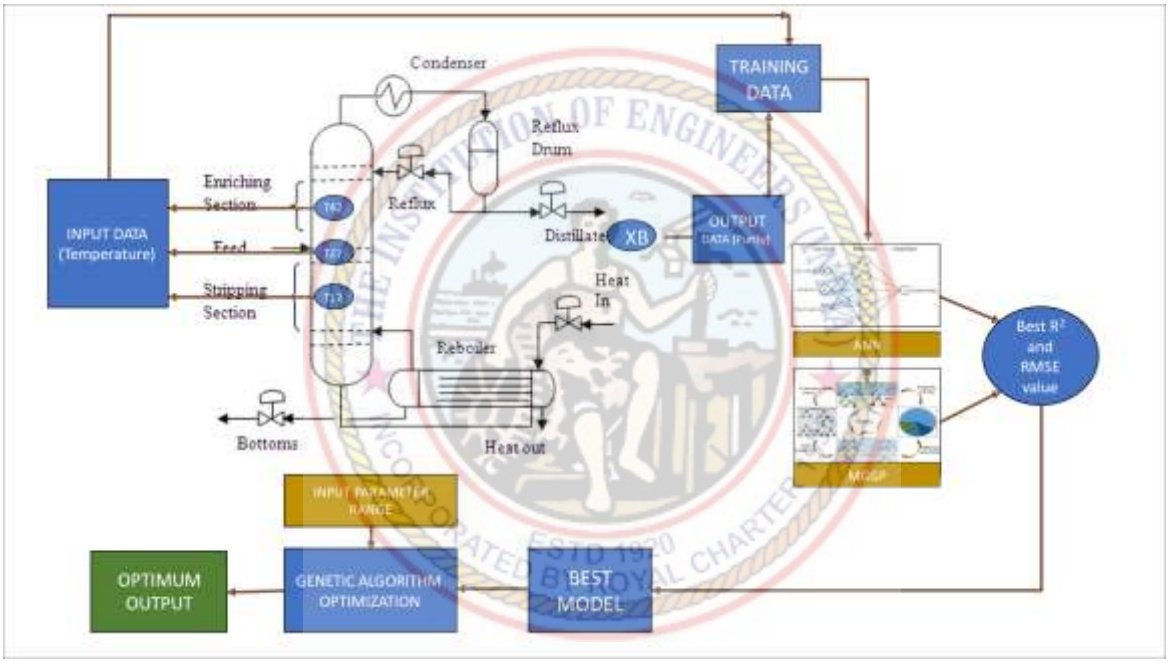


Figure 1: Schematic diagram of the study

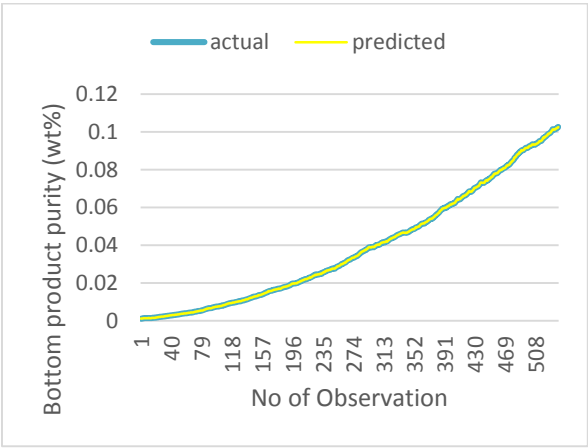


Figure 2: Predicted vs actual bottom product purity using ANN

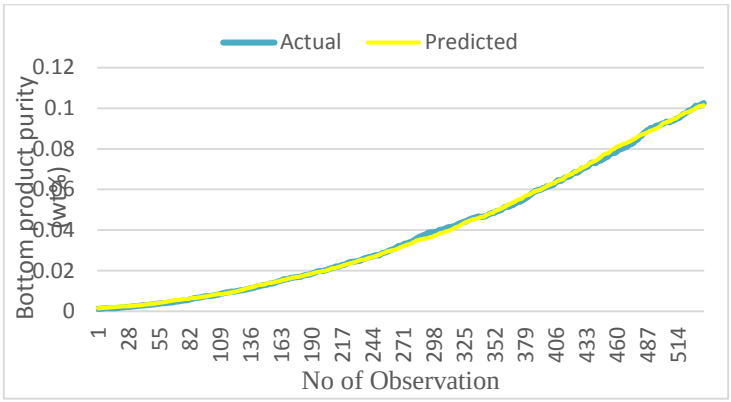


Figure 3: Predicted vs actual bottom product purity using MGGP

References

1. Shin Y, Smith R, Hwang S. Development of model predictive control system using an artificial neural network: A case study with a distillation column. Journal of Cleaner Production. 2020 Dec 20; 277:124124.
2. Searson DP, Leahy DE, Willis MJ. GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. InProceedings of the International multiconference of engineers and computer scientists 2010 Mar 17 (Vol. 1, pp. 77-80). Citeseer.

Enhancing Bioreactor Efficiency: Machine Learning Strategies for Glucon-ic Acid Yield Optimization

Srinjay Rakshit[1], Saswata Banerjee[1] and Sandip Kumar Lahiri[1*]

[1]Department of Chemical Engineering, National Institute of Technology Durgapur, Durgapur, West Bengal, India
sklahiri.che@nitdgp.ac.in

Extended Abstract

This study presents a comprehensive framework for the modeling and optimization of a bioreactor designed specifically for the production of gluconic acid—a process characterized by intricate and insufficiently understood kinetics. Leveraging a microbial process, the bioreactor converts glucose into gluconic acid, a valuable compound with widespread industrial applications. Experimental data were collected from [1] from an experimental bio-reactor where biomass, glucose and dissolved oxygen were reacted to produce gluconic acid. To tackle this challenge of poor understanding of reaction kinetics, genetic programming (GP) [2] and multi-gene genetic programming (MGGP) methodologies [3] were employed, aiming to capture the gluconic acid yield % of the bioreactor based on experimental observations. Both GP and MGGP are machine learning techniques that evolves computer programs using principles inspired by biological evolution. It starts with a population of randomly generated programs, evaluates their fitness based on a predefined objective function, and iteratively evolves the programs through selection, crossover, and mutation to find the most fit solution.

The chosen GP and MGGP methodologies are distinguished by their proficiency in discerning intricate patterns within data, thus facilitating the prediction of the bioreactor's performance i.e. gluconic acid yield % across diverse operational conditions. Essential input parameters, including initial glucose concentration, biomass concentration, and dissolved oxygen levels, were incorporated into the model to predict gluconic acid yield. Figure 1 summarize the whole process. Noteworthy is the introduction of a novel hybrid MGGP-genetic algorithm (hybrid MGGP-GA), which autonomously refines the MGGP model's complexity and meta parameters, streamlining the modelling process and alleviating users from the arduous trial-and-error methodology.

The performance of the developed models was meticulously evaluated, considering key factors such as predictability, interpretability, complexity, and adherence to the underlying kinetics of the bioreactor. Results revealed impressive accuracy, with an R^2 of 0.98 and RMSE of 0.002 for GP, and an R^2 of 0.99 and RMSE of 0.002 for MGGP [Figure 2]. Subsequently, upon establishing a reliable MGGP model, genetic algorithms were employed to optimize bioreactor process parameters, culminating in a noteworthy 25% increase in product yield facilitated by the hybrid MGGP-GA algorithm [Figure3].

Of the various models explored, the MGGP model emerged as the most effective, showcasing superior predictive capabilities. Surface plots derived from the MGGP model underscored a robust correlation between experimental results and model predictions, offering invaluable insights into the system's behaviour.

A notable advantage of the developed MGGP model over artificial neural networks (ANN) lies in its provision of a closed-loop equation deployable within plant control loops (DCS), furnishing industrial settings with heightened control and optimization capabilities. Furthermore, the transparency inherent in the MGGP model enhances its explainability compared to the black-box nature of ANN models.

Harnessing the developed models, genetic algorithms were enlisted to optimize reactor process parameters, leading to a substantial upsurge in product yield. The efficacy of the hybrid MGGP-GA algorithm in boosting product yield by 25% underscores its potential for augmenting plant profitability.

This research underscores the efficacy of machine learning methodologies in overcoming the challenges posed by poorly understood reaction kinetics, charting a trajectory towards enhanced efficiency and profitability in industrial processes. Additionally, optimization of the MGGP model identified optimal process conditions to maximize gluconic acid production rates, ultimately enhancing bioreactor efficiency and overall yield. These

insights offer valuable guidance for operational settings, presenting an avenue to bolster profitability in the bioprocess industry by optimizing product yield while minimizing resource utilization.

Keyword(s): Genetic Programming; Multi-gene Genetic Programming; Bioreactor; Modelling and Optimization; Gluconic acid production

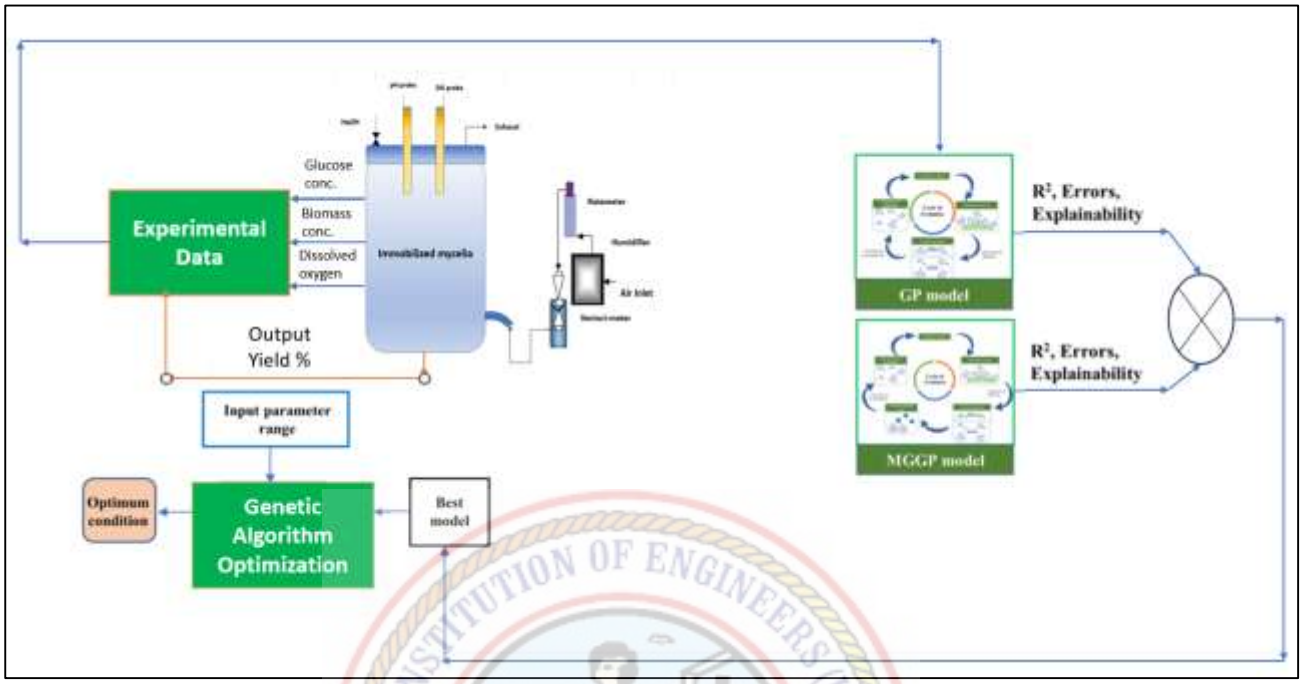


FIGURE 1 Schematic diagram of the study

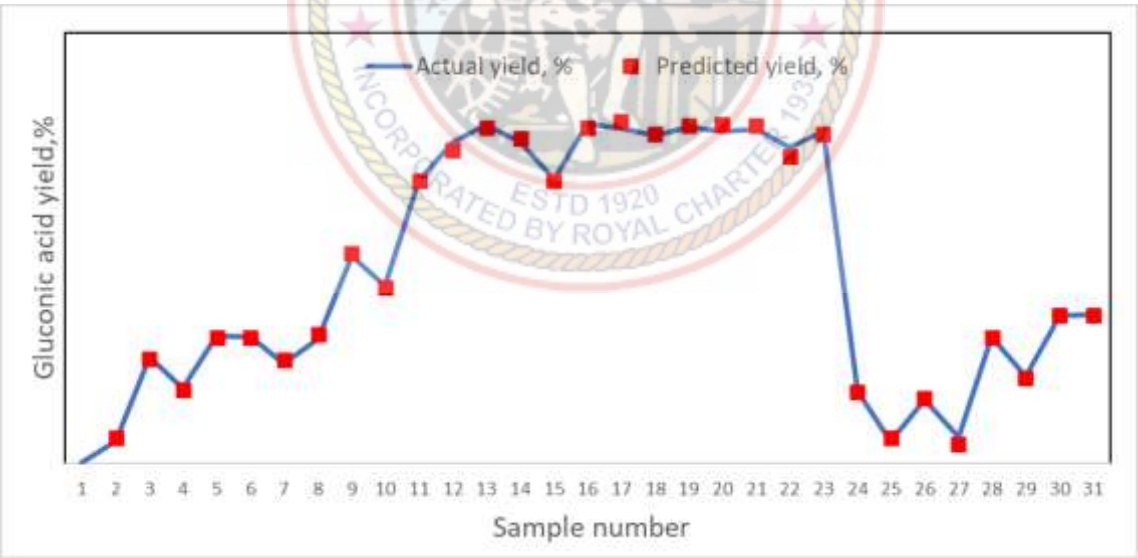


Figure 2: Experimental vs model predicted gluconic acid yield percent

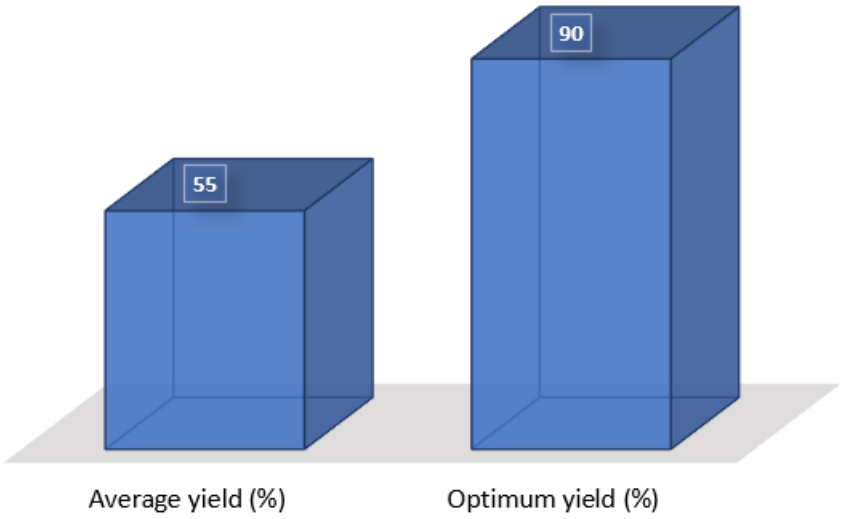


Figure 3: Gluconic Acid yield percentage before and after optimization

REFERENCE:

1. Cheema JJ, Sankpal NV, Tambe SS, Kulkarni BD. Genetic programming assisted stochastic optimization strategies for optimization of glucose to gluconic acid fermentation. Biotechnology progress. 2002;18(6):1356-65.
2. Koza JR. On the programming of computers by means of natural selection. Genetic programming. 1992.
3. Searson DP, Leahy DE, Willis MJ. GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. InProceedings of the International multiconference of engineers and computer scientists 2010 Mar 17 (Vol. 1, pp. 77-80). Citeseer.

Digital Twins: Transforming the Chemical Process Industry

Pratyush Kumar Pal[1], Sandip Kumar Lahiri[2]

[1]Information Technology Group, CSIR-CMERI, Durgapur, India
p_pal@cmeri.res.in
[2]Chamical Engineering, National Institute of Technology, Durgapur, India
sklahiri.che@nitdgp.ac.in

Extended Abstract

The chemical industry, a cornerstone of modern manufacturing, is undergoing a digital revolution. Digital twins, virtual replicas of physical assets and processes, are emerging as a powerful tool to optimize operations, enhance safety, and unlock new business models. This paper explores the application of digital twins in the chemical process industry (CPI) from an operational perspective [1].

Firstly, we present a comprehensive framework for developing an operational digital twin. This framework emphasizes the crucial role of data management, encompassing real-time and historical data from sensors, process control systems, and other sources. Process modeling, utilizing physics-based and data-driven techniques, translates this data into actionable insights. Advanced optimization algorithms leverage these insights to identify the most efficient operating conditions. Production scheduling takes these optimized settings into account to create a dynamic plan that adapts to real-time changes [2].This study proposes how digital twin can be integrated with existing process control system as summaried in Figure 1.

Central to this framework is the concept of advanced process control (APC). We delve into the APC hierarchy, highlighting how digital twins can be used to implement sophisticated control strategies, ensuring process stability and product quality.

Artificial intelligence (AI) plays a critical role throughout the development and deployment of digital twins. Techniques like surrogate modeling and predictive maintenance enable faster and more accurate simulations [3]. Additionally, AI-powered optimization and control algorithms further enhance process performance. This study proposes the steps to develop a digital twin for industrial chemical processes as summaried in Figure 2.

Furthermore, the paper delves into the challenges hindering widespread adoption of digital twins in the chemical industry. Despite rapid growth in practical applications across various sectors, the unique complexities of chemical processes pose significant hurdles to implementation. To address this gap, the study organizes existing research on digital twins, focusing specifically on barriers and enablers within the chemical domain. Through the synthesis of literature and proposal of conceptual models, this paper contributes to a deeper understanding of the factors shaping successful digital twin implementation strategies.

However, realizing the full potential of digital twins requires continuous research and development. The paper concludes by outlining key areas for future research, including the integration between virtual entities and the development of standardized frameworks for wider adoption across the CPI.

Keyword(s): Arificial Intelligence, Digital Twin, Advance Process Control, Process Industry

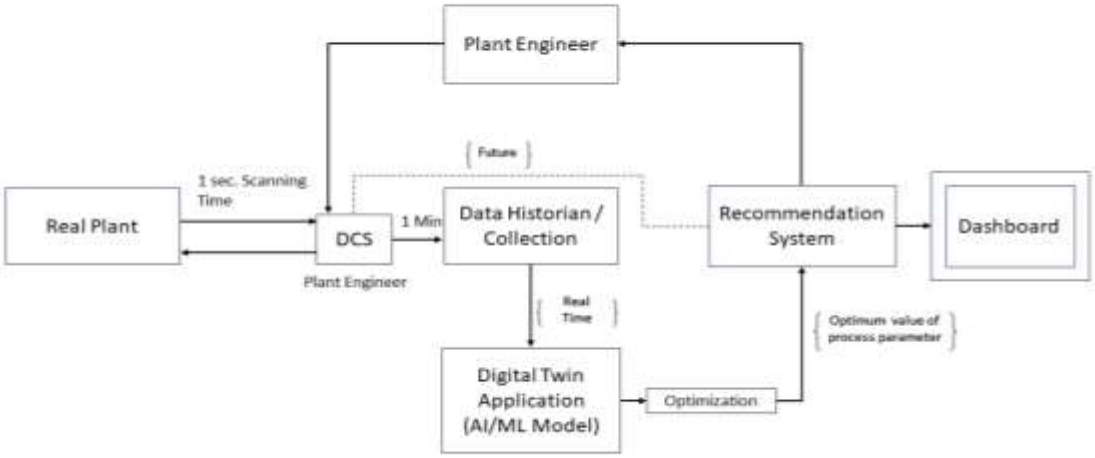


Figure 1: Digital Twin Application in Chemical Process Industry

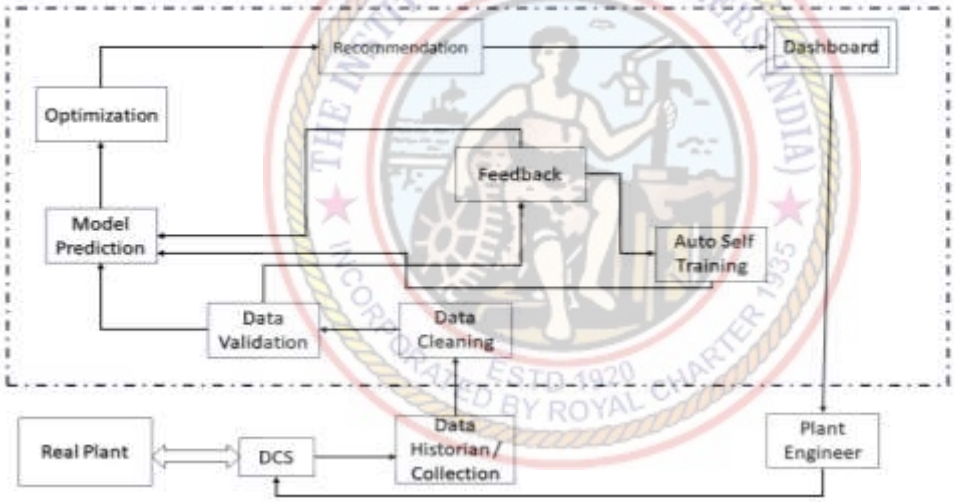


Figure 2: Digital Twin Architecture in Chemical Process Industry

References

1. Perno, M., Hvam, L., & Haug, A. (2022). Implementation of digital twins in the process industry: A systematic literature review of enablers and barriers. Computers in Industry, 134, 103558.
2. Örs, E., Schmidt, R., Mighani, M., & Shalaby, M. (2020, June). A conceptual framework for AI-based operational digital twin in chemical process engineering. In 2020 IEEE international conference on engineering, technology and innovation (ICE/ITMC) (pp. 1-8). IEEE.
1. Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. CIRP journal of manufacturing science and technology, 29, 36-52.

Machine Learning-Driven Optimization for Process Enhancement: A Case Study in Polyethylene Transition Processes

Bitopama Modak[1], Fahim Ahmed[1], Sandip Kumar Lahiri[2*]

[1]Department of Chemical Engineering, Jadavpur University, Jadavpur, Kolkata
{bitopama.modak, fahimahmed0576}@gmail.com
[2]Chamical Engineering, National Institute of Technology, Durgapur, India
sklahiri.che@nitdgp.ac.in

Extended Abstract

The Melt Flow Index (MFI) is pivotal in the polymer industry for assessing molten polymer flow properties, aiding in production optimization, product consistency, and overall operational efficiency [1]. Traditional MFI measurement methods involve time-consuming offline laboratory analyses, leading to production inefficiencies and losses during polymer grade transitions. This research leverages Multi Gene Genetic Programming (MGGP), an advanced machine learning technique, to develop a soft sensor for MFI using industry-specific data.

MGGP evolves computer programs through principles akin to biological evolution, starting with a random program population and refining them through selection, crossover, and mutation to identify the most suitable solution. Initially, potential input parameters influencing MFI, such as reactor diameter, temperature, pressure, gas flow rate, and various chemical concentrations, were considered. After feature engineering, the model was streamlined to incorporate four key parameters: reactor temperature, ethylene, butene, and hydrogen concentrations at the reactor inlet, with MFI as the output.

A significant advantage of MGGP is its ability to produce a closed-form equation, easily deployable in a plant's Distributed Control System (DCS). This equation, selected based on maximizing the R-squared value while minimizing complexity, underwent fine-tuning to strike a balance between performance and simplicity. The resulting MGGP model exhibited remarkable predictive accuracy, boasting an R-squared value of 0.99 and a low root mean squared error (RMSE) of 0.0218.

Implementing this soft sensor facilitates real-time MFI monitoring and prediction, negating the need for frequent manual measurements. This advancement enhances process control, responsiveness to process fluctuations, product consistency, and reduces production costs by minimizing material waste based on continuous, precise MFI data. Furthermore, the study investigates leveraging this real-time MFI soft sensor to expedite grade changeovers, crucial transitions between different polymer grades in the production line. Grade changeover operation in the polymer industry refers to the process of transitioning from producing one polymer grade to another within the same production line. Efficient and effective grade changeovers are essential for minimizing downtime, reducing material waste, and ensuring product quality, as it involves adjusting process parameters, and verifying that the new grade meets the required specifications before resuming production.

In polyethylene manufacturing, precise grade transitions require meticulous adjustments. A targeted single optimization genetic algorithm was employed to identify optimal temperature and concentration ratios, achieving significant reductions in transition times. Transition time reductions of up to 95% were observed for specific transitions, such as from grade LL0209 to LL0410 in Linear low-density polyethylene (LLDPE), and up to 99% for transitions like HD5211 to HD6070 in High-Density Polyethylene (HDPE).

Unlike conventional offline methods that prolong transitions due to delayed sample analyses, the MGGP-based MFI soft sensor offers instantaneous, real-time results, enabling engineers to promptly implement corrective actions. This optimization approach not only addresses operational challenges but also yields substantial cost savings, significantly enhancing manufacturing effi-ciency and overall operations in the polymer industry.

Keyword(s): Melt Flow Index (MFI); Linear low-density polyethylene (LLDPE); High-Density Polyethylene(HDPE); Multi Gene Genetic Programming (MGGP); Genetic Algorithm ;Feature Selection.

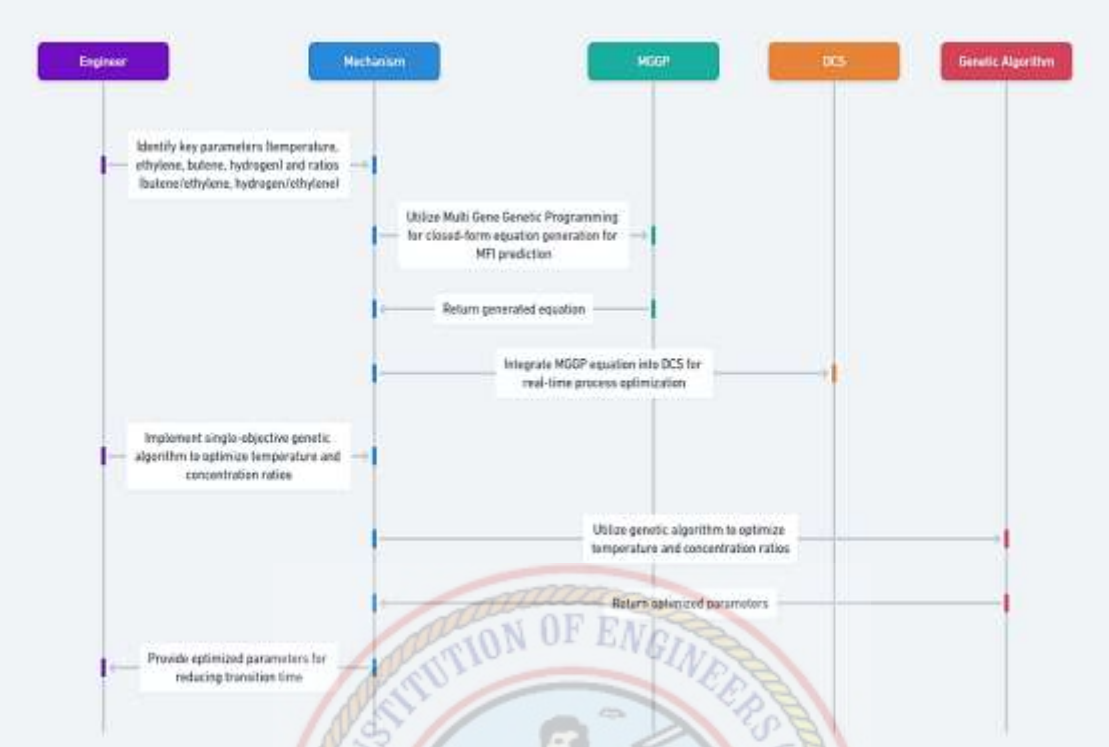


Figure 1: Schematic Diagram of the Study.

References

1. Kiashemshaki, A., Mostoufi, N., Sotudeh-Gharebagh, R., & Pourmahdian, S . Reactor modeling of gas-phase polymerization of ethylene. Chemical Engineering & Technology: Industrial Chemistry-Plant Equipment-Process Engineering-Biotechnology, 27(11), 1227-1232(2004).
2. Zhang, M., & Liu, X. A soft sensor based on adaptive fuzzy neural network and support vector regression for industrial melt index prediction. Chemometrics and Intelligent Laboratory Systems, 126, 83-90(2013).
3. Chen, Jizhong. "Optimal grade transition in continuous polymerization process." Journal of Chemical Industry and Engineering, null (2001).

From Data to Yield: Optimizing Industrial Reactions through Advanced Machine Learning Techniques

Shazeena Hussain[1], Aditi Mahto[1], Sneha Singh[1], Bushra Saad[1], Shruti Tiwari[1] and Sandip Kumar Lahiri[1*]

[1]Chamical Engineering, National Institute of Technology, Durgapur, India
sklahiri.che@nitdgp.ac.in

Extended Abstract

In the sphere of industrial reactions, grappling with the intricacies of reaction kinetics poses a formidable challenge. Nonetheless, the wealth of operational data presents an opportune arena for innovation. This study delves into the realm of machine learning methodologies, seeking to model and optimize the hydromethylation of toluene—a reaction characterized by its intricate kinetics.

The dearth of detailed insight into the mechanisms and kinetics within industrial reactors impedes the development of precise models for predicting and optimizing product yields. This knowledge gap renders the fine-tuning of operational parameters—a crucial aspect directly influencing the efficiency and effectiveness of chemical processes—a daunting task. Consequently, optimizing product yield often devolves into a laborious trial-and-error process, marked by extensive experimentation and escalated costs.

Focusing on a reactor configured for the hydromethylation of toluene under specific temperature and pressure conditions, this investigation scrutinizes the intricate progression wherein toluene reacts with hydrogen to yield benzene and methane.

The input parameter taken into consideration for modelling were the concentrations of toluene, hydrogen and benzene and output of the model is rate of reaction.

Employing various data-driven modelling methodologies, including artificial neural networks (ANN), genetic programming (GP), and Multi-Gene Genetic Programming (MGGP), predictive models are crafted based on empirical data, with the rate of reaction serving as the output parameter.

Through the amalgamation of artificial neural networks (ANN), genetic programming (GP), and Multi-Gene Genetic Programming (MGGP), predictive models are formulated, grounded in empirical data with the reaction rate as the focal output. These models undergo rigorous evaluation, appraised for their elucidation, intricacy, and fidelity to the reaction's underlying phenomenology.

Artificial neural networks (ANN) emerge as a potent tool, adept at deciphering intricate data patterns and guiding decision-making processes. By leveraging backpropagation and optimization algorithms such as gradient descent, ANNs glean insights from labelled datasets, thereby facilitating the development of predictive models. In this study, ANN models exhibit exceptional accuracy, boasting an R^2 value of 0.999 and an RMSE of 0.001.

Inspired by biological evolution, Genetic Programming (GP) orchestrates the generation of computer programs aimed at tackling complex problems through processes like selection and crossover. Multi-Gene Genetic Programming (MGGP) extends this paradigm, encoding multiple solution components to navigate the solution space with heightened efficiency. Herein, both GP and MGGP models showcase commendable accuracy, with GP recording an R^2 value of 0.99 and an RMSE of 0.002, and MGGP mirroring similar performance.

A notable advantage of the developed MGGP model over ANN lies in its provision of a closed-loop equation deployable within plant control loops (DCS), furnishing industrial settings with enhanced control and optimization capabilities. Further distinguishing itself through its transparency, the MGGP model offers heightened explainability compared to the black box nature of ANN models.

Harnessing the developed models, genetic algorithms (GA) are enlisted to optimize reactor process parameters, culminating in a substantial upsurge in product yield. The hybrid MGGP-GA algorithm illustrates its mettle,

boasting 35% enhancement in product yield, thereby fortifying the potential for augmented plant profitability.

This research underscores the prowess of machine learning methodologies in surmounting the challenges posed by poorly understood reaction kinetics, thereby charting a course towards heightened efficiency and profitability in industrial processes.

Keyword(s): ANN; GP; MGGP ; GA

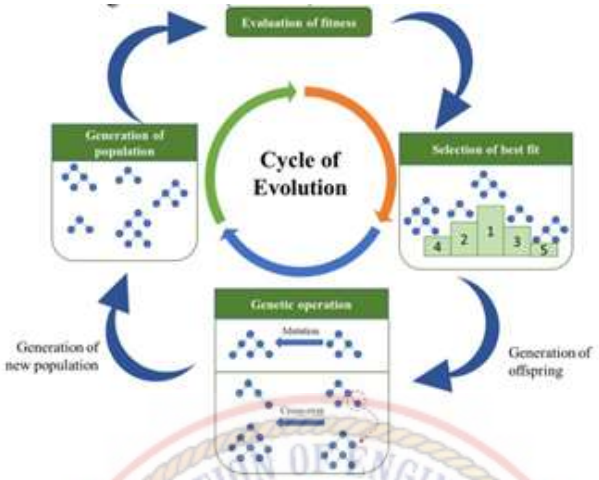


Figure 1: GP at a glance

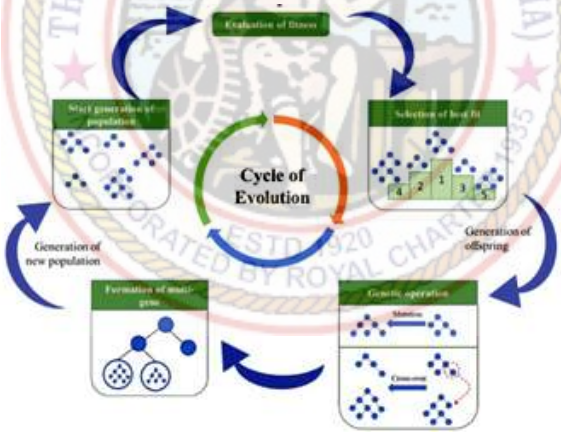


Figure 2: MGGP at a glance.

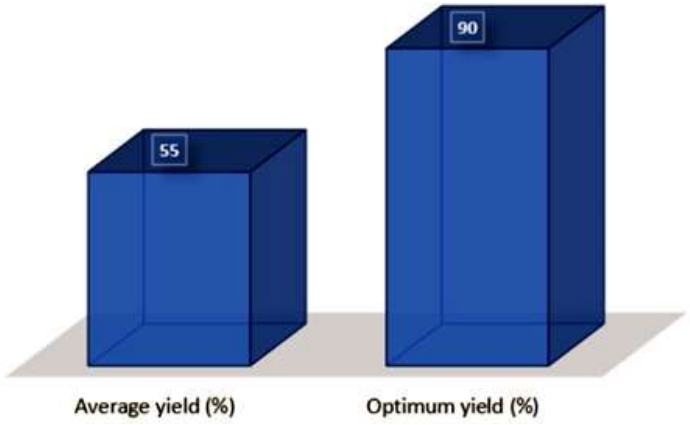


Figure 3: Average yield v/s Optimum yield

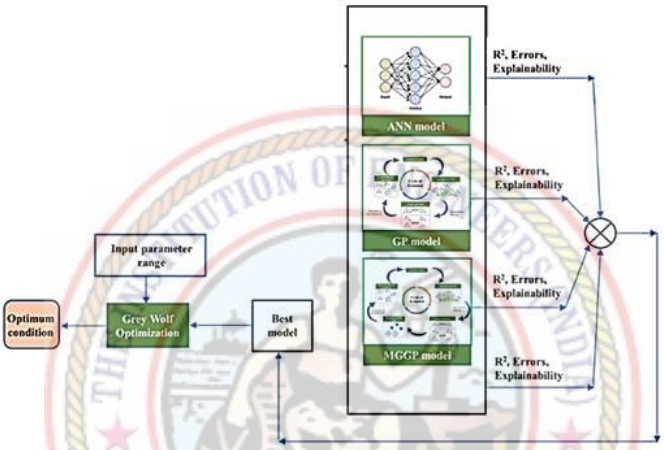


Figure 4: Schematic diagram of the study.

References

1. Cheema JJ, Sankpal NV, Tambe SS, Kulkarni BD. Genetic programming assisted stochastic optimization strategies for optimization of glucose to gluconic acid fermentation. Biotechnology progress. 2002;18(6):1356-65.
2. Koza JR. On the programming of computers by means of natural selection. Genetic programming. 1992.
3. Searson DP, Leahy DE, Willis MJ. GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. InProceedings of the International multiconference of engineers and computer scientists 2010 Mar 17 (Vol. 1, pp. 77-80). Citeseer.
4. Shin Y, Smith R, Hwang S. Development of model predictive control system using an artificial neural network: A case study with a distillation column. Journal of Cleaner Production. 2020 Dec 20;277:124124.
5. Searson DP, Leahy DE, Willis MJ. GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. InProceedings of the International multiconference of engineers and computer scientists 2010 Mar 17 (Vol. 1, pp. 77-80). Citeseer.

Optimising Profitability and Environmental Impact in Chemical Plant Operations: A Machine Learning Approach for Commercial Ethylene Oxide Reactors

Soumi Ghosh[1], Milita Ghosh[1], Saheli Saren[1], Sandip Kumar Lahiri[1*]

[1]Department of Chemical Engineering, National Institute of Technology Durgapur,
Mahatma Gandhi Rd, A-Zone, Durgapur, West Bengal 713209
sklahiri.che@nitdgp.ac.in

Extended Abstract

The challenging task of balancing environmental concerns with profitability optimization is faced by chemical plant operations, particularly in ethylene oxide reactors where dependable models are hard to come by. The limitations of optimization stem from the inability of conventional techniques to fully understand reactor complexities. This study proposes the use of machine learning to develop a robust model for ethylene oxide reactors. In this work, a range of machine learning techniques, such as decision trees, neural networks, and regression, are used to establish a data-driven model correlation between catalyst selectivity and operational parameters. In trying to balance the effects on the environment and profitability, the methodology offers insights into complex chemical processes.

The final model allows us to weigh both profitability and environmental impact, which are crucial but often at odds. Eco-indicator 99 helps us measure the environmental effects, including those on human health and nature. Then, we use the Multi-Objective Genetic Algorithm (MOGA) to create a Pareto diagram, comparing profit with Eco-99.This gives plant engineers valuable insights to fine-tune operating conditions, ensuring a good balance between profit and environmental well-being. With the help of MOGA, the model helps us navigate the conflicting goals, guiding us toward the best possible conditions. Eco-indicator 99 plays a key role in quantifying environmental impact, ensuring we consider both profit and environmental factors.

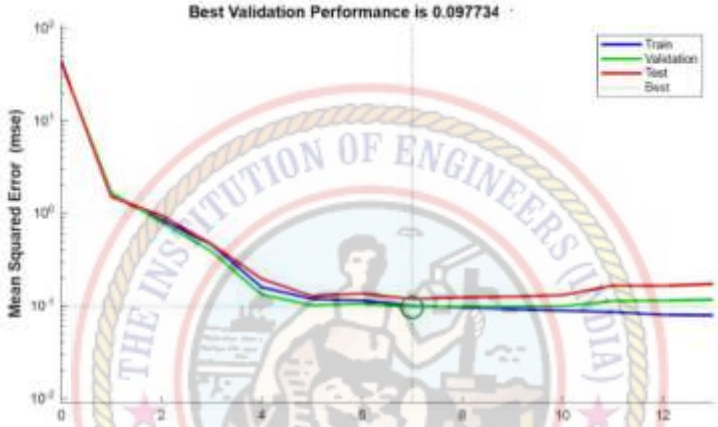
This study presents a fresh approach for chemical plants to boost profits while also making significant strides towards environmental sustainability. Through the use of sophisticated machine learning (ML) techniques, a reliable model was crafted for a commercial ethylene oxide reactor, tackling the challenge of balancing economic success with environmental responsibility.

Different ML models, such as decision trees, artificial neural networks, and support vector regression, were employed to develop a data-driven model linking catalyst selectivity with operational factors. Integrated with Eco-indicator 99, the resulting model effectively distinguishes between profit and environmental impact, guiding the creation of a Pareto diagram using the Multi-Objective Genetic Algorithm (MOGA). This diagram aids plant engineers in pinpointing optimal operating conditions that strike a balance between profit and environmental considerations. Remarkably, for a 150 kTA ethylene oxide plant, this approach led to an estimated decrease of 7548 MT/year in carbon dioxide emissions while maximizing profit.

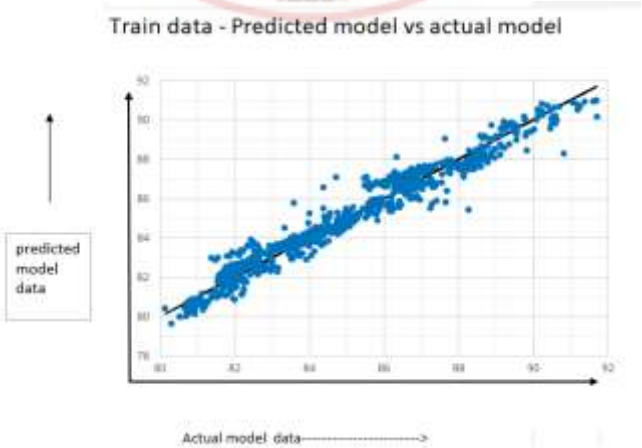
By incorporating AI/ML-based modelling and optimization techniques, this study makes a valuable contribution to sustainable chemical plant operations, offering a versatile methodology applicable to various chemical plants. This innovative approach highlights the significance of harnessing machine learning and optimization strategies to address both economic and environmental goals in chemical plant operations. Through the integration of advanced methodologies, this research provides valuable insights into enhancing profitability while minimizing environmental risks, thereby advancing discussions on sustainable chemical plant operations.

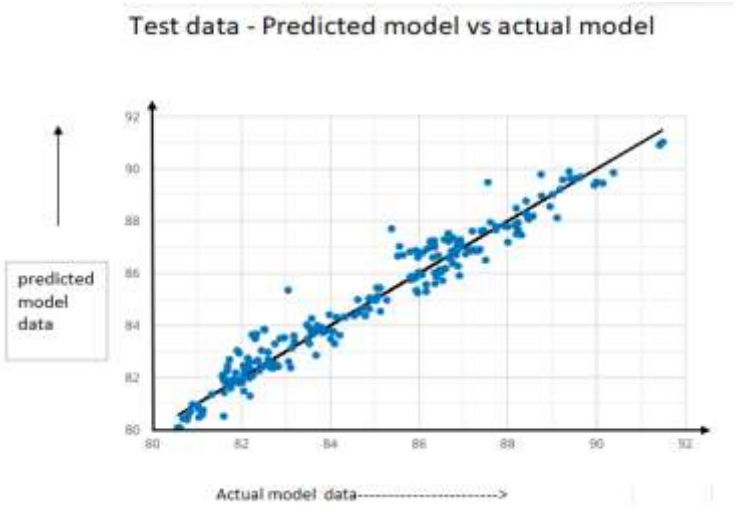
Genetic Programming (GP) and its Application in Multi-Objective Optimization in EO Reactor.

Genetic Programming (GP) is an artificial intelligence technique that evolves programs from a population of random programs to solve specific tasks. It mimics natural genetic processes, with operations including selection of fit programs (crossover), replication, and mutation based on a predefined fitness measure. Crossover swaps parts of selected programs to create new offspring, while some programs are copied to the next generation. Mutation randomly changes parts of a program. This process iterates, with each generation typically more fit than the previous. The best program from each generation often improves, and evolution stops when a program reaches a desired proficiency. One of the key applications of genetic programming is in multi-objective optimization, where the goal is to find solutions that balance multiple conflicting objectives. The Pareto solution, or Pareto front, represents a set of solutions where no single solution is superior in all objectives, showcasing the trade-offs inherent in the problem. By employing genetic programming, researchers and engineers can efficiently explore this Pareto front, gaining insights into the optimal solutions that best satisfy the given objectives. In our experiment, we partitioned the data into training and test sets, with the test set comprising approximately 15% of the total data. After applying Genetic Programming, we obtained 12 different models or Pareto solutions. To select the best model, we prioritized lower complexity and higher accuracy, focusing on an R-squared value close to 1. Subsequently, we conducted multiple iterations in MATLAB to fine-tune the model. The chosen model exhibited strong performance on both the training and test data sets.



For the training data, the model achieved an impressive R-squared value of 0.96144, The Root Mean Square Error (RMSE) for the training data was 0.5415, suggesting that, on average, the model's predictions were 0.5415 units away from the actual values. Similarly, for the test data, the model performed well, achieving an R-squared value of 0.9518 and an RMSE of 0.578. These results demonstrate that the model is robust and effective in capturing the underlying patterns in the data, indicating its potential for practical applications.





Artificial Neural Networks: Revolutionizing EO Reactor Optimization

Artificial Neural Networks (ANNs) are computational models inspired by the structure and function of the human brain. They consist of interconnected nodes called neurons, organized into layers including input, hidden, and output layers. ANNs process information by passing signals through these layers, with each neuron applying a weighted sum of inputs and an activation function to produce an output. Through a process known as training, ANNs adjust these weights based on the input data to improve their performance on tasks such as pattern recognition, classification, and prediction. ANNs have found widespread applications in various fields, including image and speech recognition, natural language processing, medical diagnosis, and financial forecasting. Their ability to handle complex, non-linear relationships in data makes them particularly suitable for tasks involving large datasets and intricate patterns.

Then, we used the command >>nnstart in the command window to open the Neural Network Start GUI. (Neural Network Start GUI is a user-friendly interface in MATLAB that simplifies the initialization and training of neural networks.)From there, we accessed the Fitting app to create and train our ANN using the data. After training the ANN with our data, we obtained the Root Mean Square Error (RMSE) and Mean Squared Error (MSE) values. In our case, the RMSE was 0.98 and the MSE was 0.021. These values indicate the accuracy of the ANN in predicting the output based on the input data.

Genetic Programming (GP) vs Artificial Neural Networks (ANNs) in Optimization

Comparison between Genetic Programming (GP) and Artificial Neural Networks (ANNs) with a focus on ANN's superiority:

Genetic Programming (GP) evolves programs from a population of random programs using genetic operators like crossover and mutation, aiming to evolve executable program solutions. In contrast, ANNs are computational models inspired by the human brain's structure, learning from data by adjusting neuron connection weights.

GP represents solutions as programs, often complex and challenging to interpret, while ANNs represent solutions as interconnected nodes (neurons), visualised in layers, making them more intuitive.

GP suits problems where executable program representation is ideal, like symbolic regression or control systems, while ANNs excel in pattern recognition, classification, and prediction tasks, such as image and speech recognition.

GP and ANNs can achieve high performance, but the choice depends on the problem's nature. ANNs are often preferred for optimization due to their ability to handle complex, non-linear relationships in data, making them versatile and powerful tools in optimization and modelling.

ANN

GP

Artificial Nural Network

	R-Squared	RMSE
TRAIN DATA	0.996	0.096
TEST DATA	0.976	0.094

Genetic programming

	R-Squared	RMSE
TRAIN DATA	0.964	0.541
TEST DATA	0.967	0.578

A Case Study Demonstrating ANN Superiority

While GP and ANNs can achieve high performance, the choice between them depends on the problem's nature. ANNs are often preferred for optimization due to their ability to handle complex, non-linear relationships, making them versatile for large datasets and intricate patterns. In a specific modelling process, although GP may provide multiple equations as potential models, the ANN model is considered the best fit. This conclusion is based on the ANN's high R-squared value and low Root Mean Square Error (RMSE), indicating its superior performance compared to GP. The high R-squared value suggests that the ANN can explain a large portion of the variance in the data, indicating a good fit. Additionally, the low RMSE indicates that the ANN's predictions are close to the actual values, demonstrating high accuracy. Therefore, the ANN's superior performance in capturing underlying patterns and making accurate predictions makes it the preferred model for this particular modelling process.

Additionally, employing ANN for modelling offers several advantages, including improved accuracy, scalability, and adaptability to dynamic operational conditions. Real-time monitoring and predictive maintenance are made possible by the use of ANN, which increases operational effectiveness and decreases downtime. Chemical plants can successfully handle real-world problems such as optimization, product quality control, and equipment failure prediction—hold promise for resolving issues and advancing sustainability.

Keyword(s): Machine learning models, Multi-Objective Genetic Algorithm, Pareto front, Ethylene oxide reactor, Eco-indicator 99, Environmental effects, Artificial Neural Network, Optimizing profit and emission hazards

References

1. Lahiri, S. K., & Khalfe, N. (2009). Process modelling and optimization of industrial ethylene oxide reactor by integrating support vector regression and genetic algorithm. *The Canadian Journal of Chemical Engineering*, *87*(1), 118-128.

2. Rahimpour, M. R., Shayanmehr, M., & Nazari, M. (2011). Modeling and simulation of an industrial ethylene oxide (EO) reactor using artificial neural networks (ANN). *Industrial & engineering chemistry research*, *50*(10), 6044-6052.
6. Chowdhury, S., Lahiri, S. K., Hens, A., & Katiyar, S. (2022). Performance enhancement of commercial ethylene oxide reactor by artificial intelligence approach. *International Journal of Chemical Reactor* Engineering, *20*(2), 237-250.
3. Khalfe, N. M. (2010). Modelling of Commercial Ethylene Oxide Reactor: A Hybrid Approach by Artificial Neural Network & Differential Evolution.
4. Picos, A., & Peralta-Hernández, J. M. (2018). Genetic algorithm and artificial neural network model for prediction of discolouration dye from an electro-oxidation process in a press-type reactor. Water Science and Technology, 78(4), 925-935.

Exploration of antibiotic sensitivity in breeder chickens and antimicrobial potential of poultry-associated microorganisms: a statistical comparative study of six different drugs

Anindya Singha Thakur[1], Sanjukta Dasgupta[1], Bimal Das[2], Deepshikha Datta[3]

[1]Department of Biotechnology, Brainware University, Barasat, Kolkata, West Bengal-700125
[2]Department of Chemical Engineering, National Institute of Technology Durgapur, West Bengal, India - 713209
[3]Department of Chemistry, Brainware University, Barasat, Kolkata, West Bengal-700125

Extended Abstract

The poultry industry plays a pivotal role in meeting global food demands, but it faces formidable challenges due to infectious diseases caused by a myriad of pathogens. One of the most concerning issues is the rampant emergence of antibiotic-resistant bacteria, largely attributed to the excessive use of antibiotics in poultry production (1). This study delves into the effects of six different drugs–amoxicillin, gentamicin, cephalexin, co-trimoxazole, oxytetracycline, and ciprofloxacin on breeder chickens, with a particular focus on understanding their impact and the ensuing implications for combating antibiotic resistance in poultry farming.

Through statistical analysis using GraphPad Prism 7 software, it was discerned that breeder chickens exhibited the highest sensitivity to amoxicillin among the tested drugs (2). Further elucidation of amoxicillin's properties using the PubChem tool revealed critical characteristics such as its molecular weight of 365 g/mol and heavy atom count of 24. This comprehensive understanding enhances our insight into the drug's efficacy and its potential implications for poultry health, thereby contributing to informed decision-making regarding antibiotic usage in poultry production.

Moreover, this study embarks on the isolation and identification of diverse microorganisms sourced from various niches within poultry birds, including the gut, skin, and litter. These isolates underwent rigorous screening to assess their ability to inhibit the growth of select poultry pathogens such as Salmonella spp., Escherichia coli, and Campylobacter jejuni. Employing agar diffusion or broth microdilution assays, researchers evaluated the antimicrobial potential of these isolates, with a primary objective of identifying novel agents capable of combating antibiotic-resistant bacteria.

Statistical analyses generated compelling insights into the antimicrobial activity of the isolated microorganisms, with a notable emphasis on their efficacy against Escherichia coli. This discovery holds significant promise for the development of targeted interventions aimed at addressing prevalent bacterial pathogens in poultry production systems. By unraveling the antimicrobial capabilities of poultry-associated microorganisms, this study paves the way for innovative strategies to counter antibiotic resistance and promote sustainable practices within the poultry industry.

The implications of these findings transcend the confines of poultry farming, offering a paradigm shift in antimicrobial discovery and intervention strategies. By harnessing the diverse microbial communities inherent in poultry birds, researchers can potentially tap into a vast reservoir of untapped resources for combating antibiotic resistance. This approach aligns with the overarching imperative of developing forward-thinking solutions to confront the escalating threat posed by multidrug-resistant bacteria. Furthermore, this research underscores the imperative of adopting holistic approaches to safeguard both animal and public health. By elucidating the intricate interplay between antibiotic use, microbial ecology, and pathogen dynamics within poultry production systems, this study furnishes invaluable insights into the multifaceted challenges confronting modern agriculture. Through interdisciplinary collaboration and concerted efforts, stakeholders can strive towards implementing effective strategies to mitigate antibiotic resistance while ensuring the sustainability and resilience of poultry farming practices.

In conclusion, this study represents a significant milestone in advancing our understanding of antibiotic resistance dynamics within the poultry industry. By unraveling the efficacy of various antibiotics and exploring novel antimicrobial agents derived from poultry-associated microorganisms, researchers have laid a solid foundation for future interventions aimed at mitigating the threat of antibiotic-resistant bacteria. These findings underscore the imperative of adopting a multifaceted approach to address the complex challenges facing poultry production, with the ultimate goal of safeguarding public health and fostering sustainability in agriculture.

Keyword(s): Poultry industry, statistical analysis, microorganisms, computational biology

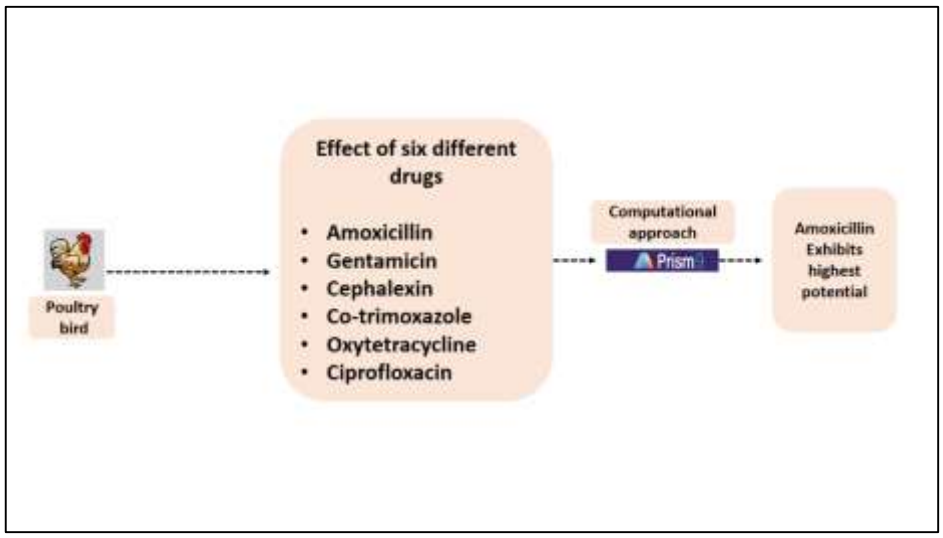


Figure 1: Workflow of the Proposed Study

References

1. Bobkov M, Zbinden P. Occurrence of Veterinary Drug Residues in Poultry and Products Thereof. A Review. Chimia (Aarau). 2018 Oct 31;72(10):707-712. doi: 10.2533/chimia.2018.707. PMID: 30376920.
2. Mitteer DR, Greer BD. Using GraphPad Prism's Heat Maps for Efficient, Fine-Grained Analyses of Single-Case Data. Behav Anal Pract. 2022 Jan 3;15(2):505-514. doi: 10.1007/s40617-021-00664-7. Erratum in: Behav Anal Pract. 2022 Jan 18;15(2):515. PMID: 35692516; PMCID: PMC9120324.

Optimization of Sensitivity of Nano silica Modified Glassy Carbon Electrode towards Arsenic (III) ions in Ground Water through Parametric Modulation

Sampurna Santra1,2, Koushik Sardar2, Deepshikha Datta3*, Bimal Das2*

1Department of Electronics and Communication Engineering, Murshidabad College of Engineering & Technology, Berhampore, India

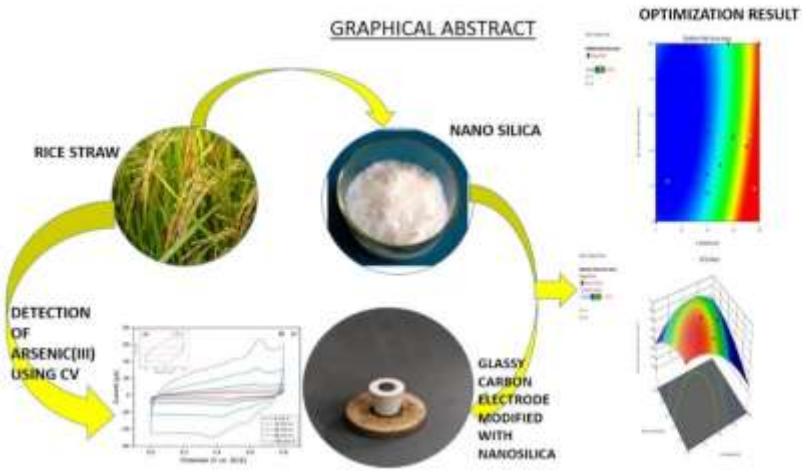
2Department of Chemical Engineering, National Institute of Technology Durgapur, West Bengal, India
bdas.che@nitdgp.ac.in

3Department of Chemistry, Brainware University, Barasat, Kolkata, West Bengal, India
deepa.datta81@gmail.com

Extended Abstract

Agro waste has a significant contribution in the production of value added product. The present work emphasize on the use of raw rice straw for the fabrication of electrodes using precipitation process. The obtained nano silica had a large surface area of 189.64 m2/g and a three-dimensional structure consisting of strongly linked open spaces, supporting its potential for application in the production of electrodes. Chemical processing was used to create nanosilca (200 nm), which was then further reduced in size using a cryomill. Using a modified glassy carbon electrode and a bare glassy carbon electrode as a comparison, the cyclic voltammetry (CV) technique was utilized to find harmful Arsenic(III) ions in the ground water. When compared to the bare electrode, the modified glassy carbon electrode performs better, exhibiting distinct oxidation/reduction peaks. The glassy carbon electrode modified with silica nanoparticles at varying scan speeds between 5 and 100 mv/s in the potential range of 0.00 V to 0.8 V SCE in 1 mM arsenic (III) chloride solution produced the results of cyclic voltammetry (CV). At the scan rate of 100 Mv/s, oxidation peaks were seen at 0.65 V, while reduction peaks were noticed at 0.36 V. Further, process parameter optimization has been carried out using a central composite design (CCD) strategy. This strategy takes into account the interaction between quantitative and process variables. The model fit statistics of the ANOVA (Analysis of variance) study suggested that the model corresponds to the quadratic model in nature and the coefficient of variance (R2) value for both the responses are 0.9834 (for oxidation peak) and 0.9754 (for reduction peak), which indicates the model accuracy. Thus the developed electrode produced from agro waste has a significant role in the detection of arsenic content in waste water.

Keyword(s): Nanosilica; electrode; arsenic; glassy carbon electrode; oxidation peak; ANOVA; reduction peak



References

1. Carmona, V. B., Oliveira, R. M., Silva, W. T. L., Mattoso, L. H. C., & Marconcini, J. M.: Nanosilica from rice husk: extraction and characterization. Industrial Crops and Products, 43, 291-296 (2013)
2. Reddy, M., Shetti, N., Nayak, D., Malode, S., & Muddapur, U.: Electrochemical oxidation of food dye at nanosilica modified carbon electrode. Materials Today Proceedings, 18, 798-805 (2019)
3. Babar, N. U. A., Joya, K. S., Tayyab, M. A., Ashiq, M. N., & Sohail, M.: Highly sensitive and selective detection of arsenic using electrogenerated nanotextured gold assemblage. ACS omega, 4(9), 13645-13657 (2019)

Parametric optimization for fixture design in the field of resistance spot welding in sheet metal through central composite design approach towards the application of metal welding.

Ishan Gupta[1], Sumit Mahato[1], Debayan Bhowmick[2], Bimal Das[1*], Deepshikha Dutta[3*]

[1]Department of Chemical Engineering, National Institute of Technology Durgapur, West Bengal, India
bdas.che@nitdgp.ac.in
[2]Department of Mechanical Engineering, Brainware University, Barasat, Kolkata, West Bengal, India
[3]Department of Chemistry, Brainware University, Barasat, Kolkata, West Bengal, India
deepa.datta81@gmail.com

Extended Abstract

Due to its smallest thickness variation at the time of assembly, resistance spot welding is a technology that is frequently employed in industry to join sheet metal. Fixturing is essential for increasing the weld efficiency of sheet assembly systems. Around 73% of the variation issues from pre-production to production were caused by fixture-based issues, according to a survey conducted by American automotive weld quality of the sheet metal assembly process. This work presents a novel fixture design method for the sheet metal spot welding process using Unigraphics CAD Software. It focuses on how the position of the fixture affects the dimensional quality of sheet metal components by taking the locator's position, welding force, and clamping force into account. Research Surface Methodology (RSM) is used with a hybrid optimization technique to find the best position and clamping points for achieving the least amount of distortion.

Keyword(s): Resistance Spot welding, Sheet metal, Fixturing layout, Optimization, Deformation.

References

1. B. Li, H. Tang, X. Yang, H. Wang, Quality design of fixture planning for sheet metal assembly, Int. J. Adv. Manuf. Technol. 32 (7-8) (2007) 690–697.
2. B. Li, B.W. Shiu, K.J. Lau, Principle and Simulation of Fixture Configuration Design for Sheet Metal Assembly with Laser Welding, Part 1: Finite-Element Modelling and a Prediction and Correction Method, Int. J. Adv. Manufact. Technol. 18 (4) (2001) 266–275.
3. G. Prabhaharan, K.P. Padmanaban, R. Krishnakumar, Machining fixture layout optimization using FEM and evolutionary techniques, Int. J. Adv. Manuf. Technol. 32 (11-12) (2007) 1090–1103.
4. K.P. Padmanaban, K.P. Arulshri, G. Prabhakaran, Machining fixture layout design using ant colony algorithm based continuous optimization method, Int. J. Adv. Manuf. Technol. 45 (9-10) (2009) 922–934.
5. N. Siva Shanmugam, G. Buvanashekaran, K. Sankaranarayanasamy, Some studies on weld bead geometries for laser spot welding process using finite element analysis, Mater. Des. 34 (2012) 412–426.
6. B. Li, B.W. Shui, K.J. Lau, Fixture Configuration Design for Sheet Metal Assembly with Laser Welding: A Case Study, Int. J. Adv. Manufact. Technol. 19 (7) (2002) 501–509.
7. B. Li, B.W. Shiu, K.J. Lau, Weld Pattern Design for Sheet Metal Laser Welding Considering Fixturing Quality, Int. J. Manufact. Technol. 19 (6) (2002) 418–425.
8. W. Zhu, H. Wang, X. Lai, Z. Lin, Y. Chen, A New Method for Fitting Car Doors Allowing for the Thermal Effects of Welding, Proceed. Institut. Mechan. Eng. , Part D: J. Automob. Eng. 220 (7) (2006) 891–900.
9. W. Cai, Fixture optimization for sheet panel assembly considering welding gun variations, Proceed. Institut. Mechan. Eng., Part C: J. Mechan. Eng. Sci. 222 (2) (2008) 235–246.
10. J.A. Camelio, S.J. Hu, D. Ceglarek, Impact of fixture design on sheet metal assembly variation, J. Manuf. Syst. 23 (3) (2004 Jan 1) 182–193.
11. N.I. Afzeri, Hybrid optimization of pin type fixture configuration for free form workpiece, Niruth Prombutr, Prayoot Akkaraaektharin. 2 (3) (2008) 32.

Analyze the sensing performance of passivated (Al_xGa_{1-x})2O3/Ga2O3 Schottky diode gas sensor with catalytic metals

Shefali Mamataj[1], Deepshikha Datta[2],Bhaskar Roy[3], Bimal Das[4], Ritam Dutta[5], Sourav Karmakar[4]

[1]Department of Chemical Engineering& ECE, NIT Durgapur & MCET, Berhampore, West Bengal, India
[2]Department of Chemistry, Brainware University, West Bengal, India
[3]Department of AIML, Asansol Engineering College, West Bengal, India
[4]Department of Chemical Engineering, NIT Durgapur, West Bengal, India
[5]Department of CSE, Poornima University, Rajasthan, India
bimal.das@che.nitdgp.ac.in
{deepa.datta81, bhaskar.ec11, ritamdutta1986, debsuvojit12}@gmail.com

Extended Abstract

Gallium oxide (Ga_2O_3) has emerged as an optimistic substance for applications like gas sensors due to its wide bandgap, high electron mobility, excellent thermal stability, and chemical resistance. During the last decade, Ga_2O_3 has found major applications in power electronics, ultraviolet photodetectors, radiation detectors, as well as gas sensors. Among various Ga_2O_3-based gas sensors, Schottky diodes have demonstrated superior sensing performance for a wide range of gases, including hydrogen, oxygen, nitrogen oxides, and volatile organic compounds. The effectiveness of highly significant technologicalfunctionality in enhancement-mode high voltage rectifiers and other technologically significant effects of the Metal-Oxide Field Compared to SiC or GaN, β-Ga_2O_3 has a higher critical electric field, which is advantageous for transistors. Out of the five polymorphs of Ga_2O_3, the β-polymorph is the one that has been studied and utilized the most, and in most cases, it is the stable form. So, Schottky rectifiers are a desirable use for β-Ga_2O_3, as they have a fast-switching speed that is critical for increasing the inductive motor controllers' efficiency, in addition to their low turn-on voltages as compared to p-n junction rectifiers, and power supply.The popular sensor used for the detection of decreasing gases, such as H_2, is a Schottky-type sensor. The Schottky diode's structure and interfacial properties provide spontaneous and piezoelectric polarisation effects. Consequently, a highlydense 2-DEG is produced that is extremely responsive to variations in surface states. Regarding Schottky sensors with different Pt thicknesses, active regions, and operating temperatures, it may be investigated how adding noble metals such as Pt, Pd, etc. affects their capacity to detect gases. The metal's reaction with the semiconductor is one of the most crucial factors in determining Schottky barriers. Therefore, before selecting an electrode metal for sensors, it is essential to understand the reactive and non-reactive metals associated with the particular semiconductor. The surface is often passivated using scandium oxide (Sc_2O_3), and the passivated device has a greater 2DEG density than the un-passivated device. In this proposed work, a physics-based analytical modelling technique has been used to inspect the AlxGa1-x)$_2O_3$/Ga_2O_3 Schottky hetero-structure diode for applications that require extremely linear and sensitive gas sensing. Designing a sensor that has great linearity, a high level of sensitivity, and excellent detectivity at low gas concentration levels and lower temperatures is the main goal of the effort. The surface characteristics and hetero-interface are taken into account here. The primary method for modelling the device is to use 2-DEG's dependence on charges on the surface that are dependent on the passivation layer of Sc_2O_3. Determining the sensitivity and other essential properties of the sensors is largely dependent on the gas sensor electrode. On Ga_2O_3-based sensors, the suggested Schottky contact with catalytic electrodes (Pd, Pt, and other metals) has been investigated. We simulated the sensor model based on Schottky diodes using Silvaco TCAD and were able to get V-I curves for various gas concentrations. The response obtained from the V-I curves demonstrates a reaction of more than 75% under the influence of gas concentration (500 ppm) at a 0.95 V bias voltage. Additionally, we observed that temperature affected the sensor's response for various Schottky contacts at different gas concentrations.

Keyword(s): Schottky diode; 2DEG, Gas sensor; $Al_xGa_{1-x})_2O_3/Ga_2O_3$hetero-structure; Passivation; Modeling.

References

1. Pearton, S., Ren, F., Mastro, M.: Gallium Oxide: Technology, Devices and Applications. Elsevier: Amsterdam, The Netherlands, pp. 439–464 (2019)

2. Higashiwaki, M., Fujita, S.: Gallium Oxide: Materials Properties, Crystal Growth, and Devices. Springer Nature: Cham, Switzerland, pp. 1–11 (2020).
3. Xue, H., He, Q., ian, G.J, Long, S., Pang, T., Liu, M.: An overview of the ultrawide bandgap Ga_2O_3 semiconductor-based Schottky barrier diode for power electronics application. Nanoscale Res. Lett. 2018, 13, 290 (2018)
4. Liu. Z., Li. P., Zhi. Y., Wang. X., Chu. X., Tang. W.: Review of gallium oxide-based field-effect transistors and Schottky barrier diodes. Chin. Phys. B, 28, 017105(2019)
5. Jun, L.W. et al.: AlGaN/GaN Based Diodes for Liquid Sensing. Chin. Phys. Lett., vol. 30, pp. 037301-1-037301-4 (2013).
6. K.H. Baik et al.: AlGaN/GaN heterostructure based Pt nanonetwork Schottky diode with water-blocking layer. Sensors and Actuators B: Chemical 317, 128234 (2020) https://doi.org/10.1016/j.snb.2020.128234, 045010
7. Lee, Sung Pil.: Electrodes for Semiconductor Gas Sensors. Sensors, 17, 683 (2017) doi: 10.3390/s17040683

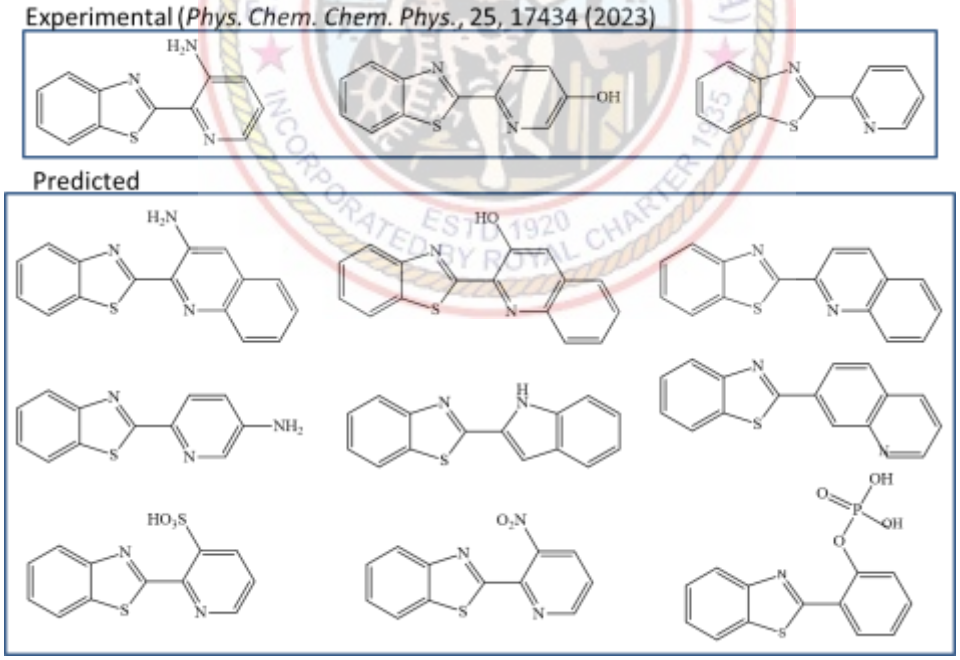
Machine learning based QSAR approach to predict corrosion inhibition performance of benzothiazole derivatives for mild steel in aqueous HCl medium

Pradeep Kumar Rout[1], Dipankar Sukul[1] and Subhas Ghosal[1*]
[1]Department of Chemistry, National Institute of Technology Durgapur, M G Avenue, Durgapur, WB 713209
sghosal.ch@nitdgp.ac.in

Extended Abstract

Heterocyclic organic bases like indole, imidazole, triazole, tetrazole, pyrazole and their derivatives have been known for their good corrosion inhibition properties in acidic aqueous media [1]. Most of these molecules are cheap, have relatively low toxicity and easily bio-degradable. Recent experimental studies from our laboratory [2, 3] shows that various benzothiazole molecules can also be used for similar anti-corrosion activities on mild steel in aqueous HCl medium. However, screening a large number of such molecules for their anti-corrosion activates is experimentally challenging. Therefore, we have used a machine learning based QSAR approach on a set of 19 benzothiazole derivatives to predict new molecules that may have better or similar corrosion inhibition efficiency without performing the experiment.

To overcome the small number of experimentally available data, theoretical calculations using high level Density Functional Theory (DFT) and Molecular Dynamic (MD) simulations on structurally similar benzothiazole molecules were added to the training sets using chemical intuition and other Virtual Sample Generation (VSG) method. Adding theoretical data in the training set helps the Machine Learning algorithm to recognize how the chemical and quantum parameters are correlated with the corrosion inhibition efficiency. The proposed model is verified on a small number of Benzothiazole derivatives and we found this model can make predictions with best accuracy, particularly, for systems with small experimental datasets.



Keyword(s): Corrosion inhibition; Machine learning; QSAR; Benzothiazole.

References

1. Finšgar, M., Petovar, B., Xhanari, K., Maver, U.: The corrosion inhibition of certain azoles on steel in chloride media: electrochemistry and surface analysis. Corros. Sci. 111, 370–381 (2016)
2. Suhasaria, A., Senapati, R., Satpati, S., Ghosal, S., Dey, S., and Sukul, D.: Experimental and theoretical

investigation on the anti-corrosion characteristics of pyridine-substituted benzothiazole derivatives for mild steel in aqueous HCl. Phys. Chem. Chem. Phys., 25, 17434 (2023)

3. Suhasaria, A., Satpati, S., Ghosal, S., Dey, S., and Sukul, D.: Effect of the Heterocyclic Groups on the Anti-corrosion Performance
of Heterocyclic Schiff Bases of Benzothiazole for Mild Steel in 1 M Aqueous HCl. J Bio Tribo Corr, 9, 26 (2023)

GMORF: A GAN-Based Multi-Occlusion Removal Framework for Biometric Verification

Parul Choudhary[1], Phalguni Gupta[1], and Pooja Pathak[1]

[1]Department of Engineering & Applications, GLA University, Chaumuhan, Mathura, U.P,
cparul2605@gmail.com, {pg@,pooja.pathak}@gla.ac.in

Extended Abstract

An occlusion plays a vital role in the accurate analysis of images and the identification of their features in face recognition. The occlusion occurred due to random objects (single or multiple) on the image, which may create a problem of matching the occluded image with the registered image in the database. To the best of the authors' knowledge, all well-known methods have considered the problem of single occlusion in biometrics. To address the problem of multi-level occlusion, a novel feature-oriented GAN-based multi-occlusion removal framework (GMORF) has been proposed in this paper. The proposed framework, GMORF, shown in Figure 1, consists of four components. The first component is to cope with face alignment via a spatial transformer net. The second one produces a qualitative binary map through the QUnet++ module. The next component addresses the problem of multi-occlusion removal technique through the proposed framework to enforce pixel level similarity, and the last component uses ResNet, which enforces feature level similarity from feature space, to determine better inpainting results for verification. For efficient optimization, all four components are implemented in an end-to-end network.

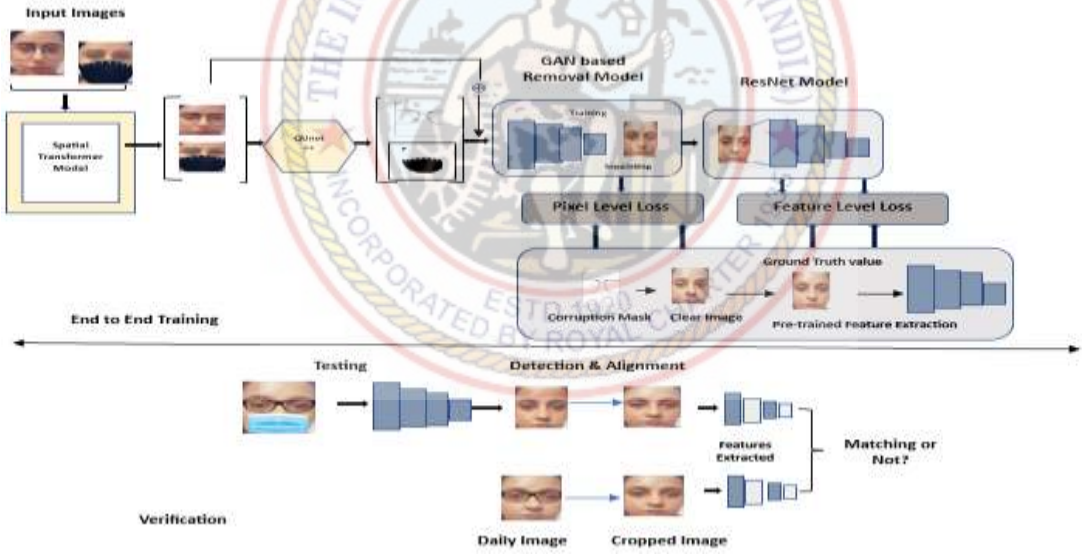


Figure 1: GMORF Frame-

Since generative adversarial network (GAN)-based occlusion removal algorithms have shown exceptional results in tasks involving dense predictions, such as depth prediction, it was inspired to employ GAN as the non-linear function to enhance the performance of blind face inpainting, which is the third component of the proposed GMORF.

The main difference between the existing model and the proposed model is the qualitatively segmented image given to GAN for image inpainting rather than simply applying FCN. The proposed GAN is the combination of a generator, a discriminator, and a perceptual network, as shown in Figure 2.

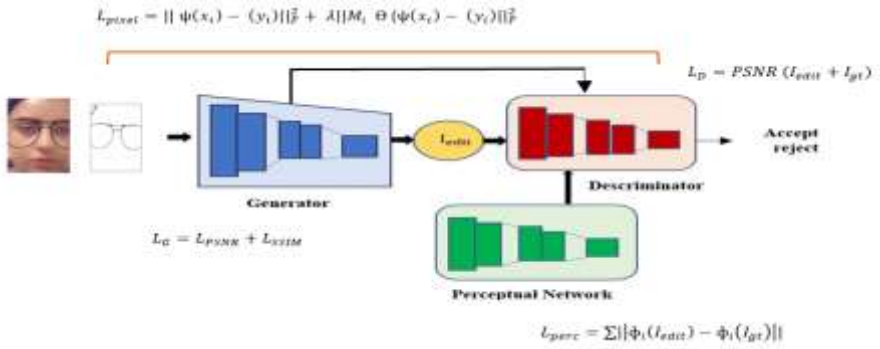
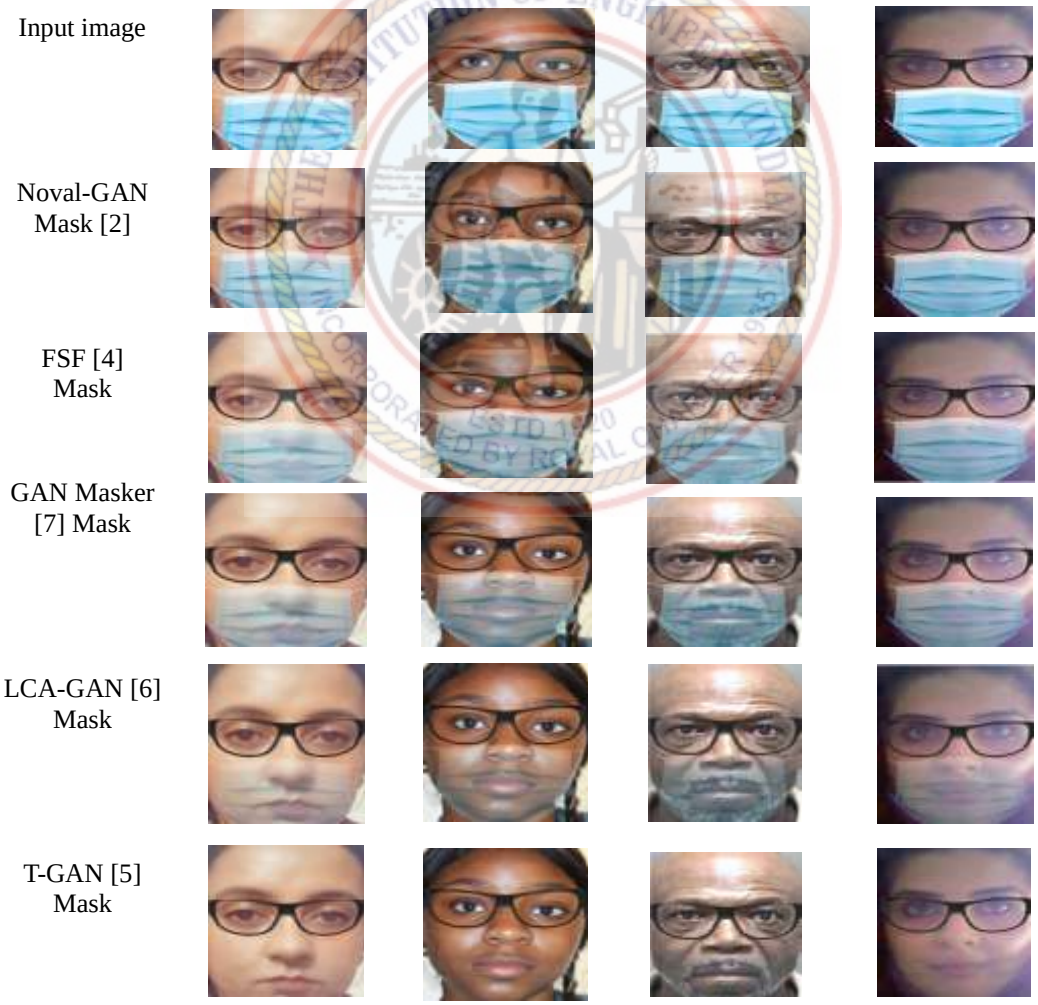


Figure 2: Proposed GAN Multi Occlusion Removal Framework

The performance of the proposed GMORF is visually compared with existing methods for face mask and glass removal: GAN Masker, LCA-GAN, T-GAN, Noval-GAN, Bye Glasses, FSF, and Ours GMORF. The generated de-occluded images with GMORF and other methods are shown in Figure 3 for masks and glass. This comparison clearly shows that GMORF not only produces the best result in terms of a realistic and closer-to-the-ground truth image but is also capable of removing multi-occlusion. These qualitative comparisons demonstrate the improved performance of GMORF in producing de-occluded faces that look attractive and are free of distortions.



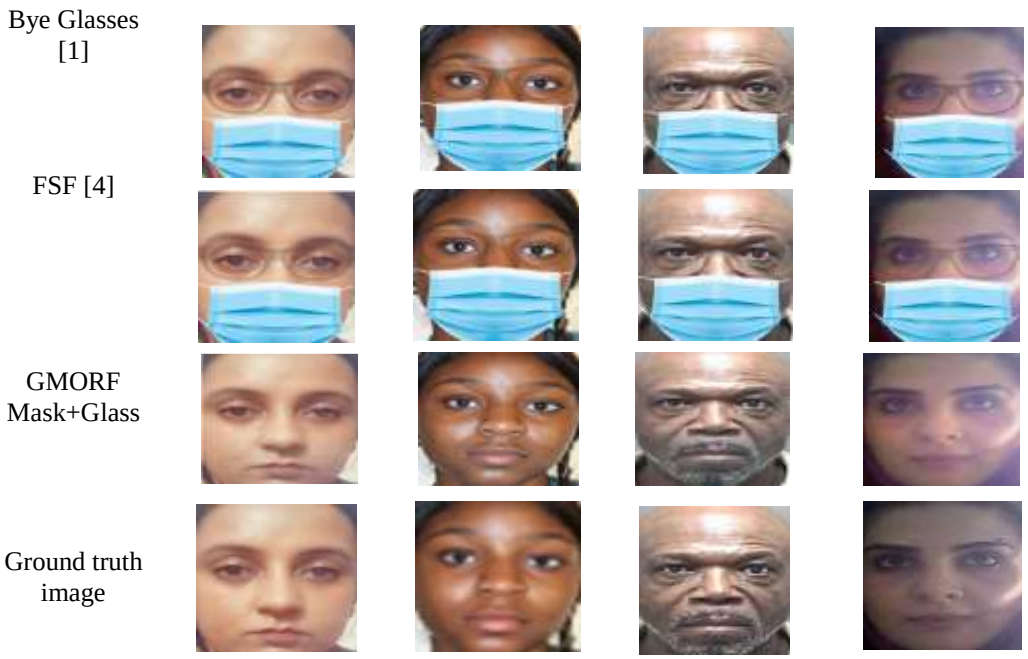


Figure 3: Visual comparison with existing methods using all dataset. Row-1 shows Input image, Row 2-6 interprets removal of Mask and Row 7-8 illustrates removal of glass. Row-7 displays combine removal of mask and glass. Last row shows ground truth images.

The quantitative comparison is done in Table 1 between all methods to ensure an objective evaluation. This comparison involves assessing the outcomes between GMORF and other existing methods by utilizing the quantitative reference metrics SSIM and PSNR. The GMORF achieves higher scores in PSNR and SSIM than other existing methods, indicating similarity to the ground truth image.

Table 1. Quantitative comparison of GMORF with existing methods in terms of PSNR and SSIM. The bold result shows the best performance.

Methods	Object	Coverage	PSNR	SSIM
GAN Masker [7]	Mask	High	30.96	0.95
LCA GAN [6]	Mask	High	19.031	0.696
T-GAN [5]	Mask	High	25.27	0.907
Noval GAN [2]	Mask	High	26.19	0.864
FSF [4]	Mask	High	22.42	---
ByeGlasses GAN [1]	Glass	Medium	21.22	0.91
FSF [4]	Glass	Medium	22.11	---
GMORF	Glass	Medium	20.22	0.87
	Mask	High	32.93	0.98
	Mask+Glass	High	35.62	0.96

An extensive experiment during training considered faces with occlusion (e.g., masks, eyeglasses) and without occlusion and demonstrated the effectiveness of the proposed framework for verification of biometric data.

Keyword(s): Biometrics, face occlusion, face recognition, generative adversarial network, image inpainting.

References

1. Lee, Yu-Hui, and Shang-Hong Lai.: Byeglassesgan: Identity preserving eyeglasses removal for face images. Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16. Springer International Publishing, (2020).
2. N. Ud Din, K. Javed, S. Bae and J. Yi.: A Novel GAN-Based Network for Unmasking of Masked Face in IEEE Access, vol. 8, pp. 44276-44287, 2020, doi: 10.1109/ACCESS., 2977386 (2020).
3. Roychowdhury, Sohini.: QU-net++: Image Quality Detection Framework for Segmentation of Medical 3D Image Stacks. arXiv preprint arXiv:2110.14181 (2021).
4. C. Botezatu, M. Ibsen, C. Rathgeb and C. Busch, "Fun Selfie Filters in Face Recognition: Impact Assessment and Removal," in IEEE Transactions on Biometrics, Behaviour, and Identity Science, vol. 5, no. 1, pp. 91-104, Jan. 2023, doi: 10.1109/TBIOM.2022.3185884 (2023).
5. Nam, S.H.; Kim, Y.H.; Choi, J.; Park, C.; Park, K.R. LCA-GAN: Low-Complexity Attention-Generative Adversarial Network for Age Estimation with Mask-Occluded Facial Images. Mathematics 2023, 11, 1926. https://doi.org/10.3390/math11081926
6. Man, Q.; Cho, Y.-I. Efficient Face Region Occlusion Repair Based on T-GANs. Electronics 2023, 12, 2162. https://doi.org/10.3390/ electronics12102162
7. Mahmoud, M.; Kang, H.-S. GANMasker: A Two-Stage Generative Adversarial Network for High-Quality Face Mask Removal. Sensors 2023, 23, 7094. https:// doi.org/10.3390/s23167094

An Efficient Face Image Quality Assessment Technique

Parul Choudhary[1], Phalguni Gupta[1], and Pooja Pathak[1]

[1]Department of Computer Engineering & Applications, GLA University, Chaumuhan, Mathura, U.P India
{cparul2605@gmail.com, pg@gla.ac.in, pooja.pathak@gla.ac.in}

Extended Abstract

Face recognition systems have gained significant attention in recent years due to their applications in various domains such as security, authentication, and surveillance. The performance of face recognition is affected by analysing facial quality with a single property. However, there are still some attributes, such as occlusion and entropy that are not well studied but play an important role in face recognition. Measuring the quality of images is likely to filter out poor-quality images, which can improve the performance of downstream tasks. This paper presents a comparative study of various face image quality assessment (FIQA) techniques to select the best image for better recognition. Fig. 1. Illustrates an overall FIQA framework divided into three sections. The first section explores the benefits of incorporating and measuring image quality factors such as contrast (C), brightness (B), illumination (I), sharpness (S), focus (F), entropy (E), and occlusion (O), followed by preprocessing steps. In the second section, we determine an efficient fusion technique using clustering to combine all these factors to get a single face image quality index. The correlation is evaluated between the fusion technique and the human observer. An image with the highest correlation scores is assumed to be of good quality. Finally, the biometric face recognition system makes the decision.

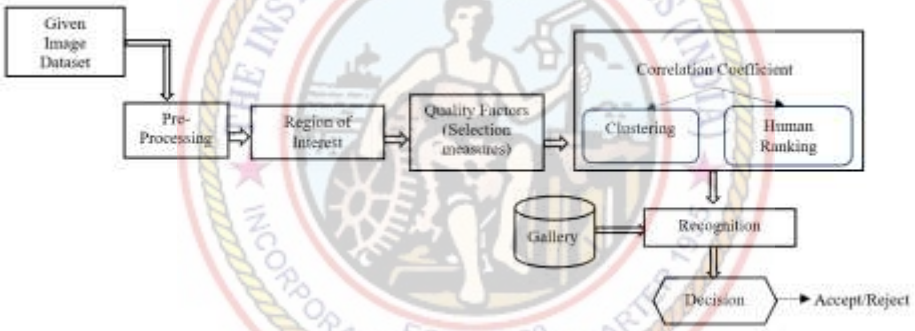


Figure 1: FIQA Framework

The performance of the above-mentioned quality factors is evaluated by choosing 10 random images from a color ferret database, as shown in Table 1. Few changes are made in contrast; brightness, focus, and sharpness are mentioned below, while no changes in other factors such as entropy or occlusion are shown in Table 2. by considering only five random images.

Contrast-images are saturated at low and very high intensities in steps of 10%, i.e., map-ping of low [in] = 0, high [out] = 1, to low [in] = 0.05, high [out] = 0.95.

Brightness: verifying the gamma parameter to change brightness in an image, such as $\gamma = 0.5, 0.6...1.4$. Here, $\gamma<1$ indicates a bright image, and $\gamma>1$ indicates a dark image.

Focus and Sharpness- circular average filter applied by putting disk value d= 3,5….19. The effect of illumination can be seen using different sets of Yale DB.

Table 1. Quality factors : C, B, F, I, S

DB	B	C	F	S	I
Normal	0.49	0.32	0.03	0.84	0.87
10%	0.46	0.48	0.02	0.56	0.82
20%	0.44	0.52	0.006	0.48	0.84
30%	0.39	0.56	0.003	0.37	0.76
40%	0.37	0.62	0.002	0.24	0.62
50%	0.32	0.66	0.003	0.23	0.59
60%	0.23	0.73	0.002	0.20	0.25
70%	0.25	0.78	0.0013	0.19	0.91
80%	0.22	0.85	0.0012	0.15	0.82
90%	0.15	0.92	0.0014	0.13	0.72

Table 2. Quality factors: E, O measures by considering random 5 images of DB.

DB	E	O
I1	2.4	0.95
I2	3.45	0.85
I3	2.96	0.76
I4	2.74	0.72
I5	2.45	0.56

The next step is to combine the quality factors to get a single index value. Due to this, a clustering technique is used that is able to analyse intra-cluster and inter-cluster differences in the context of data analysis. When inter-cluster distance is greater and intra-cluster distance is less, it is found to be a good cluster. Comparative analysis is done with respect to several clustering techniques to determine the effectiveness of the better one, as illustrated in Table 3. Out of them, DBSCAN gives the better result.

Table 3. Inter-cluster and Intra-cluster distance for each clustering technique

Algorithm/Dataset	Color-Feret		Kinect-Face	
	Inter	**Intra**	**Inter**	**Intra**
K-Means	75.69	80.95	59.08	88.65
Agglomerative	89.85	93.23	75.07	94.22
DBSCAN	**91.49**	**25.02**	**72.38**	**54.48**
Birch	89.83	96.23	60.0	94.22
Gaussian Mixture	16.05	57.31	19.22	56.41
Affinity Propagation	59.78	70.57	70.52	98.12
Bisect-Means	75.69	90.95	59.08	88.65

A comparative analysis with other existing methodologies is shown in Table 4, which determines that PittPatt has remarkable recognition performance. Further, human quality assessment (HQA) becomes essential after system evaluation. Typically, we interview a panel of three people to assess images and determine subjective scores based on individual visual perception. Subsequently, to accomplish the correlation between system scores and human scores, the Spearman correlation coefficient (SROCC), also known as Spearman's rho (ρ), is used. Therefore, with results of 0.88, 0.90, and 0.95, it is evident that the score of the third HQA performs considerably better.

Table 4. Comparison of various face detection, fusion schemes to classify the input images, where B: brightness, C: contrast, E: entropy, F: focus, I: illumination, O: occlusion, P: pose S: sharpness, R: resolution.

Database	Description	Face Detection Rate	Fusion Rule	Correct Matching (%)	Correlation	Recognition
FRI CVL	S, R, P, B	94.3%	$S = \sum_{i=1}^{4} W_i * S_i$	92.1%	--	
Hermes project		90.5% (Cascade Classifier)		87.1%	--	
FOCS	C, B, F, S, I	91.3% (LBP)	Neural Network [1]	---	--	
Color-Feret	C, B, S, P, O	92.7% (DCNN)	DCNN [2]	---	0.85	
Kinect Face					0.76	
Clustering	C, B, F, S, I, O, E	93.2 %	DBSCAN	---	0.88 0.90 0.95	PittPatt (98.6%)

The system processes images as inputs and generates quality assessments using a single quality index ranging from q0 to q4. In this scale, q0 represents excellent image quality, q1 indicates good quality, q2 suggests fair quality, q3 implies better quality, and q4 signifies the worst image quality, as illustrated in Table 5. The first two grayscale images are taken from the Color-Ferret database, while the next two RGB images are taken from the Kinect Face database.

Table 5. Quality based rankings

Score/Data				
System Score	**0**	**1**	**3**	**2**
Human Score-1	2	1	2	2
Human Score-2	1	1	3	2
Human Score-3	**0**	**1**	**3**	**4**

This work highlights the importance of IQA in enhancing the performance of face recognition systems. The proposed technique has been tested statistically to obtain the confidence level between the FIQA technique and human observers, and it has demonstrated better performance in face recognition techniques.

Keywords: Biometric, clustering, face recognition, image quality assessment, image qual-ity factor, neural network

References

1. Abaza, Ayman, et al.:. Design and evaluation of photometric image quality measures for effective face recognition. IET Biometrics 3.4, pp. 314-324, (2014).
2. Zhuang, Ning, et al.: Recognition oriented facial image quality assessment via deep convolutional neural network. Neurocomputing 358, pp. 109-118, (2019).

Optimization of crop production in the regions of India: A Hybrid Technique Using SimulatedAnnealing and XGBoost for Crop Profit Maximization

Malayalla Nissie[1], Bellam Keerthana[1], Perumalla Mounika Reddy[1], Mekala Sravani[1], Amrutha Varshini[1], G Madhukar Rao[1]

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, India
{nissiehft, keerthanabellam2310, Mounikareddy1459, Sravanichowdary1459, amruthavarshini9966}@gmail.com

Abstract

Within the tremendous and changed rural scene of India, where differing climatic conditions, soil sorts, and edit inclinations win, the journey for feasible and productive cultivating hones remains ever-pressing. Recognizing the complex exchange of variables affecting edit generation, this consider endeavors to present and explain the viability of a novel cross breed strategy pointed at optimizing agrarian results over distinctive locales of the nation.

At the heart of this imaginative approach lies the integration of two effective techniques: XGBoost and Mimicked Tempering. These strategies, in spite of the fact that unmistakable in their instruments, meet to make a comprehensive system for maximizing trim benefits whereas exploring the complexities characteristic in Indian horticulture.

Mimicked Toughening, drawing motivation from the tempering prepare in metallurgy, serves as a metaheuristic optimization calculation custom-made to fine-tune choice factors basic for asset assignment and trim arranging. Its versatile nature permits for the investigation of arrangement spaces, empowering iterative alterations to agrarian hones in reaction to changing natural conditions and showcase elements. By probabilistically tolerating adjustments based on a temperature parameter, Mimicked Strengthening optimizes asset allotment methodologies, guaranteeing effective utilization of inputs and maximizing the financial returns from trim development.

Complementing this optimization process is the utilization of XGBoost, a vigorous machine learning calculation eminent for its prescient modelling capabilities. Leveraging an gathering learning system and slope boosting strategy, XGBoost exceeds expectations in capturing complicated connections between agrarian parameters and edit abdicate. By analyzing authentic information and important highlights, XGBoost encourages exact determining of edit results, giving important bits of knowledge for educated decision-making in trim administration and generation arranging.

The collaboration between Reenacted Toughening and XGBoost causes an all-encompassing approach to rural optimization, interestingly custom fitted to the multifaceted challenges experienced in Indian horticulture. By harmonizing the qualities of these techniques, the crossover procedure adeptly navigates the energetic scene of edit generation, accommodating the goals of surrender maximization with the imperatives of asset accessibility and natural supportability.

Central to the viability of this crossover approach is its flexibility to the assorted rural settings winning over diverse locales of India. Recognizing the nuanced varieties in climate, soil richness, and edit inclinations, the technique utilizes a custom-made optimization methodology for each district, guaranteeing significance and adequacy in tending to neighbourhood agrarian challenges.

The results of this ponder verify to the adequacy of the cross breed procedure in upgrading edit benefit and sustainability over shifted agrarian scenes in India. Through thorough experimentation and examination, the inquire about illustrates how the integration of Recreated Toughening and XGBoost yields substantial changes in trim yields and financial returns, subsequently advertising a promising pathway towards economical escalated of farming.

Besides, past its quick suggestions for edit generation, the cross breed technique holds broader importance for the headway of agrarian science and hone. By cultivating development and intrigue collaboration, it embodies a

proactive approach to tending to the complex challenges going up against cutting edge horticulture, from climate alter adjustment to nourishment security enhancement.

The integration of Mimicked Tempering and XGBoost speaks to a spearheading endeavor in optimizing edit generation in India. By tackling the synergies between optimization calculations and prescient modeling strategies, this crossover approach offers a vigorous system for maintainable and productive horticulture, balanced to catalyze transformative alter within the Indian rural scene and past.

Keyword(s): Agricultural optimization, Hybrid techniques, XGBoost, Simulated Annealing, Crop profitability

References

1. Odegard, I. Y. R., & Van der Voet, E. (2014). The future of food—Scenarios and the effect on natural resource use in agriculture in 2050. Ecological Economics, 97, 51-59.
2. Ederer, P., Baltenweck, I., Blignaut, J. N., Moretti, C., & Tarawali, S. (2023). Affordability of meat for global consumers and the need to sustain investment capacity for livestock farmers. Animal Frontiers, 13(2), 45-60.
3. Zeferino, J. A., Antunes, A. P., & Cunha, M. C. (2009). An efficient simulated annealing algorithm for regional wastewater system planning. Computer- Aided Civil and Infrastructure Engineering, 24(5), 359-370.
4. Kuo, S. F., Liu, C. W., & Merkley, G. P. (2001). SW—Soil and water: application of the simulated annealing method to agricultural water resource management. Journal of Agricultural Engineering Research, 80(1), 109-124
5. Glória, A., Cardoso, J., & Sebastião, P. (2021). Sustainable irrigation system for farming supported by machine learning and real-time sensor data. Sensors, 21(9), 3079
6. Bassine, F. Z., Epule, T. E., Kechchour, A., & Chehbouni, A. (2023). Recent applications of machine learning, remote sensing, and iot approaches in yield prediction: a critical review. arXiv preprint arXiv:2306.04566.
7. Madhukar Rao, G., & Ramesh, D. (2016). Supervised learning techniques for big data: a survey. IJCTA. Int Sci Press, 9, 3811-3891

AI Enabled Stress Monitoring using PPG and GSR

Swagata Barik[1], Vinay Kumar Thakur[1], Mohasin Ali Miah[1], and Saurabh Pal[2]

[1]Department of Applied Physics, University of Calcutta, Kolkata, India
{swagatabarik1997b,thakur.vinay06,alimohasin62}@gmail.com
[2]Department of Applied Physics, University of Calcutta, Kolkata, India
spaphy@caluniv.ac.in

Extended Abstract

Aim: Stress is an integral part of modern life. It impacts over quality of living and worsens mental and physiological condition in the long run. So smart healthcare system requires monitoring and analysis of the mental stress for overall wellbeing.

Modern wearable sensing technology enables researchers to develop AI-based data driven algorithms for such purposes [1]. Present work reports a preliminary study conducted over 15 subjects where Photoplethysmography (PPG) and Galvanic skin response (GSR) signals are used to study the stress level of the subjects.

Methods: Laboratory dataset of PPG and GSR signal is developed using BIOPAC MP160 data acquisition system from 15 volunteers (Age= 20-31 years, Male: Female= 60: 40). Data was collected under different stressors under controlled room temperature between 25 ± 2 °C. All participants were instructed to undergo through three different conditions like resting (3-minutes duration), Math test (calculating backward arithmetic subtraction for 3-minutes) and Stroop color word test (2-minutes) while PPG and GSR were captured with a sampling rate of 500 Hz. A visual illustration of the methodology is presented in Figure 1. Artifacts and noise were removed from the raw PPG and GSR signals using standard pre-processing techniques. Peak- to-peak interval (PPI) from PPG signal were used to compute temporal and statistical features like pulse rate variability (PRV), mean, standard deviation (SD), root mean square of successive differences (RMSSD), standard deviation of successive differences (SDSD) of PRV and Pulse Transit Time (PTT) computed with ECG R-peak reference.

Tonic and phasic components of GSR signal were differentiated using Empirical Mode Decomposition (EMD). Features like phasic energy per phase (from phasic components) and sectional slopes from tonic components were extracted. Feature optimization from the combined feature set was performed by Random Forest (RF) feature importance estimation technique based-on Gini impurity [2]. Finally, Various classifiers [3] were used to detect stress-related activities compared to resting states. Further, classification between two stressors were also performed for realization of the impact of stress sources over subject.

Results: PPG and GSR based features were independently analysed to identify the most suitable features from each group. Being the most significant features, Mean of Pulse rate (PR), RMSSD of PRV from PPG and Number of peaks, Energy per peak of phasic components from GSR were used for classification. Conventional model performance analysis parameters [4] were calculated to quantify the results. The optimal results achieved

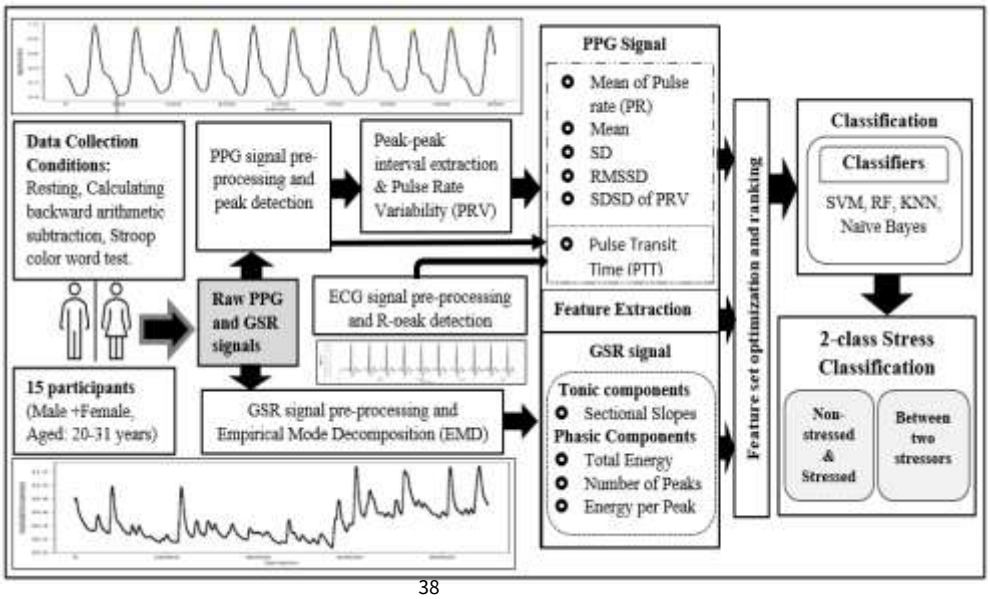


Figure 1 Graphical depiction outlining the proposed system configuration

from the best two features of PPG and GSR signal using an RF classifier with GridSearchCV, yielding Accuracy (80± 8.31), Precision (86.43± 12.04), Recall (86.67± 12.47), Specificity (66.67± 29.81), and F1-score (85.17± 5.98), with hyperparameters set to max_leaf_nodes=9, min_samples_split=5, and n_estimators=75.It was also obtained that PPG-based features predominantly differentiated stressors with (76.67± 13.33) accuracy.

Conclusion: In summary, the existing evidence on the system indicates that stress affects PPG and GSR. Exploitation of such autonomic biosignals can be used with proper AI based feature optimization and machine learning algorithms for dynamic monitoring of stress. Machine learning classifiers are utilized for detecting mental stress due to their ability to identify underlying pattern of huge dataset generated due to continuous acquisition of PPG and GSR. Thus, it provides objective assessments, facilitate real time monitoring, offers personalized solutions and integrate seamlessly with wearable devices for convenient and continuous stress management. This underscores their potential utility in objectively evaluating psychological wellbeing and stress levels differentiating from resting condition as well as between two stressors. By analysing the deviation of features from resting value can provide the impact of stressors used in this study on subject-to-subject basis. The study can be further extended to diverse datasets and towards development of the smartphone-based mobile healthcare applications.

Keyword(s): Mental Stress Analysis; PPG and GSR; Time domain module; Tonic and Phasic Components; Machine Learning Application.

References

1. Zhu, L., Spachos, P., Ng, P. C., Yu, Y., Wang, Y., Plataniotis, K., Hatzinakos, D.: Stress Detection Through Wrist-Based Electrodermal Activity Monitoring and Machine Learning. IEEE journal of biomedical and health informatics, vol. 27, issue.5, pp. 2155–2165 (2023)
2. Disha, R.A., Waheed, S.: Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity 5, 1 (2022)
3. Martinez-Ríos, E., Montesinos, L., & Alfaro-Ponce, M.: A machine learning approach for hypertension detection based on photoplethysmography and clinical data. Computers in biology and medicine, 145, 105479 (2022)

Telecom Churn Stacking-Classifier Ensemble Model with Higher AUROC

Vankayala Sri Sailaja[1], Rajput Sahil[1], and Goyal Mayank Kumar[1]

[1]School of Engineering and Technology, Sharda University, Greater Noida, India
{vankayala, sahil, mayank.goyal2}@ug.sharda.ac.in

Extended Abstract

Subscribers integrated to telecom domain have many options available to choose service providers and to swap. It is one of the fastest growing city in eastern zone of India. This contributes to the per annum churning rate of 15-25 percent in this immense competitive presence, Subscriber churn is a critical parameter and metric as it is proven to cost less to retain existing subscriber than in turn to acquire new one. Various studies have been conducted to achieve predictive models and algorithms to accurately predict customer churn in the telecommunications industry. Keeping the necessary parameters under consideration, we conducted our study on to build a Telecom Churn Stacking Classifier Ensemble Model with Higher AUROC incorporating predictions from one machine learning model become predictors for another, with which aim was to develop solution and to take advantage of complex composite structures to reduce prediction error and variance. We have tried to address imbalanced dataset with SMOTE resampling technique along with feature selection and post algorithm implemenation of Random Forest Classifier and Gradient Boosting Algorithm. Also, we have provided an overview to compare the traditional prediction with ensemble classifier as XGBOOST and Stacking Classifer in respect to the model presented based on Accuracy and Precision, ROC Curve and AUC. We observerd that XGBoost has the ROC AUC: 0.9823 after class balancing, although conventional Random Forest Classifier and Gradient Boosting Algorithm proves more efficient and outperfomed with the accuracy improvised upto 95% , with class imbalance dataset addressed.

Keyword(s): Telecom Churn, Ensemble Model Classifier, Higher AUROC, Comparative Analysis

References

1. Torsten, J., Gerpott., Wolfgang, Rams., Andreas, Schindler. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. Telecommunications Policy, (2001)
2. Aryan, Raj., D., Vetrithangam. Prediction of customer churn using resampling and ensemble classification algorithms on telecom dataset. (2023).
3. Akinyemi, Omololu, Akinrotimi., Roseline, Oluwaseun, Ogundokun., M., A., Mabayoje., R., A., Oyekunle., Marion, Olubunmi, Adebiyi. A Smote-Based Churn Prediction System Using Machine Learning Techniques. (2023).
4. Changqing, Su., Xianzhong, Xie. Churn Prediction in Telecommunications Industry Based on Conditional Wasserstein GAN.(2022).
5. Teuku, Alif, Rafi, Akbar., Catur, Apriono. Machine Learning Predictive Models Analysis on Telecommunications Service Churn Rate. Green Intelligent Systems and Applications, (2023).
6. R., Raja, Subramanian., M., Lakshmi., M., Lavanya., M., K., G., Roy., K., Krishnaveni., T., Yogeswari. Leveraging Machine Learning Algorithms for predicting Churn in Telecom industries.(2022).
7. Almadhor, A., Sampedro, G.A., Abisado, M., Abbas, S.; Kim,Y.-J., Khan, M.A., Baili, J., Cha, J.-H.: Wrist-Based Electrodermal Activity Monitoring for Stress Detection Using Federated Learning. Sensors , 23, 3984 (2023)

A Hybrid GAN-UNet for Lung Segmentation from Interstitial Lung Disease (ILD) Affected High-resolution CT Images

Biswadev Goswami[1,2], Tapas Pal[1], Sagnik Das[2], and Rajesh P. Barnwal[1,3]

[1]Faculty of Engineering Sciences (CSIR-CMERI Campus)
Academy of Scientific & Innovative Research (AcSIR),
Sector 19, Kamla Nehru Nagar, Ghaziabad – 201 002, U.P., India

[2]Department of CSE, Dr. B.C. Roy Engineering College, Jemua Road, Fuljhore, Durgapur – 713 206, India

[3]AI & IoT Lab, IT Group, CSIR-CMERI, Mahatma Gandhi Avenue, Durgapur – 713 209, WB, India
r_barnwal@cmeri.res.in

Extended abstract

Significant advancements in medical imaging over the past two decades, employing techniques like CT scans, MRI, fluoroscopy, etc., have played a vital role in delivering critical diagnostic information essential for saving lives. Although the age-old practice of manually reviewing these images has successfully saved numerous lives, its efficacy is confined to the expertise of radiologists and clinical specialists, resulting in subjective and offline interpretations. It has already been identified in different literature that computer-based medical image processing can reduce the treatment planning time and increase the accuracy of disease diagnosis. Interstitial Lung Disease (ILD) is the general term for various lung parenchymal disorders that impair the gas exchange process in the lungs. In individuals with ILDs, breathing problems arise. The prevalence of ILDs is very high in India, and it is increasing rapidly due to air pollution and lifestyle disorders. When diagnosing ILD, doctors commonly order a chest X-ray or HRCT scan to acquire a more comprehensive examination of the lungs. If the affected lungs can be appropriately segmented, the doctors can better view the diseased area. Lesion segmentation, an integral aspect of image interpretation, offers additional information that can profoundly impact disease detection, staging, and treatment planning. Deep Learning models are data-hungry. Most deep learning-based segmentation researchers use natural datasets that are publicly available, like PASCAL VOC or COCO.The process of medical image segmentation encompasses the isolation of regions of interest (ROIs) from image datasets, as seen in Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) scans.

Keyword(s): Segmentation; HRCT; ILD; GAN; U-Net

This research paper aims to present a deep learning-based lung segmentation method that demonstrates high efficiency, mainly when dealing with minimal amounts of image data. We have used data from the MedGIFT[1] database and the publicly available Kaggle OSIC pulmonary fibrosis progression dataset of lungs HRCT. MedGIFT's database comprises DICOM files, each containing multiple CT slices. We have converted selected slices of CT images to JPEG images of 512 x 512. The database also includes masks corresponding to each image's region of interest (ROI). Subsequently, with the assistance of radiologists, we selected and categorized CT images that included the presence of ILDs. U-Net and its sub-types exhibit high efficiency in image segmentation. However, they come with the drawbacks of substantial data requirements and a lengthy training process. GANs can produce synthetic instances, thereby contributing to dataset balance and enhancing the performance of the segmentation model on classes that are typically underrepresented [2]. Our model utilizes a generator with a UNET-like architecture and includes encoder and decoder blocks with skip connections, chosen for superior image generation in our ROI.

NETWORK ARCHITECTURE

The hybrid setup merges generator and discriminator networks in a GAN framework. Discriminator layers are fixed, while a separate trainable discriminator is used—original and generated CT scans with masks from input pairs. Generator losses include L1 loss (Mean Absolute Error) and are updated with a 100:1 ratio (MAE: BCE). Training alternates gradient descent on the discriminator and generator. The discriminator's accuracy ceiling is 50%, with higher values suggesting bias or overfitting. Optimal performance is near 50% accuracy.

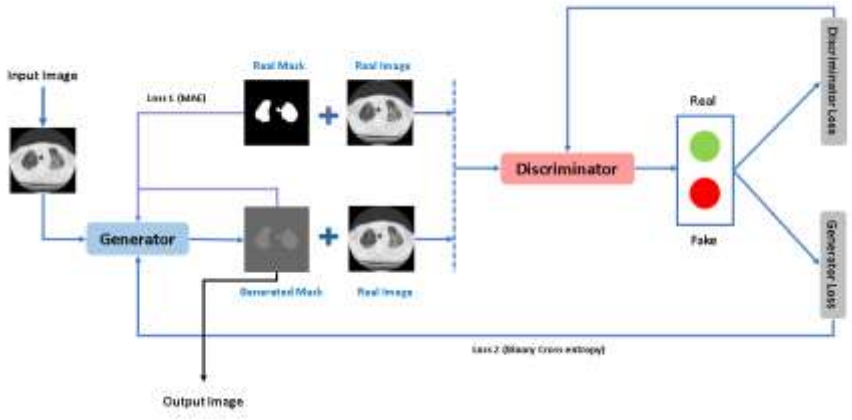


Figure 1: Network architecture of the proposed hybrid model for Lung Segmentation

RESULT

We evaluated our combined GAN model using Accuracy, Binary Cross Entropy(BCE), and Mean Absolute Error(MAE) metrics. MAE measures absolute variance between predicted and actual values, while BCE, also known as Log Loss, assesses binary classification distinctions. During training, we applied a weighted sum of BCE and MAE, with a 100:1 ratio favouring MAE, optimizing for semantic segmentation tasks. Our experimental results yielded 84.39% accuracy for the Generator and 46.78% for the Discriminator, with an MAE of 0.008.

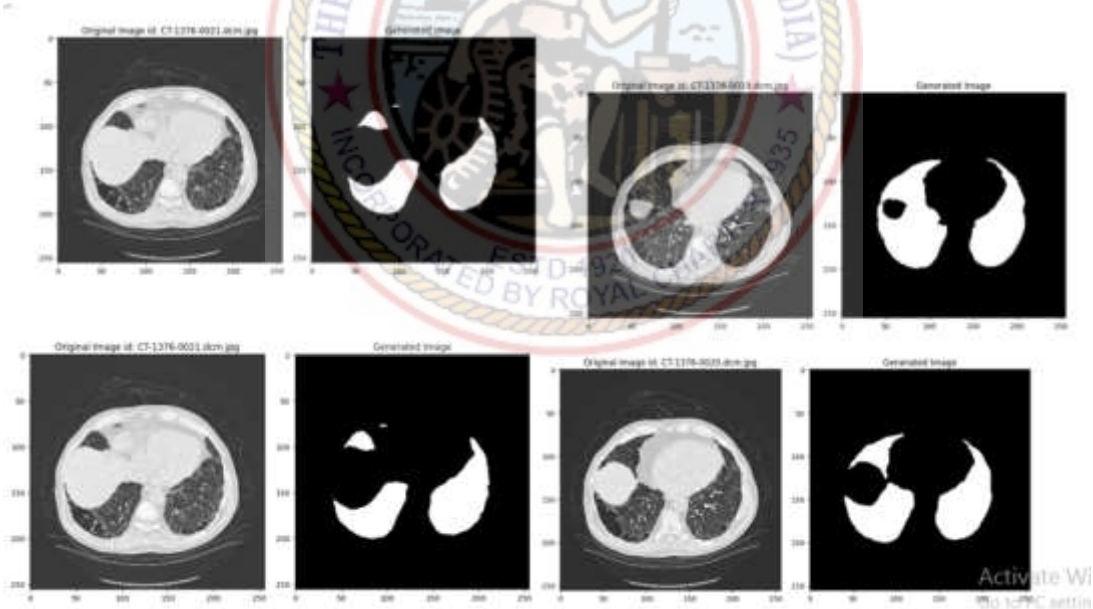


Figure 2: Original CT scan images (left) paired with segmented images (right) after segmentation.

References

1. Depeursinge, A and Vargas, A and Platon, A and Geissbuhler, A. and Poletti,PA and Müller,H, "Building a reference multimedia database for interstitial lung diseases", Computerized Medical Imaging and Graphics,2012,36(3),227-238,10.1016/j.compmedimag.2011.07.003

2. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ''Generative adversarial nets,'' in Proc. 27th Int. Conf. Neural Inf. Process. Syst., 2014, pp. 2672–2680.
3. Murugappan, M., et al. "Automated semantic lung segmentation in chest CT images using deep neural network." Neural Computing and Applications 35.21 (2023): 15343-15364.
4. Shoaib, Muhammad, Rabia Naseem, and Amir H. Dar. "Automated segmentation of lungs in computed tomographic images." Eur J Sci Res 98 (2013): 45-54.
5. Liu, Wufeng, et al. "Automatic lung segmentation in chest X-ray images using improved U-Net." Scientific Reports 12.1 (2022): 8649.

Early Detection of Knee Osteoarthritis: A Simple and Effective Approach using Local Directional Octa Pattern (LDOP)

Saksham Garg[1], Rinku Datta Rakshit[1], Deep Suman Dev[2] and Dakshina Ranjan Kisku[3]

[1]School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, India
gargsaksham16@gmail.com, rinku.rakshitfcs@kiit.ac.in

[2]Department of Computer Science and Engineering, School of Science and Technology, Neotia University, West Bengal, India
deepsumandev@yahoo.co.in

[2]Department of Computer Science and Engineering, National Institute of Technology Durgapur, Durgapur, India
drkisku.cse@nitdgp.ac.in

Extended Abstract

Knee osteoarthritis (KOA) is a degenerative joint condition affecting millions worldwide. The KOA offers a serious healthcare concern owing to the steady degeneration of knee joints. This condition causes pain and inhibits the movement of a person. Early identification of this condition is critical for preventing disease progression and implementing successful treatment strategies. With the evolution of computer vision (CV), Computer-Aided Detection (CAD) technologies have demonstrated a new path to success in KOA diagnosis. This study investigates the potential of Local Directional Octa Pattern (LDOP), a simple and powerful local descriptor to identify reliable early KOA. LDOP gathers essential information from images, such as joint space narrowing and osteophyte development, to successfully discriminate between healthy and KOA-affected knees. In this study, LDOP outperformed deep learning based approaches for KOA detection. This finding suggests that LDOP has the potential to be an effective tool in clinical settings for accurate and timely early KOA diagnosis, paving the way for better patient outcomes.

The proposed method consists of three stages (Input, Feature Extraction, and Classification). The input stage collects X-ray images and resizes them. The Feature extraction stage extracts features of Knee images in eight directions to identify discriminatory patterns and generate feature vectors. The Classification stage classifies healthy vs. KOA using three classifiers (Histogram Intersection (HI), Euclidean Distance (EuD), and Sum of Absolute Differences (SAD)). The Kaggle's freely accessible Knee Osteoarthritis Dataset is used for the experiment. Images were classified into five severity levels using Kellgren and Lawrence (K&L) grading. The dataset was divided into training (5778 pictures) and testing (1656) sets. The model's accuracy was assessed using three classifiers (HI, EU, and SAD) with varying numbers of retrieved images (Top 5, Top 10, and Top 20) for comparison. When applied to the top 20 most similar photos, the HI classifier had the best accuracy (97.8%). Deep learning technologies, like Convolutional Neural Networks (CNNs), have demonstrated potential in KOA diagnosis. Studies with publicly available datasets show accuracy ranging from 61% to 97.4% [1]-[5]. However, deep learning models are computationally costly and can lack interpretability, making it difficult to comprehend how they make judgments.

The LDOP provides promising findings for early KOA detection via X-ray imaging, additional research may look at several areas for improvement like Multimodal Integration (Combining X-ray data with additional modalities, like as MRI scans, can provide more detailed information for KOA diagnosis) and Clinical Integration (Additional research is needed to determine the effectiveness of LDOP in a clinical setting). This study highlights the potential of LDOP as an accurate and effective technique for early KOA diagnosis. Its simplicity and interpretability provide advantages compared to complicated deep learning systems.

The Figure 1 represents a knee image and its transformed images using LDOP and Figure 2 represents the block diagram of LDOP.

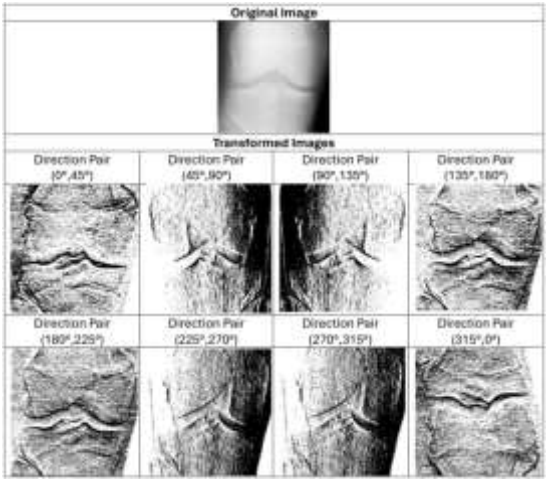


Figure 1: The original image and the corresponding transformed images using LDOP.

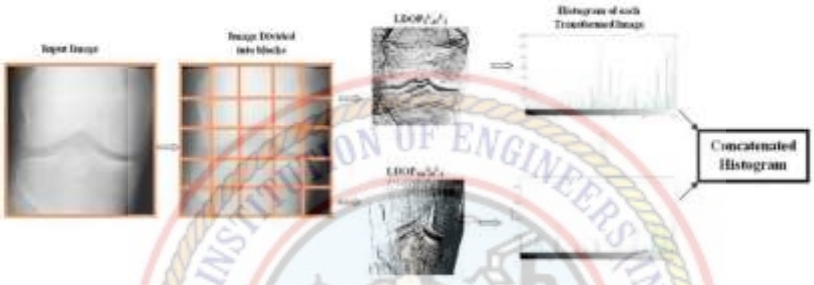


Figure 2: Block diagram of local directional octa pattern (LDOP).

Keyword(s): Knee osteoarthritis; computer vision; Computer-Aided Detection; Local Directional Octa Pattern; deep learning.

References

1. Guan, B., Liu, F., Mizaian, A.H., Demehri, S., Samsonov, A., Guermazi, A. and Kijowski, R., 2022. Deep
2. learning approach to predict pain progression in knee osteoarthritis. *Skeletal radiology*, pp.1-11.
3. K. Üreten and H. H. Maras, ''Automated classification of rheumatoid arthritis, osteoarthritis, and normal hand radiographs with deep learning methods,'' J. Digit. Imag., vol. 35, no. 2, pp. 193–199, Apr. 2022.
4. Arumugam, S. R., Balakrishna, R., Rajeshram, V., Gowr, S., Karuppasamy, S. G., & Premnath, S. P. (2022, November). Prediction of severity of Knee Osteoarthritis on X-ray images using deep learning. In *2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon)* (pp. 1-5). IEEE.
5. Ahmed, N., Saeed, M., Aftab, M., Mehmood, A. and Ilyas, Q.M., 2023, September. Knee Osteoarthritis Detection And Classification Using Transfer Learning. In *2023 3rd International Conference on Computing and Information Technology (ICCIT)* (pp. 365-369). IEEE.
6. Cueva, J. H., Castillo, D., Espinós-Morató, H., Durán, D., Díaz, P., & Lakshminarayanan, V. (2022). Detection and classification of knee osteoarthritis. *Diagnostics*, *12*(10), 2362.

Machine Learning and its Revolution in Healthcare Services: A Survey

Ashish Padhi[1], Saksham Garg[1], Rinku Datta Rakshit[1], and Dakshina Ranjan Kisku[2]

[1]School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, India
ashishpadhi2004@gmail.com,gargsaksham16@gmail.com,rinku.rakshitfcs@kiit.ac.in
[2]Department of Computer Science and Engineering, National Institute of Technology Durgapur, Durgapur, India
drkisku.cse@nitdgp.ac.in

Extended Abstract

Healthcare is an inevitable part of our daily life. In our society, the Healthcare sector is a major field which provides varied healthcare services like healthcare related medical services, medical insurance to patients, manufacturing of drugs or medical equipment to make our daily life much easier. The increasing populations and reports of different new diseases created tremendous pressure on healthcare sectors. The COVID-19 pandemic brought enormous challenging needs in healthcare sectors due to staffing shortage and patient surges. The demand for changes inspires healthcare personnel to make changes at every level of the healthcare environment. Machine Learning (ML) is a kind of technology that can alleviate the tremendous pressure of healthcare personnel and is capable of fulfilling the expectations of patients. The ML models have the potential to extract discriminatory patterns from medical data and medical images, those are very useful to detect and predict a disease at its early stage efficiently. This research study has reviewed the background, current scenarios, and future scope of ML based services. This study also discusses the need, benefits and challenges of ML based services in healthcare sectors. ML has potential to add a new milestone in healthcare services by providing Computer-Aided Diagnosis (CAD), and Computer-Aided Prognosis (CAP).

Machine Learning has several benefits in healthcare services. Disease progression prediction algorithms may be used to forecast the course of an illness and the chances of a patient's mortality. The aforementioned data may be utilized to make treatment and care decisions, such as when to start therapy early or send a patient to a medical specialist. Novel Drug Discovery algorithms may be used to find different drug targets and create new medications. This has the power to change the way illnesses are treated. ML is capable to increase patient safety by detecting prescription mistakes or identifying patients at danger of falling.

Artificial intelligence (AI)-powered systems can help doctors diagnose illnesses more accurately and quickly by processing vast amounts of data from patient surveys, electronic health records, and medical imaging. It has been shown that deep learning algorithms are quite effective in identifying medical images. Machine learning will be able to further revolutionize healthcare in the years to come. Machine learning may be used to create personalized medicine, which is medical therapy that is tailored to the particular patient. This may be accomplished by analyzing a patient's medical history, genetic information, and other characteristics to determine the best therapy choices for that patient. ML models such as neural networks were inspired by the human brain. They are composed of linked artificial neurons arranged in layers, and they are excellent at jobs requiring pattern recognition, data analysis, and judgment. Neural networks' adaptability and versatility have contributed to the development of numerous AI applications, and can be used in the healthcare field. The table 1 lists accuracy of different disease detection using ML.

TABLE 1: Accuracy of different Disease Detection using ML

Sl no.	Name of Disease	Published work	Year	Techniques used	Accuracy
1	Knee osteoarthritis	[1]	2022 - 2023	Radon features	85%
2	Stroke detection	[2]	2021 - 2023	BrainNet (BrN)	95%
3	Tumor	[3]	2023	Deep CNNs	95% - 99%
4	Dibeatic retinopa-thy	[4]	2022 - 2023	Machine learning and deep learning algorithms	90%-95%

Keyword(s): Machine Learning; Healthcare; Healthcare services; Computer-Aided Diagnosis; Computer-Aided Prognosis.

References

[1] Roomi, S. M. M., Suvetha, S., Maheswari, P. U., Suganya, R., & Priya, K. (2023, May). Radon Feature Based Osteoarthritis Severity Assessment. In *2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)* (pp. 1-5). IEEE.

[2] Tripura, D., Haque, I., Dutta, M., Dev, S., Jahan, T., Ghosh, S. K., & Islam, M. A. (2023, May). A BrainNet (BrN) based New Approach to Classify Brain Stroke from CT Scan Images. In 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT) (pp. 1-6). IEEE.

[3] Babu, K. R., Teja, S. S. B., Reddy, V. S., Deep, K. M., Ramesh, M. N., & Rambabu, K. (2023, January). Analytical review of deep neural techniques for breast tumor detection. In *2023 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-7). IEEE.

[4] Hatode, P., & Edinburgh, M. M. (2022, December). Machine learning & deep learning algorithm reviews for diabetic retinopathy detection. In *2022 5th International Conference on Advances in Science and Technology (ICAST)* (pp. 164-169). IEEE.

A CNN-Trasnformer Approach for Bengali Handwritten Text Recognition

Vishal Bania[1], and Sujan Kumar Saha[1]

[1]Department of Computer Science and Engineering, NIT Durgapur, West Bengal
{vb.22p10377@mtech, sksaha.cse}@nitdgp.ac.in

Extended Abstract

Handwritten Text Recognition (HTR) is the process of converting handwritten textual content into machine-readable texts. The challenges in handwritten character recognition arise from the inherent complexity and variability in individual handwriting. When we focus on HTR in Indian languages, the difficulty is much more incredible compared to English because of the script-specific challenges like a more extensive set of characters, diacritic forms of vowels, and a few consonants, consonant conjuncts, and allographs. Although a substantial amount of work has been carried out to develop Bengali HTR, the publicly available systems did not give satisfactory accuracy when we tested them. For instance, when we tested a publicly available HCR model on a dataset of Bengali handwritten text images collected from various schools, we found that the recognition rate was lower than expected. This is because the test dataset is not similar to the training dataset. Then, we started developing a Bengali HTR system with higher accuracy in cross-dataset settings. We used the publicly available Bengali BanglaWriting [1] and BN-HTRd [2] datasets for the development. Initially, we developed the baseline system using a typical CNN-LSTM model. The first model is built using a 7-layer CNN architecture, fed into two bi-directional LSTM layers for making the final prediction. The CNN architecture uses a combination of batch normalization and max pooling layers. Our model internally uses the CTC loss function as an endpoint layer to calculate the error rate between the predicted and actual labels. We performed several hyper-parameter tuning experiments on our Baseline Model to obtain the best model using the CNN-LSTM architecture. In our experiments using the BN-HTRd dataset, we achieved the highest word recognition rate (WRR) of 81.81%.

Later, we performed experiments using the Transformer encoder to achieve better accuracy. During these experiments, we achieved the best performance using a model that uses a custom 13-layer CNN architecture followed by a positional encoding layer and then fed into the encoder module of our Transformer, as shown in Fig 1. The output from the encoder is then finally fed into a linear layer, which produces the corresponding predictions. The Transformer-based model achieved a WRR of 95.18% on the BN-HTRd dataset.

When we performed a cross-dataset evaluation by training the system using the BN-HTRd dataset and tested it on the Bangla Writing dataset, our model achieved a Character Recognition Rate (CRR) of 89.48% and WRR of 70.15%. Hossain et al. [3] also conducted similar cross-dataset experiments using the same datasets, achieving a CRR of 84.46% and a WRR of 63.22%. Therefore, the proposed CNN-transformer model outperforms the model developed by Hossain et al.

Keyword(s): Handwritten Bengali Text Recognition; CNN; LSTM; Transformer.

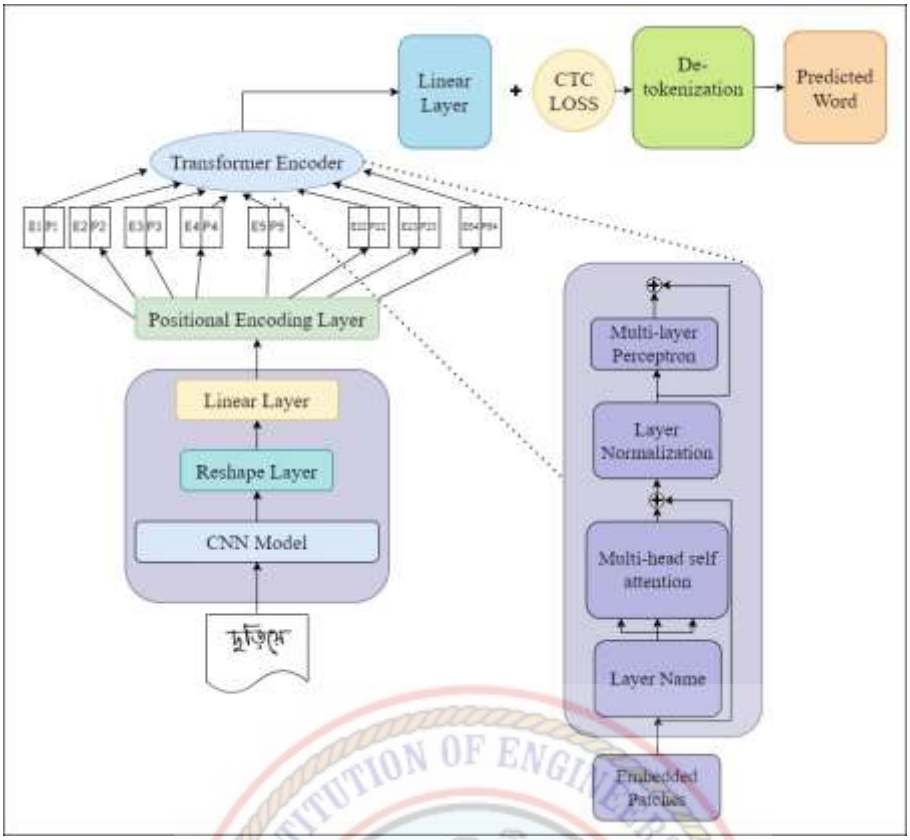


Figure 1: Proposed CNN-Transformer model

References

1. M.F. Mridha, "BanglaWriting: A Multi-Purpose Offline Bangla Handwriting Dataset", 2021, [online] Available: https://data.mendeley.com/datasets/r43wkvdk4w/1
2. M. A. Rahman, N. Tabassum, M. Paul, R. Pal, M. K. Islam, "BN-HTRd: A Benchmark Dataset for Document Level Offline Bangla Handwritten Text Recognition (HTR) and Line Segmentation", arXiv:2206.08977, https://doi.org/10.48550/arXiv.2206.08977, 2022
3. M. I. Hossain, M. Rakib, S. Mollah, F. Rahman, and N. Mohammed, "LILA-BOTI: Leveraging Isolated Letter Accumulations By Ordering Teacher Insights for Bangla Handwriting Recognition", 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 2022, pp. 1770-1776.

Relevance of Artificial Intelligence in Marketing: A Narrative Review and Future Research Directives

K. Srikanth[1], and Sneha De[2]
[1]Department of Hospital Management, Brainware University, Kolkata
dks.hm@brainwareuniversity.ac.in

[2]Department of Food & Nutrition, Brainware University, Kolkata
sd.fnt@brainwareuniversity.ac.in

Extended Abstract

Technological innovations for instance, internet of things, big data analytics, block chain and artificial intelligence led to a paradigm shift in the ways the businesses operate. Amongst all the technological innovations, Artificial Intelligence (AI) is the contemporary technological disruptor and acquires an immense marketing transformation potential. Theorists across the globe are attempting to decipher the gold-standard AI solutions for their marketing functions. In specific, Artificial Intelligence agents driven by the machine learning algorithms are rapidly transforming the business world, generating heightened interest from the researchers. In the current paper, a qualitative content analysis was done to analyze the secondary data sources and the findings of the content analysis resulted in a deeper understanding on the significance of Artificial Intelligence in the discipline of marketing. Deriving upon the evidence from the text analysis, the current paper provides a narrative review of various machine learning tasks which includes – supervised learning, unsupervised learning, semi-supervised and transfer learning, active learning and lastly reinforcement learning. By synthesizing the observations from the secondary data sources, the significance of AI in the context of traditional marketing mix - product management, price management, place management, and lastly in promotion management were portrayed. This article further uncovers the AI driven marketing industry trends which includes - Interactive and media-rich, Personalization and targeting, Real-time optimization and automation and lastly customer-journey focus. We then conclude with the future research directions of AI & machine learning in the marketing context.

Keyword(s): Artificial Intelligence, Content Analysis, Machine Learning, Marketing, Technological Disruptors.

References

1. Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A.: Customer relationship management and big data enabled: Personalization & customization of services. Applied Computing and Informatics. 15(2), 94-101 (2019)
2. De Haan, E., Kannan, P. K., Verhoef, P. C., & Wiesel, T.: Device switching in online purchasing: Examining the strategic contingencies. Journal of Marketing. 82(5), 1-19 (2018)
3. Huang, M. H., & Rust, R. T.: A strategic framework for artificial intelligence in marketing. Journal of the Academy of Marketing Science. 49, 30-50 (2021)
4. Zhang, H., Cao, X., Ho, J. K., & Chow, T. W.: Object-level video advertising: an optimization framework. IEEE Transactions on industrial informatics. 13(2), 520-531 (2016)

Covid -19 Detection from Chest X-ray using Deep Learning

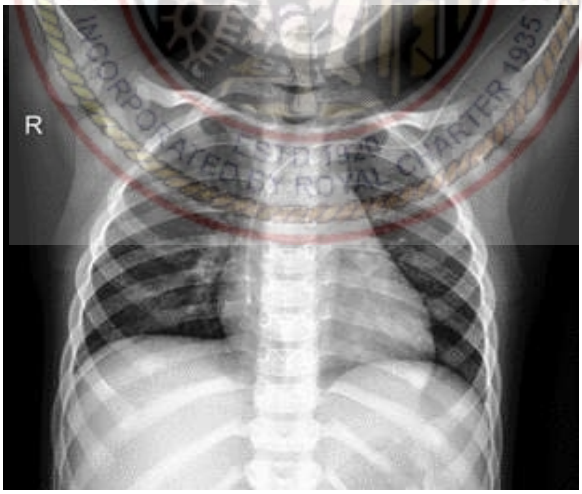
Nilabja Dutta

Department of IT, Dr. B C Engineering College, Durgapur, India
nilabjadutta1234@gmail.com

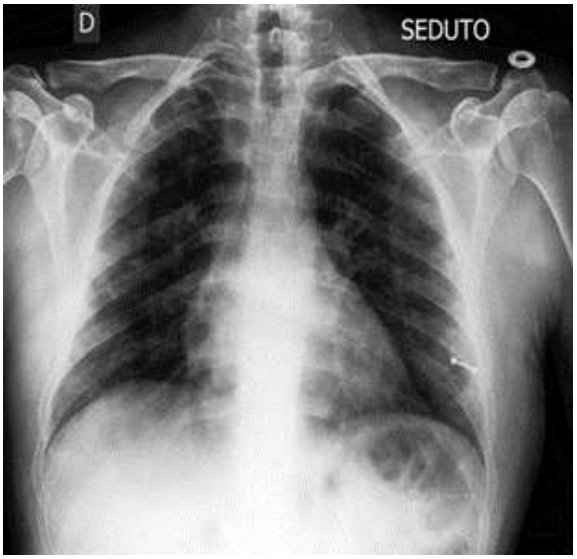
Extended Abstract

We all know that covid-19 was the most controversial and concerned topic in the last 4 to 5 years. The whole world had seen the pandemic situation that occurred due to the coronavirus disease. If we see the situation of India only then the second wave reports tells that in 2021 when the second wave came lakhs of people were getting infected every day. Many people lost their lives, their way of income. Peoples were affected not only physically but also mentally. Many people had lost their lives in corona. This disease also affects organs like lungs, brain, etc [1]. Though it has reduced now, through vaccines and other protections, but it is not completely destroyed. It can again occur in the future also. So a well-defined and appropriate covid 19 detection system and technology is very important. Here I introduced a covid 19 detection system from chest X-Ray using deep learning techniques. Currently deep learning is a very powerful machine learning technique to classify images. It is basically a neural network technique [2]. Basically if a person is suspected to have covid -19 then it can be confirmed by checking his chest X-Ray. So I collected a dataset of chest X-Ray having 111 covid positive patient X-ray and 70 covid negative X-Ray. Data augmentation is done on this dataset to increase the dataset size. And finally with the help of deep learning techniques I trained the dataset and made a covid 19 detection algorithm. The training accuracy I got is 99.13% , validation accuracy of about 99.37% and testing accuracy about 97.75% which is pretty good.

Keywords: Deep Learning; Data Augmentation; Class Balancing



Normal person's Chest X-ray



Covid-19 affected person's Chest X-ray

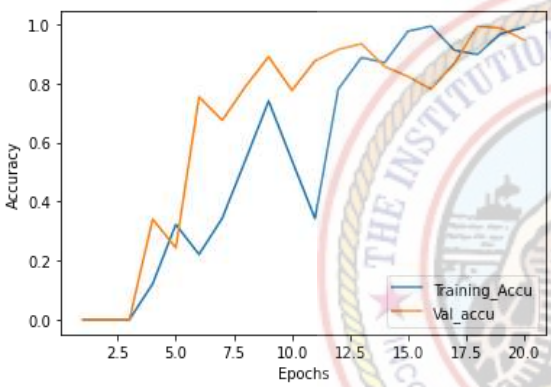


Figure 1: training accuracy and validation accuracy

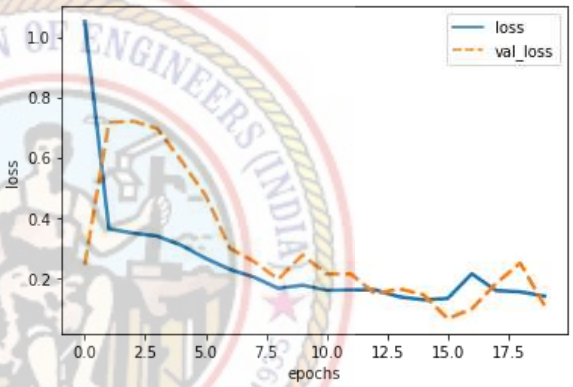


Figure 2: Training loss and validation loss

References

1. Krishna Mohan Agarwal , Swati Mohapatra , Prairit Sharma , Shreya Sharma , Dinesh Bhatia , Animesh Mishra, "Study and overview of the novel coronavirus disease (COVID-19)", Sensors InternationalVolume 1, 2020, 100037
2. Kavi B. Obaid, Subhi R. M. Zeebaree, Omar M. Ahmed, "Deep Learning Models Based on Image Classification: A Review", International Journal of Science and Business,Volume: 4, Issue: 11 Page: 75-81 2020

Solar Prognostication: Machine Learning based PV Power Generation in Kolkata (22.57'N, 88.41'E), India

Akash Layek[1], Ayan Ghosh[2], Ramij Mallick[3], Moumita Ghosh[4], B. Tudu[5], K.K. Mandal[6]

Power Engineering Department, Jadavpur University, Kolkata, India
[1]akashlayek70@gmail.com,[2]ghoshayan3705@gmail.com, [3]ramijn7@gmail.com,
[4]moumitaghosh1095@gmail.com, [5]bhimsen.tudu@jadavpuruniversity.in,
[6]kkm567@yahoo.co.in

Extended Abstract

Intermittency and volatility in photovoltaic (PV) generation presents great challenge for grid operators. Ensuring the security and reliability of electricity dispatch demands operators to have precise visibility into the capacities of both synchronous and asynchronous generators. The present work aims to provide machine learning-enabled forecasting for PV power generation for Kolkata (22.57'N, 88.41'E), a bustling urban center in India. The study has been conducted in Power Engineering Department, Jadavpur University, Salt Lake Campus (22.57'N, 88.41'E), a location endowed with abundant renewable resources, particularly solar energy [1]. The department has PV panels deployed at rooftop for catering partial load of the department. The input data and power output were directly measured from the panels, enabling precise energy yield forecasts. Solar power generation prediction utilizes variables like solar irradiance and temperature as input parameters discarding other parameters like rainfall and relative humidity due to sparse data availability. Fig. 1 shows the site location and system components deployed in site [2].

Figure 1: Site location and system architecture [2]

Various machine learning techniques, like Linear Regression (LR), Polynomial Regression (PR), Decision Tree Regression (DTR) using GridSearchCV optimizer with calculated best parameters, 'max_depth' as 7, 'max_features' as 10 and 'max_leaf_nodes" as 52, Support Vector Regression (SVR), Random Forest Regression (RFR), along with deep learning approaches of Multilayer Perceptron Regression (MLP) using two hidden layers of 1000 and 100 neurons respectively with "relu" activation function and data preprocessing using MinMaxScaler and 'Adam' optimizer iterated 100 times have been explored in this study. The block digram of adopted aproach for energy prediction is depicted in Figure 2.

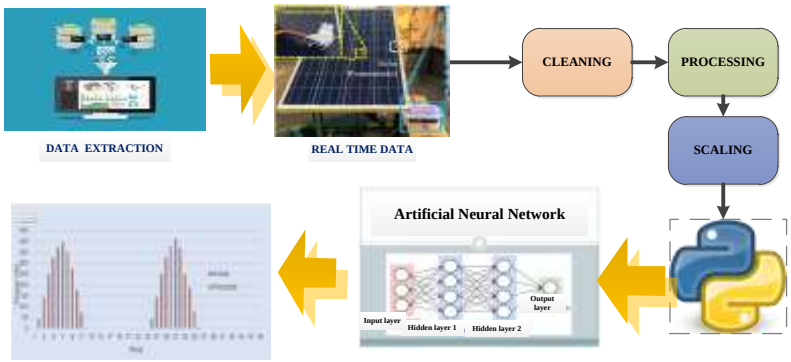


Figure 2: Block diagram of adopted machine learning model [3]

Comparative performance analyses using various performance matrices like R^2-score, mean absolute error (MAE), mean squared error (MSE), median absolute error (MedAE) have been conducted to assess their effectiveness [4]. The total data set of 25,560 with training-test ratio of 85:15 has been considered for entire dataset. Various performance matrices for different algorithm are tabulated in Table 1. Polynomial Regression emerges as best fit model with R^2-score of perfect 1.0 for training and test data among machine learning methods and Multilayer Perceptron Regression (MLP) comes out as superior performer amongst all for the presented dataset. R^2 Matrıx variation for different algorithms is shown in Figure 3.

Table 1: Performance matrices for different algorithms for entire dataset.

Algorithm	MAE	MSE	MedAE	R^2 Score
LR	3.28238	21.97904	2.31097	0.99928
PR	2.22904e-11	6.51019e-22	1.90141e-11	1.0
SVR	2.86974	34.60694	1.64334	0.99909
DTR	2.72212	23.579	0.36922	0.99928
RFR	0.20629	0.3371	0.01058	0.99998
ANN	0.13662	0.0359	0.07618	0.9999

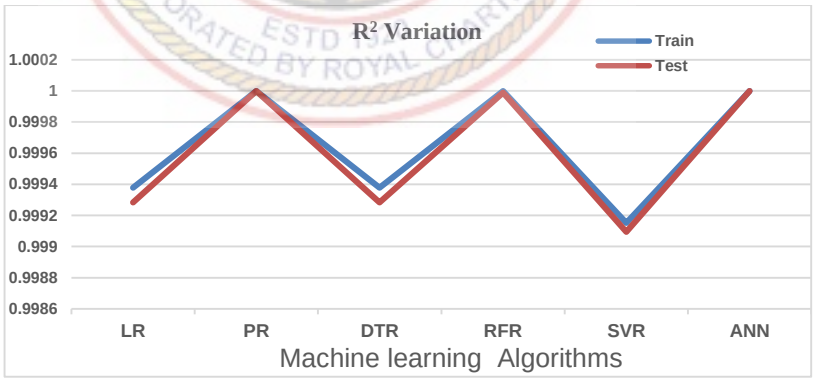


Figure 3: R^2 Matrıx variation for different algorithms

Various works related to energy prediction were carried out worldwide [4-5] However, as far as authors' knowledge, the study and application of machine learnings in Indian context is sparse [6] and present study provides valuable insights to assist Indian grid operators in choosing appropriate PV power forecasting algorithms and preparing for potential generation volatility in the future.

Keyword(s): Renewable energy; PV power generation, Artificial intelligence; machine learning; Deep learning

References

1. Solar Radiation Hand Book. A joint Project of Solar Energy Centre, MNRE Indian Metrological Department, http://www.indiaenvironmentportal.org.in/files/srd-sec.pdf; 2008 [Accessed on 18.03.2024]
2. Tudu, B., Mandal, K.K., Chakraborty, N. Optimal design and development of PV-wind-battery based nano-grid system: A field-on-laboratory demonstration. Front. Energy 13, 269–283 (2019). https://doi.org/10.1007/s11708-018-0573-z
3. Hassan, M. K., Youssef, H., Gaber, I. M., Shehata, A. S., Khairy, Y., El-Bary, A. A.: A predictive machine learning model for estimating wave energy based on wave conditions relevant to coastal regions. Results in Engineering, 21, 101734 (2024)
4. Mahmud, K., Azam, S., Karim, A., Zobaed, S., Shanmugam, B., Mathur, D.: Machine learning based PV power generation forecasting in Alice Springs. IEEE Access, 9, 46117-46128 (2021)
5. López-Flores, F. J., Ramírez-Márquez, C., Rubio-Castro, E., Ponce-Ortega, J. M.: Solar photovoltaic panel production in Mexico: A novel machine learning approach. Environmental Research, 246, 118047 (2024)
6. Garg, I., Raj, A., & Jajoria, A.: Solar Power Production Forecasting in India using Machine Learning Approach. In: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1265-1272. IEEE, (2021, May)

Birch Clustering Approach for Detection of Random Valued Impulse Noise

Sulagna Basu[1],and Aritra Bandyopadhyay[1]

[1]Department of Computer Science and Engineering, Supreme Knowledge Foundation Group of Institutions, Mankundu,West Bengal, India
sulgnabasu9@gmail.com, aritra.d90@gmail.com

Extended Abstract

An essential component of computer vision is image processing, which has uses in a variety of fields including autonomous driving, surveillance, and medical imaging. Detecting and segmenting image data depend heavily on clustering, an essential unsupervised learning activity. Due to scalability and efficiency, Birch (Balanced Iterative Reducing and Clustering utilizing Hierarchies) clustering has become a popular and promising method for image processing applications. Large datasets are challenging to handle with constrained resources, such as memory or a slower CPU, and clustering techniques such as K-means clustering do not conduct clustering very effectively. Consequently, as dataset sizes grow, standard clustering methods do not scale well in terms of running times or quality. Birch clustering is useful in this situation. Birch initially creates a tiny, compacted brief of the bulky dataset while retaining ample statistics as feasible. Afterward, the smaller summary is grouped rather than the larger dataset. Birch is frequently used in conjunction with other clustering algorithms, providing a summary of the dataset that the other method can use. Birch clustering has benefits like scalability, efficiency in high-dimensional areas, and handling of huge datasets. Birch clustering effectively arranges images into a hierarchy of sub-clusters by iteratively creating a tree structure, enabling efficient segmentation and representation. Image segmentation, which divides an image into discrete areas based on similarities in pixel intensities or characteristics, is one important use of Birch clustering in image processing. Furthermore, by locating representative centroids inside clusters, Birch clustering makes feature extraction easier and allows for concise and insightful representations of picture data. Additionally, Birch clustering can be incorporated into a number of pipelines for image processing, such as those for texture analysis, object recognition, and picture retrieval. It is appropriate for a variety of image processing tasks in both academic and industrial applications due to its versatility in handling noise and adaptability to varying data distributions. Birch clustering arranges data into a hierarchical tree structure for scalability, in contrast to k-means, which divides data into predefined groups based on centroids. Birch clustering is a more scalable method. Alternative clustering method DBSCAN differentiates between neighbors around core points and clusters based on density in contrast to Birch and k-means, which are reliant on centroids and fixed cluster formations. However memory management is much better in BIRCH clustering. As the data are first subdivided into sub-clusters by an iterative process, and they are condensed into centroids with the use of Birch clustering, which results in the formation of a hierarchical tree, it saves data in memory in an effective manner at the foundation enabling it to manage massive datasets in an effective manner. Through the process of splitting and combining data in order to establish a hierarchy, makes it possible to achieve scalable and effective clustering with reduced processing complexity. For the purpose of evaluating the effectiveness of the proposed technique, a variety of typical test photographs, including Baboon, Peppers, Lena, Barbara, Boat, Bridge, and Fingerprint, are utilized. The proposed method was found to have higher accuracy in noise detection when related to three established methods that used "miss" and "false-hit" detection schemes as can be seen from the results. By comparing the result of the Proposed method with the existing methods based on the 'miss' value and 'false-hit' value marks a minimum margin of 12165-pixel values and 603-pixel values which signifies the superiority of the proposed method.

Keyword(s): Birch clustering; Digital images; Noise removal; Random Valued Impulse Noise; Sub-cluster.

References

1. Dong, Y., Xu, S.: A new directional weighted median filter for removal of random-valued impulse noise. IEEE signal processing letters 14(3), 193-196 (2007).
2. Saikrishna, P., Bora, P. K.: Detection and removal of random-valued impulse noise from images using sparse representations. In: International Conference on Image Processing, pp. 1197-1201. IEEE (2013).
3. Deka, B., Handique, M., Datta, S.: Sparse regularization method for the detection and removal of random-valued impulse noise. Multimedia Tools and Applications 76(5), 6355-6388 (2017).

Natural Language Processing in Healthcare

Shubham[1], Ms Anjali Dhamiwal[2], and Ms Monika Saini[3]

[1]Student, CSE, World College of Technology and Management, Gurgaon, India
shubh971797@gmail.com
[2]Assistant Professor, CSE, World College ofTechnology and Management, Gurgaon, India
anjalidhamiwal17@gmail.com[2]
[3]HOD , CSE, World College of Technology and Management, Gurgaon, India
mona.khushi79@gmail.com

Extended Abstract

Natural Language Processing (NLP) has emerged as a transformative technology in the healthcare industry, offering various innovative and modern solutions for long-standing challenges. By enabling computers to understand, interpret, and produce human language, NLP has the potential to revolutionize various aspects of healthcare, including clinical decision-making, patient communication, and administrative tasks.

The evolution of NLP techniques in healthcare has been remarkable, from rudimentary tasks like text parsing to sophisticated machine learning algorithms such as recurrent neural networks and transformers. These advancements have empowered NLP to handle the complexities and nuances of clinical language, enabling more accurate and context-aware analyses of textual data from electronic health records (EHRs), clinical notes, and medical literature.Applications of NLP in healthcare are diverse and impactful. From improving clinical documentation accuracy to supporting clinical decision support systems, disease surveillance, patient monitoring, and health management tasks, NLP has demonstrated its effectiveness in enhancing patient care, streamlining clinical workflows, and advancing public health initiatives.

Through comprehensive case studies and literature reviews, the effectiveness of NLP in addressing healthcare challenges has been shown. Experiment results from the implementation of NLP models in healthcare applications show significant improvements in various metrics. For instance, in a Clinical Documentation Improvement (CDI) case study, the NLP-based system achieved a 30% increase in coding accuracy and a 25% reduction in manual chart review time. Similarly, a Clinical Decision Support System (CDSS) based on NLP algorithms showed a 40% reduction in treatment variability among oncologists, leading to more consistent and evidence-based treatment decisions.

The implementation of NLP models in healthcare applications undergoes rigorous performance evaluation, utilizing metrics such as precision, accuracy, recall, F1 score, and AUC-ROC to quantify model performance. Cross-validation techniques ensure model robustness and generalizability across different datasets or settings.Despite the promising opportunities presented by NLP in healthcare, several challenges remain. Data privacy and security concerns, complexity of clinical language, lack of standardized datasets, and biases in NLP models are among the challenges that need to be solved. However, innovative solutions and future research directions offer various promising solutions to overcome these challenges.

The Indian Government also understood the power of EHR Records and their importance in Health Sector.The ABHA (Ayushman Bharat Health Account) come into existence in 2021 to achieve these goals.NLP in Healthcare can be operational at each sector and achieve its goal only in Public Private Partnership Model.

Strategies such as strong encryption technologies, access control, and anonymization methods can protect patient data and ensure compliance with regulatory frameworks such as HIPAA. Obtaining informed consent from patients before using their health data in NLP applications promotes transparency and respects patient autonomy. Additionally, addressing biases in NLP models through dataset diversification, bias-based training, and algorithmic fairness evaluation fosters equity and inclusion in healthcare.

The future prospects of NLP in healthcare are promising, with continued collaboration and recognition of the ethical implications of its application. By leveraging NLP technologies responsibly and ethically, healthcare

organizations, researchers, policymakers, and technology developers can harness its full potential to transform healthcare delivery, improve patient outcomes, and drive innovation in healthcare.

Keyword(s): Natural Language Processing, Healthcare, EHR, NLP Effectiveness

Figures and Tables

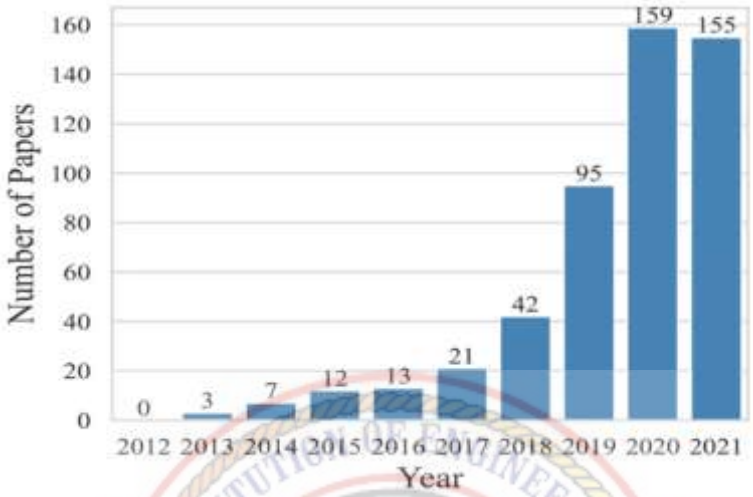


Figure 1: Distribution of number of papers from 2012 to 2021

Figure 2: Natural Language Processing for Healthcare & Life Sciences Market Value ($Million)

References

1. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. doi:10.1038/s41591-018-0300-7
2. https://towardsdatascience.com/a-decade-of-knowledge-graphs-in-natural-language-processing-5fdb15abc2b3H.
3. Liao, K., & Han, L. (2019). Natural language processing in health informatics: A review. Journal of Biomedical Informatics, 94, 103–117. doi:10.1016/j.jbi.2019.103117

Density Based Clustering Approach Combined with Fuzzy Logic for Elimination of Random Valued Impulse Noise

Sulagna Basu[1],and Aritra Bandyopadhyay[1]

[1]Department of Computer Science and Engineering, Supreme Knowledge Foundation Group of Institutions, Mankundu,West Bengal, India
sulgnabasu9@gmail.com, aritra.d90@gmail.com

Extended Abstract

Reducing impulsive noise is one of the most crucial aspect in the post-acquisition image processing steps. The reduction of impulse noise has an effect on the noise patterns in pictures, making it an essential method. This study is being conducted with the intention of developing a method that is a two-step procedure and is called as DCIFF (DBSCAN clustering identified fuzzy filter). The purpose of this method is to successfully remove impulsive noise from digital photographs while preserving the quality of the images. The DBSCAN clustering approach is utilized in order to achieve the aforementioned objectives. The subsequent part is about a technique for minimizing noise that is founded on the principles of fuzzy reasoning. The approach DBSCAN, which functions as a density-based clustering algorithm, is utilized with the intention of discovering commonalities that are present in every single 5×5 frame. In order to generate these windows, each pixel in the picture must be positioned in the exact center of the frame. The formation of a window with dimensions of 7 × 7 is done with the intention of successfully minimizing noise. In this specific frame, each and every pixel that has been identified as being disturbed is located in the middle of the window. After the computations have been finished, the non-noisy pixel intensities are multiplied by the triangle fuzzy membership values, and the products that are obtained from this process are then added to one another. There are a number of iterations of this strategy that are carried out in fast succession following one another. The total result is then divided by the sum of the fuzzy membership values of all of the pixels that are not made up of noise. For the purpose of carrying out the process of eliminating the noise from the core pixel that has been compromised by noise, it is important to make use of the outcome. In order to create an evaluation of the method, the DCIFF is reviewed using traditional image evaluation criteria. These criteria are used for a broad variety of images or photographs during the course of the evaluation process. DCIFF was able to get peak-signal-to-noise ratios (PSNR) of 25.86 dB and structural similarity indexes (SSIM) of 0.8429 for the Lena picture that had been altered by random valued impulsive noise at a noise level of 80%. These results were obtained throughout the process of analyzing the picture in different iterations. After analyzing the photograph, these findings were obtained having minimalized deviation. Both the PSNR and SSIM values were considered to be significant when viewed through the lens of statistical significance respectively. The DBSCAN Algorithm has a significantly lower susceptibility to outliers compared to previous partitioning approaches. Furthermore, considering that fuzzy logic is explicitly developed to address situations characterized by ambiguity and uncertainty, it is rational to employ it in the management of the unexpected element impacted by abrupt and disruptive sounds. The combined quantitative and qualitative results along with the comparison with current research undeniably demonstrate that the recommended process is far more effective than the most advanced filtering technologies.

Keyword(s):DBSCAN clustering; Image processing; Noise removal; Impulse noise; Random valued impulse noise.

References

1. Roy, Amarjit, et al. "Combination of adaptive vector median filter and weighted mean filter for removal of high- density impulse noise from colour images." *IET image processing* 11.6 (2017): 352-361.
2. Jin, Kyong Hwan, and Jong Chul Ye. "Sparse and low-rank decomposition of a Hankel structured matrix for impulse noise removal." *IEEE Transactions on Image Processing* 27.3 (2017): 1448-1461.
3. Roy, Amarjit, Lalit Manam, and Rabul Hussain Laskar. "Region adaptive fuzzy filter: an approach for removal of random-valued impulse noise." *IEEE Transactions on Industrial Electronics* 65.9 (2018): 7268-7278.

Enhanced Deep Learning Based Fingerprint Image Quality Assessment

Parul Choudhary[1], Phalguni Gupta[2], and Pooja Pathak[3]

[1,2,3]Department of Engineering & Applications, GLA University, Chaumuhan, Mathura, U.P, India
[1]cparul2605@gmail.com, [2]pg@gla.ac.in, [3]pooja.pathak@gla.ac.in

Extended Abstract

Fingerprint recognition has become a cornerstone technology in various applications, ranging from law enforcement to smartphone security. However, the quality of fingerprint images can significantly affect the performance of recognition systems. Traditional methods of assessing fingerprint image quality often rely on handcrafted features and simplistic models, which cannot capture the complexity of real-world scenarios. To address this limitation, in this research, an enhanced deep learning-based approach for fingerprint image quality assessment (FIQA) is proposed as shown in Fig 1.

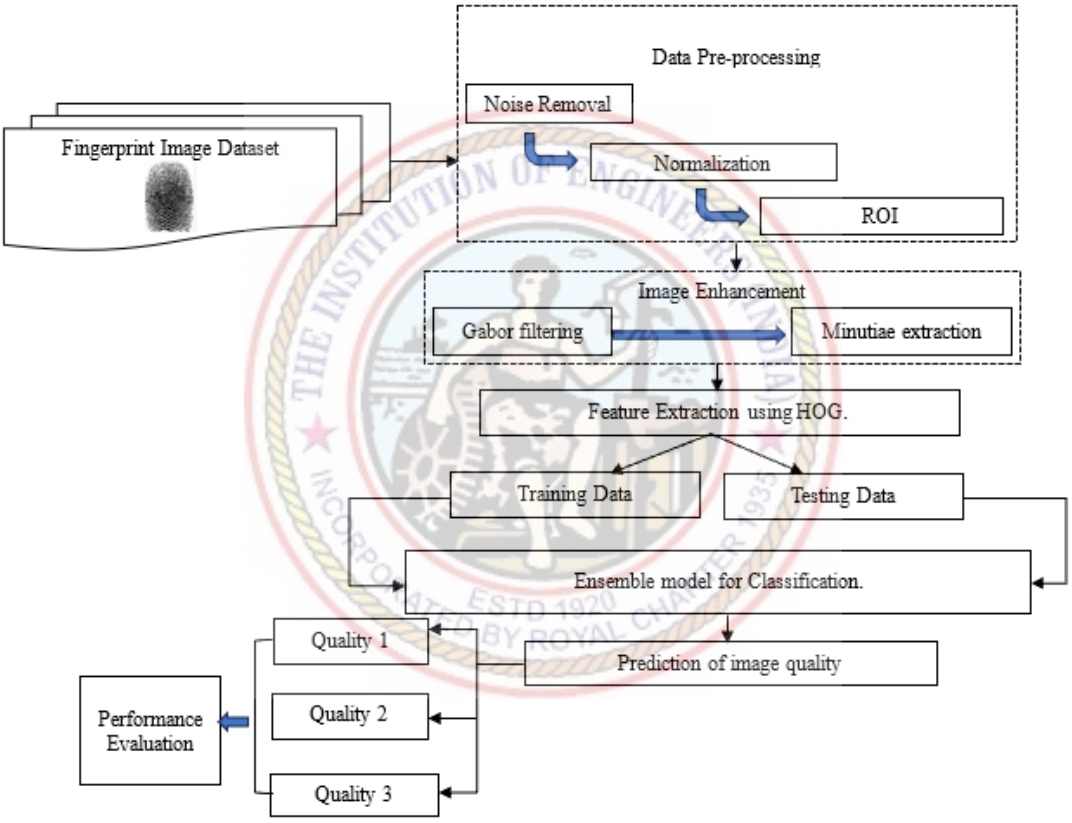


Fig 1. Fingerprint Image Quality Assessment (FIQA)

The following steps describe the methodology in detail-

Step 1: Fingerprint Image Dataset
Initially, data collected from the SOCOFing dataset is represented as FID which encompasses 6,000 fingerprint images that were obtained from 600 individuals with variations in clarity, contrast, illumination, and noise levels. Dataset consists of two classes real and altered. Fig 2 shows some examples of real images and Fig 3 shows an example of altered images.

Figure 2: Real Images

Figure 3: Altered Images

Step 2: Data Pre-processing

- **Noise Removal:** Gaussian smoothing is used to reduce noise while preserving important image details. The Gaussian smoothing operation is represented as:

$$f_{smooth}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} * f(x,y)$$

where $f(x, y)$ is the original image and σ is the standard deviation.

- **Normalization:** The initial fingerprint image goes through a sequence of pre-processing procedures. The image is first normalized. This is done to minimize the fluctuations in grey level values in peaks and valleys.

$$fnorm = \frac{f_{old} - f_{old}^{min}}{f_{old}^{max} - f_{old}^{min}}$$

ROI (Region of Interest) Selection: ROI is identifying and extracting the region containing the fingerprint from the background and helps in focusing the analysis on the relevant information.

Step 3: Image Enhancement

- **Gabor Filtering**: Gabor filter is used for enhancing texture features in images, making them useful for improving the visibility of fingerprint ridge structures, represented as:

$$g(x,y;f,\theta) = exp(-2\sigma^2 x'^2 + \gamma^2 y'^2) exp(i2\pi f x')$$

- **Minutiae Extraction**: Identifying and extracting minutiae points from the fingerprint image enables the extraction of distinctive features that are used for fingerprint matching and recognition. Let M represent the set of extracted minutiae points, defined as:

$$M = \{(x_i, y_i, \theta_i)\}_{i=1}^{N}$$

Step 4: Feature Extraction using HOG.

HOG descriptor is used to capture the local gradient information in an image, making it effective for representing texture and shape features in fingerprints. HOG descriptor is represented as:

$$HOG(x,y) = \sqrt{(G_x(x,y))^2 + (G_y(x,y))^2}$$

Step 5: Data Split into Training and Testing Data

The dataset FID is split into training and testing sets, denoted as FID_{train} and FID_{test}, respectively.

Step 6: Apply the Deep Learning Model for Classification

In the step. design and train CNN, SVM, and MLP architectures for classifying fingerprint image quality.

- **CNN:** Let X represent the data from the input fingerprint image. CNN is composed of several layers, such as fully connected, pooling, and convolutional layers.

Given an input image X, the output of a CNN can be represented as:

$$Y_{CNN} = f_{CNN}(X; \theta_{CNN})$$

where f_{CNN} represents the CNN model with parameters θ_{CNN}.

- **SVM:** Using a collection of training data$\{(X_i, y_i)\}$, where X_i is the input fingerprint image and yi is the matching label indicating image quality, SVM attempts to determine the appropriate hyperplane for classifying the data.

The decision function of an SVM can be represented as:

$$f_{SVM}(X) = sign\left(\sum_{i=1}^{N_{SV}} \alpha_i y_i K(X_i, X) + b\right)$$

- **MLP:** The input layer, hidden layer (or layers), and output layer are the three main components of an MLP feedforward neural network.

The output of an MLP can be represented as:

$$Y_{MLP} = f_{MLP}(X; \theta_{MLP})$$

where f_{MLP} represents the MLP model with parameters θ_{MLP}.

Step 7: Performance Evaluation
Model performance is evaluated based on accuracy and other relevant metrics on the testing dataset. Techniques such as cross-validation or *k*-fold validation are employed for more reliable performance estimates.

The dataset is divided into training and testing sets, with several deep learning models, such as Convolutional Neural Networks (CNNs), developed and trained to classify the proposed FIQA model. The performance metrics for each model are as follows: CNN achieves 82% accuracy, SVM achieves 72%, MLP achieves 73%, and the ensemble model achieves an accuracy of 87%. Implementation results also show that the proposed model gives a higher accuracy in comparison to the existing conventional methods.

Fig 4. Comparision of original image and qualitative image

The image quality is assessed based on a range system. If the model predicts an image quality between 80-100%, it's categorized as quality 1, indicating high quality. If the predicted quality falls between 60-80%, it's classified as quality 2, signifying good quality. A quality between 40-60% is considered better quality. The Fig 4 illustrates the results, indicating that the model predicts the image quality as good with a matching percentage of 79.45%.
In conclusion, authors' study on fingerprint image quality assessment underscores the importance of employing robust methodologies for accurately evaluating the quality of fingerprint images. Through rigorous analysis and experimentation, we have demonstrated the efficacy of proposed approach in effectively categorizing image quality levels. The findings highlight the potential of our methodology to enhance the reliability and efficiency of fingerprint recognition systems, thereby contributing to advancements in biometric authentication technology.

Keyword(s): Deep learning models, fingerprint image, feature extraction, histogram of oriented gradients, quality assessment.

References

1. Deshmukh, R., & Yannawar, P. (2024). Deep learning based person authentication system using fingerprint and brain wave. International Journal of Computing and Digital Systems, 15(1), 723-739.
2. Priesnitz, J., Weißenfeld, A., Rathgeb, C., Strobl, B., Lessmann, R., & Busch, C. (2023). MCLFIQ: Mobile Contactless Fingerprint Image Quality. arXiv preprint arXiv:2304.14123.
3. El-Naggar, S., & Bourlai, T. (2022). Image Quality Assessment for Effective Ear Recognition. IEEE Access, 10, 98153-98164.
4. Bibi, S., Khan, M. I., Mahmood, M., Gul, F., & Ullah, S. Fingerprint Verification Based on GA and SVM.
5. Aloweiwi, A. H. (2021). Fingerprint classification using transfer learning technique.
6. Giudice, O., Litrico, M., & Battiato, S. (2020, October). Single architecture and multiple task deep neural network for altered fingerprint analysis. In 2020 IEEE International Conference on Image Processing (ICIP) (pp. 813-817). IEEE.
7. Hernandez-Ortega, J., Galbally, J., Fierrez, J., Haraksim, R., & Beslay, L. (2019, June). Faceqnet: Quality assessment for face recognition based on deep learning. In 2019 International Conference on Biometrics (ICB) (pp. 1-8). IEEE.
8. Ezeobiejesi, J., & Bhanu, B. (2018). Latent fingerprint image quality assessment using deep learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 508-516).

An Improved Methodology for Analyzing Resumes and Sorting Candidates Using Statistical Approaches

Ria Roy[1],and Aritra Bandyopadhyay[1]

[1]Department of Computer Science and Engineering, Supreme Knowledge Foundation Group of Institutions, Mankundu, West Bengal, India
riaroy416@gmail.com, aritra.d90@gmail.com

Extended Abstract

One of the most essential aspects of the screening process that takes place in a recruitment consulting agency is the selection of applicants who are suited for the position. This is one of the most crucial components of candidate selection. Generally speaking, when it comes to the process of hiring new staff members, the vast majority of businesses and consulting firms use the conventional methods of applicant recruitment. The key approach that is being utilized in order to get highly competent researchers and developers for high-tech organizations is the utilization of screening procedures that are constructed on the basis of resumes. In order to pick the applicant who is the most suitable for the position, the organization will typically assess the individual's qualifications, professional background, job experience, motivations, expert talents, and previous creative efforts that are reflected in their curriculum vitae (CV). This action is taken in order to guarantee that the individual in question is the most suitable candidate for the position. The purpose of the study that was carried out was to investigate and automate screening procedures that are very effective and are applied by consulting companies. The extraction of tabular bioinformatics data from resumes is often accomplished by the employment of a variety of methodologies, such as automated image analysis and heuristic algorithms. Obtaining information that is relevant to the circumstance is the goal of these strategies, which include gathering information. Specifically, an organized survey was distributed to consultancy companies in Chennai in order to acquire the main data needed for this purpose. The additional data was gathered from the documents of the professional firms, as well as from a variety of periodicals, publications, and webpages throughout the internet. Optical character recognition (OCR) was applied in order to obtain data from the resume sheets (also known as resume sheets). As optical character recognition (OCR) is a viable option, this investigation makes use of the OCR on the data since it removes the underlying automotive engineering that is present in the text. A number of statistical techniques, including the Chi-square test, percentage studies, and the weighted average approach, are then utilized in order to conduct an analysis of the data that is relevant. When selecting these strategies, factors such as experience, technical and non-technical abilities, employability, career advancement, income expectations, and role/designation expectations are taken into consideration. The Chi-square test obtained a calculated Value of 19.7892 and the tabulated Value of 20.09, from which It can be deduced that there is a connection between the candidates' job experiences and their ability to comprehend the required abilities. Moreover, by the weighted average method, a value of 4.04 has been found on a scale of 5 which deduced that the majority of respondents are in favor of maintaining the applicants' pay expectations. Lastly percentage analysis; based on Personal Value of 90%, Interest of 85%, Abilities and Skills of 80%, Goals & Opportunities of 65%, and Employment Preferences of 75% infers recruiters' views on the fundamentals of considerable employability. By implementing the suggested strategy, the recruitment operation was able to effectively choose the appropriate resumes for the subsequent expansion of the process.

Keyword(s): Resume screening process; Optical character recognition; Recruitment; Chi-square test.

References

1. Cowgill, Bo. "Bias and productivity in humans and algorithms: Theory and evidence from resume screening." *Columbia Business School, Columbia University* 29 (2018): 679-681.
2. Berdanier, Catherine GP, Mary McCall, and Gracemarie Mike Fillenwarth. "Characterizing disciplinarity and conventions in engineering resume profiles." *IEEE Transactions on Professional Communication* 64.4 (2021): 390-406.

3. Lamba, Disha, et al. "An integrated system for occupational category classification based on resume and job matching." *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*. 2020.

Online Text Sentiment Analysis using a Hybrid Model Based on BERT and Fuzzy Logic

Sourav Nag[1],and Aritra Bandyopadhyay[1]

[1]Department of Computer Science and Engineering, Supreme Knowledge Foundation Group of Institutions, Mankundu, West Bengal, India
nag123.sourav@gmail.com, aritra.d90@gmail.com

Extended Abstract

Sentiment analysis, sometimes known as SA, is a method that involves evaluating text in order to determine the views, attitudes, and subjectivity that are communicated within it. This strategy is often referred to as "sentiment analysis." As an aspect of the procedure, Natural Language Processing (NLP) methods and procedures are used in order to extract information from records. From the perspective of the various methods, these strategies and methodologies may be classified as either amalgam-based, regulation-based, or machine learning-based. Sentiment analysis is normally used in order to evaluate the viewpoint or perspective of a customer with regard to a product or service. This is accomplished by analyzing the content of the text that is included within the product or service. In order to carry out an accurate analysis of sentiment, it is absolutely necessary to make use of advanced algorithms for text representation. It is necessary for these algorithms to have the capability of converting words into accurate vectors that can offer a description of the text that is being submitted. In spite of the fact that it is a complicated model, the BERT model has garnered a significant amount of prominence all around the world over the course of the last few years. Natural language processing (NLP) techniques, such as supervised text categorization, make it possible to carry out operations without the need for human intervention. This is made possible via the use of these techniques. The fact that this does place does not change the reality that there is still potential for development in terms of the accuracy of sentiment analysis. The purpose of this work is to propose an online text analyzer that makes use of bidirectional encoder representations employing transformers (BERT) in conjunction with a fuzzy logic-based approach in order to identify the sentiment of textual evaluations that are provided by a variety of online customers with a higher degree of precision. The aforementioned strategy will be used in order to successfully attain this goal. In the process of analyzing the feelings conveyed by online writings, the BERT model is most often used with the intention of classifying comments as either good, negative, or neutral. Therefore, this is due to the fact that the BERT model is able to differentiate between the three different kinds of comments. After that, the findings are entered into a Fuzzy Inference System by making use of a triangular fuzzy set. The purpose of this action is to clear up any confusion or uncertainty that may arise about the manner in which any particular paragraph may be categorized as either Absolutely positive, negative, or neutral within the context of the discussion.SigmaLaw-ABSA(Aspect-Based Sentiment Analysis) labeled dataset is used for experimental analysis. Within the context of determining whether or not the model is effectively applied, the metrics that are used include the F1 measure, recall, and precision of the model. The proposed model yielded F1 measures of 0.8332, 0.8325, and 0.6732, the precision of 0.9090, 0.7683, and 0.7906, and recall of 0.7692, 0.9086, and 0.5862 respectively for absolutely positive, negative, and neutral sentiment. Based on the findings of the experiment, it can be concluded that the model that was suggested is far more advanced than the most cutting-edge algorithms that are now accessible.

Keyword(s): Sentiment analysis; BERT; NLP; Fuzzy Inference System; Text classification.

References

1. Sugathadasa, Keet, et al. "Legal document retrieval using document vector embeddings and deep learning." Intelligent Computing:
   Proceedings of the 2018 Computing Conference, Volume 2. Springer International Publishing, 2019.
2. Bhoi, Amlaan, and Sandeep Joshi. "Various approaches to aspect-based sentiment analysis." arXiv preprint arXiv:1805.01984 (2018).
3. 3.https://osf.io/qvg8s/files/osfstorage, retrieved on March'24.

Symptom-Based Disease Prediction: A Comparative Exploration of Decision Trees, Support Vector Machines, and KNN in Predictive Healthcare

Sandeepan Das, Jacob Cherian, Shanmugabalaji H and Vasanthi MS

Dept. of Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, India
{sd0714, jc7514, sq2468, vasanthm2}@srmist.edu.in

Extended Abstract

The ever-expanding domain of electronic health records (EHRs) offers a plethora of data for disease prediction. This study delves into this potential, exploring the efficacy of supervised machine learning (ML) algorithms for symptom-based disease prediction. However, our focus goes beyond traditional healthcare settings.

In conflict zones, where healthcare systems crumble and access to medical professionals is limited, this study explores the potential of machine learning for a user-centric chatbot. This chatbot, designed for resource-constrained environments, goes beyond traditional disease prediction by incorporating injury prediction and treatment recommendations, catering to the diverse medical needs that arise in these areas.

To identify the optimal ML algorithm for our conflict zone chatbot, we compare the performance of Decision Trees (DTs), Support Vector Machines (SVMs), and k-Nearest Neighbors (KNN). While these algorithms have been explored for disease prediction, our research focuses on their suitability within a chatbot specifically designed for conflict zones. Key considerations include interpretability for providing clear explanations to users, efficiency for real-time interaction, and adaptability to potentially limited datasets. This chatbot, accessible through a web interface built with Flask, a Python web framework, ensures a user-friendly interaction for smooth communication between the user and the predictive model improving accessibility globally.

Our methodology begins with acquiring an open-source dataset encompassing a broad spectrum of diseases and their associated symptoms. We meticulously pre-process the data to ensure quality and prepare it for machine learning models. Pre-processing steps may involve handling missing values, encoding categorical data, and feature scaling.

Following data preparation, we evaluate the suitability of each chosen algorithm for disease prediction within the context of the chatbot. We employ a 10-fold cross-validation approach to mitigate overfitting and ensure generalizability of our findings. Here's a breakdown of the algorithms, key hyperparameters optimized, and achieved performance metrics:

Decision Trees (DTs): We explored the impact of maximum tree depth (4-10) and minimum number of samples per leaf (2-5) on classification accuracy (0.92-0.97) and model complexity. DTs achieved good accuracy but could be prone to overfitting with deeper trees.

Support Vector Machines (SVMs): We investigated the influence of the kernel function (linear and radial basis function) and the cost parameter (C) on both accuracy (0.90-0.95) and model interpretability. Linear SVMs offered better interpretability but may not capture complex relationships. Kernel SVMs achieved higher accuracy but at the cost of interpretability.

k-Nearest Neighbors (KNN): We analyzed the effect of the number of neighbors (k) on prediction performance. KNN achieved a balanced accuracy of 0.98 with k=5, demonstrating effectiveness with categorical data like symptoms. However, KNN's performance can be sensitive to noisy data and high dimensionality.

Our analysis compared Decision Trees (DTs), Support Vector Machines (SVMs), and k-Nearest Neighbors (KNN) for suitability within our conflict zone chatbot. While DTs offered good accuracy, their interpretability suffered with deeper structures. SVMs achieved competitive accuracy, especially with kernel functions, but at the cost of interpretability. KNN emerged as the most appropriate choice due to its balanced performance and the highest accuracy. KNN excels in interpretability, allowing clear explanations for users. Additionally, its efficiency and ability to handle non-linear relationships in complex medical data make it ideal for resource-constrained environments. This prioritizes user understanding and efficient chatbot operation within the limitations of the platform.

Our current approach utilizes a general disease prediction dataset. Future work involves evaluating these algorithms on a conflict zone-specific dataset if available, incorporating additional features relevant to such settings (e.g., limited access to medication, psychological distress), and exploring ensemble methods that combine the strengths of multiple algorithms for potentially higher accuracy. Additionally, integrating real-time data streams from wearable devices could further enhance the chatbot's capabilities.

Keyword(s): Machine Learning; Disease Prediction; Decision Trees; Support Vector Machine; KNN.

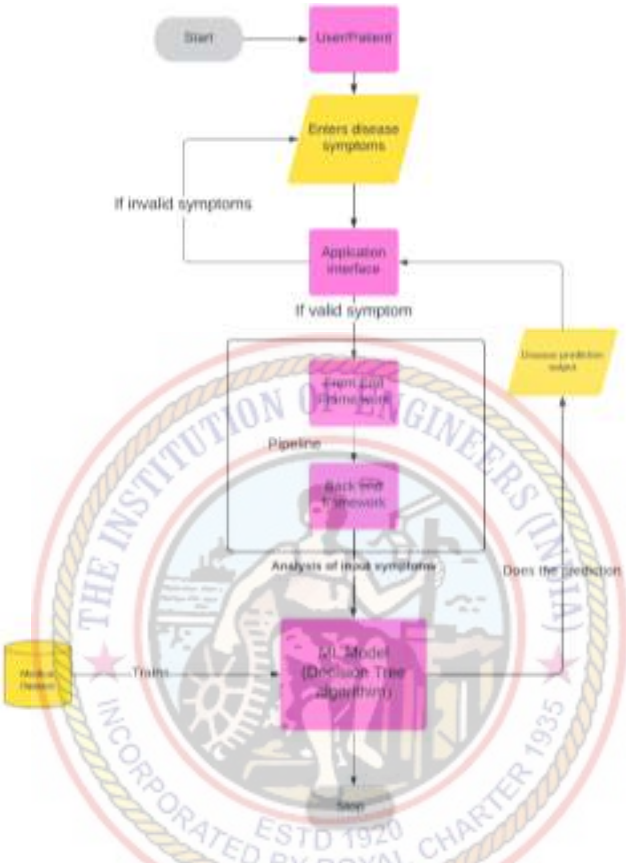


Figure 1: Flowchart of disease prediction model

References

1. Alanazi R. Identification and Prediction of Chronic Diseases Using Machine Learning Approach. J Healthc Eng. 2022 Feb 25;2022:2826127. doi: 10.1155/2022/2826127. PMID: 35251563; PMCID: PMC8896926.
2. Abdulazeem H, Whitelaw S, Schauberger G, Klug SJ. A systematic review of clinical health conditions predicted by machine learning diagnostic and prognostic models trained or validated using real-world primary health care data. PLoS One. 2023 Sep 8;18(9):e0274276. doi: 10.1371/journal.pone.0274276. PMID: 37682909; PMCID: PMC10491005
3. Noor, Nahiyan & Anwar, Mohammad & Dey, Mrinmoy. (2019). Comparative Study Between Decision Tree, SVM and KNN to Predict Anaemic Condition. 10.1109/BECITHCON48839.2019.9063188.
4. Absar N, Das EK, Shoma SN, Khandaker MU, Miraz MH, Faruque MRI, Tamam N, Sulieman A, Pathan RK. The Efficacy of Machine-Learning-Supported Smart System for Heart Disease Prediction. Healthcare. 2022; 10(6):1137.

Classification of Indian Snakes using an Attention-based Deep Neural Network

Ayush Kumar[1], Rajesh Mukherjee[2], Minakshi Sarkar[3] and Bidesh Chakraborty[4]

[1]Department of Computer Science and Engineering (Data Science), Haldia Institute of Technology, Haldia, India
ayushjha4wd@gmail.com
[2]Department of Computer Science and Engineering, Haldia Institute of Technology, Haldia, India
rajeshmukherjee.cse@gmail.com
[3]Department of Computer Science and Engineering, Haldia Institute of Technology, Haldia, India
minakshi.sarkar42@gmail.com
[4]Department of Computer Science and Engineering, Haldia Institute of Technology, Haldia, India
bidesh.me@hithaldia.ac.in

Extended Abstract

There are a wide range of snake species that are native to India. Many of them are venomous, such as cobras and kraits, while others, like keelbacks and pythons, are non-venomous. Snakes can often be found in human habitats, especially in villages. The classification of snakes is not easy because they often share similar features. It is essential to identify whether they are venomous or non-venomous, as this knowledge is crucial for reducing snakebite-related deaths by providing appropriate medicine in a timely manner [1]. An analysis of 60 articles, encompassing 363 victims, revealed that 88% of snakebites were from venomous snakes, while 12% were from non-venomous ones. When necrosis was present, 15.2% of cases reported no infection, even when antibiotics were not used, suggesting potential antibacterial properties in snake venom.

Hence, in this work, we have proposed a deep learning model utilizing Xception [2] and the Convolutional Block Attention Module (CBAM) [3] to automatically categorize snakes by their species name and determine whether the snake is venomous or nonvenomous (Figure 1). We initially compiled a comprehensive dataset of Indian snakes, consisting of nearly 3000 images distributed across fifteen distinct classes and this dataset is used in the proposed model (Figure 2). Our proposed methodology demonstrated effectiveness in efficiently integrating information from the target domain, with a specific focus on extracting crucial features from significant regions. This enhancement significantly bolstered the model's capability for feature extraction. This study presents significant contributions across three main areas: 1. It introduces an Indian snake dataset consisting of nearly 3000 images distributed among fifteen diverse classes. 2. Initially, a CNN model was developed from scratch. Subsequently, pre-trained models were utilized with fine-tuning for training purposes. This dataset was split into training and testing sets at an 85%:15% ratio, with 2482 images used for training and 446 images reserved for testing. For classification tasks, we employed a deep learning model incorporating an Exception and CBAM architecture. The model was trained using the Adam optimizer with a learning rate set to 0.0001, aiming to minimize the categorical cross entropy loss function. During training, data was processed in batches of 16 samples per iteration. Among the various pre-trained models evaluated, Xception was identified as the most suitable for our dataset. Additionally, unfreezing the last 5 blocks of Xception has yielded the best results. 3. The integration of Xception with CBAM, where the last 5 blocks of Xception are unfrozen, represents a novel approach. This allows the model to concentrate on relevant regions within input images, thereby enhancing its ability to discern intricate patterns and relationships. Table 1 describes the classification report which provides key metrics (precision, Recall, and F1 score) to evaluate the performance of our classification model. We've allocated unique numerical mappings to each snake species for classification purposes. For instance, BLACK HEADED ROYAL SNAKE NON-VENOUMOUS corresponds to class 0, while COBRA VENOUMOUS is associated with class 1 and so on. The model achieves an outstanding validation accuracy of 86%, surpassing the state-of-the-art models listed in Table 2. This outcome underscores the efficacy of employing attention mechanisms in deep learning models for intricate classification tasks, such as distinguishing between diverse snake species. As our research contributes to the broader field of wildlife conservation and public safety, the findings pave the way for enhanced snake identification systems, facilitating timely and accurate responses in snake-related emergencies. Despite the notable success achieved, future endeavors could explore further optimizations and refinements to elevate the model's performance, ensuring its robustness across various environmental conditions and species variations.

Keyword(s): CNN; Xception; CBAM; Data Augmentation; Indian Snake.

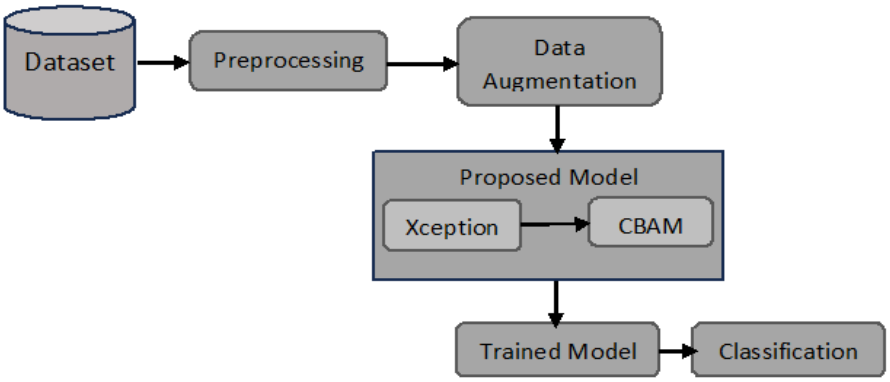


Figure 1: Overall system architecture

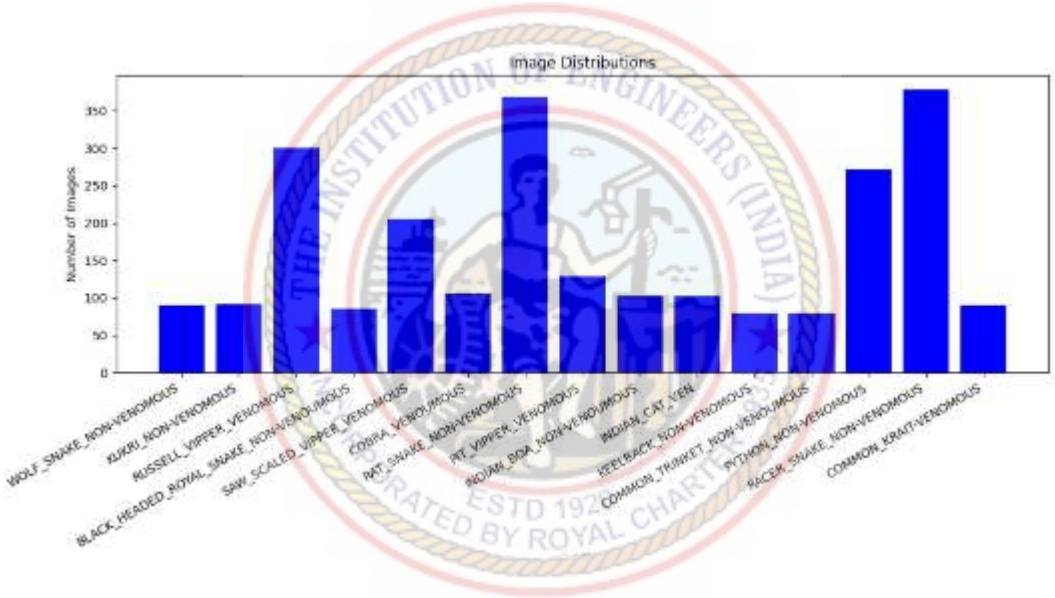


Figure 2: Number of image data breed wise

Table 1: Classification report

Different snake species	Precision	Recall	F1 Score	Support
Class 0	0.88	0.94	0.91	16
Class 1	0.94	0.79	0.86	19
Class 2	0.88	0.94	0.91	16
Class 3	0.68	0.87	0.76	15
Class 4	0.88	0.78	0.82	18
Class 5	0.75	0.63	0.69	19
Class 6	0.73	0.73	0.73	15
Class 7	0.80	0.71	0.75	17
Class 8	0.89	0.70	0.78	23
Class 9	0.87	0.98	0.92	48
Class 10	0.88	0.85	0.86	67
Class 11	0.85	0.83	0.84	66
Class 12	0.93	0.96	0.94	53
Class 13	0.89	0.89	0.89	37
Class 14	0.81	1.00	0.89	17
accuracy			0.86	446
macro avg	0.84	0.84	0.84	446
weighted avg	0.86	0.86	0.86	446

Table 2: Performance of deep learning models on validation data

Model Network	LOSS	ACC	F1 Score	PRECISION	RECALL
CNN (from scratch)	2.42	0.33	0.24	0.48	0.16
DenseNet201	1.20	0.69	0.63	0.75	0.55
InceptionV3	1.06	0.68	0.67	0.78	0.60
MobileNetV2	1.04	0.73	0.72	0.83	0.64
InceptionResNetV2	1.21	0.74	0.76	0.80	0.74
Xception	0.86	0.75	0.74	0.79	0.70
Xception + ChannelAttention	1.11	0.79	0.80	0.82	0.79
Xception + CBAM	1.26	0.80	0.79	0.82	0.77
(Unfreeze last 4 blocks of Xception)					
Xception + CBAM (Unfreeze last 5 blocks of Xception)	**1.20**	**0.86**	**0.85**	**0.87**	**0.85**

References

1. Vasmatkar, Mrugendra, et al. "Snake species identification and recognition." 2020 IEEE Bombay Section Signature Conference (IBSSC). IEEE, 2020.
2. Chollet, François. "Xception: Deep learning with depthwise separable convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
3. Woo, Sanghyun, et al. "Cbam: Convolutional block attention module." Proceedings of the European conference on computer vision (ECCV). 2018.

Machine Learning about Electrical Engineering Self-Healing Grid and Smart Power Grid

Siddhartha Sow[1], Amit Sinha[2], and Bhaskar Roy[2]

[1]Department of Computer Science & Engineering, Asansol Engineering College, Asansol, India
siddharthasow007@gmail.com
[2]Department of Artificial Intelligence & Machine Learning, Asansol Engineering College, Asansol, India
sinha.amit84@gmail.com,bhaskar.ec11@gmail.com

Extended Abstract

The integration of machine learning algorithms into smart grid management systems is explored in this research, which has the potential to completely transform the way energy is distributed. It offers a thorough analysis of the characteristics, advantages, and difficulties of smart grid management systems and delves into the several machine learning methods used in them, including support vector machines, decision trees, and neural networks. There are numerous benefits to using machine learning algorithms in smart grid management. First off, by maximising energy distribution and reducing waste, they provide cost savings. They also improve efficiency and reliability by allowing the grid to react dynamically to changes in supply and demand for energy. Moreover, energy demand forecasting, predictive maintenance, and client usage identification can all be made easier by machine learning algorithms. Nevertheless, there are difficulties in integrating machine learning into smart grid management. Scalability, privacy, and data security are important issues that need to be taken care of. Smart grids require strong cybersecurity safeguards to reduce the hazards associated with their heavy reliance on wireless technology. In order to improve system security overall, machine learning can be extremely helpful in data authentication as well as the detection of unusual activity, infiltration, cyberattacks, and harmful behaviours. The increased difficulty of preserving power system stability in the face of distributed and renewable energy sources is highlighted in the study. With the integration of AI and IT, conventional power systems can be transformed into adaptive smart grids. The paper highlights how machine learning can enable predictive maintenance, energy demand forecasting, and identification of customer consumption trends - critical capabilities for optimizing smart grid operations. Predictive maintenance leverages machine learning to analyze data from sensors and equipment to predict failures before they occur, allowing pre-emptive repairs and minimizing downtime. Energy demand forecasting employs machine learning models to accurately predict future energy needs based on historical data and various factors like weather, enabling better resource planning and load balancing. Furthermore, analysing customer consumption patterns through machine learning provides valuable insights for tailoring energy services, implementing demand-side management strategies, and driving energy conservation efforts. Collectively, these machine learning applications enhance grid reliability, reduce waste and costs, and enable a more efficient and responsive smart grid that can dynamically adapt to evolving energy needs. The study also highlights the significance of machine learning algorithms in assisting with predictive maintenance, energy demand forecasting, and customer consumption trend analysis. These features improve energy conservation and management while also boosting the smart grid's general responsiveness and efficiency. The research addresses various obstacles to the practical deployment of machine learning in smart grid management, despite the significant potential benefits. These include the requirement for strong cybersecurity defences, legal compliance, the complexity of technology, and interoperability issues. For machine learning-based smart grid management solutions to be widely adopted and optimised, several challenges must be overcome. The study concludes by highlighting the revolutionary potential of machine learning in transforming smart grid management. Advanced algorithms and approaches can be used to improve the responsiveness, efficiency, security, and stability of smart grids. However, realising the full potential of this technology and enabling its widespread implementation in the energy sector require tackling the accompanying obstacles, which include data security, privacy, scalability, and regulatory compliance.

Keyword(s): Smart grid; Energy distribution; Machine learning algorithms; Support vector machines (SVM); Neural networks, Decision trees; Reliability improvement; Energy efficiency.

References

1. Chen, Yijie, et al.: Mathematical model and performance calculation of a double-wishbone independent suspension. Journal of Physics: Conference Series. Vol. 1074. No. 1. IOP Publishing (2018)
2. Mu, Mingfei, et al.: Study on a new gasoline particulate filter structure based on the nested cylinder and diversion channel plug. Energies 12.11: 2045 (2019)
3. Chen, Y., Zhang, Y., Du, F., Gui, P., & Nie, W.: Mathematical model and performance calculation of a double-wishbone independent suspension. In Journal of Physics: Conference Series (Vol. 1074, No. 1, p. 012057). IOP Publishing (2018)

A Review on Sentiment Analysis Algorithm for Determining the Best Stock Market Investing Sector in Difficult Economic Times

Sk Majar Ali[1], Sandipan Biswas[1], Bhaskar Roy[2], Amit inha[2], Trisha Das[3]

[1]Department of Computer Science, Gitaram Institute of Management, West Bengal, India
majar.ali.durgapur@gmail.com,sandipan_diet@rediffmail.com
[2]Department of Artificial Intelligence & Machine Learning, Asansol Engineering College, West Bengal, India
bhaskar.ec11@gmail.com,sinha.amit84@gmail.com
[3]Department of CSE, Asansol Engineering College, West Bengal, India
trishadas1805@gmail.com

Extended Abstract

Since the begining of civilization, the concept of money and economic systems has been intertwined with human progress and societal development. Over time, currencies have evolved, and economic expansion has become a defining factor in shaping the future of nations and the world. In the modern era, the ebb and flow of stock markets have emerged as a key indicator of economic health and a driver of global economic trends.Despite the apparent unpredictability of stock markets, researchers have shown that with sufficient historical data, it is possible to make projections about future stock and share prices. This ability to forecast, albeit with varying degrees of accuracy, is crucial in understanding and navigating the complexities of modern economies.The importance of stock markets in today's world cannot be overstated. They serve as a barometer of economic activity, reflecting the collective sentiment of investors and impacting everything from consumer confidence to government policy. Stock market speculation, while inherently risky, can also be highly profitable for those who can accurately forecast future market behavior.The model utilized by the authors of this study represents a significant advancement in the field of stock market forecasting. By leveraging historical data and employing sophisticated analytical techniques, the model can reliably predict stock market movements with a small margin of error. This capability has profound implications for investors, policymakers, and economists alike, as it offers a glimpse into the future trajectory of economic growth and prosperity.In addition to its predictive capabilities, the study also delves into the realm of sentiment analysis. By analyzing the tweets of political leaders from various nations, the study seeks to identify correlations between sentiment regarding the rise in drug and medication use for COVID-19 treatment and stock market behavior. This innovative approach sheds light on the interconnectedness of political discourse, public sentiment, and economic activity, offering valuable insights into the factors that influence stock market trends.Furthermore, the study explores how stock markets may respond to external shocks such as terrorist attacks or epidemics. By examining the behavior of stock markets in the aftermath of such events, the study aims to identify patterns and trends that can inform future research and policy decisions. This analysis not only enhances our understanding of the dynamics of stock markets but also provides a framework for evaluating the potential impact of unforeseen events on economic stability.In conclusion, the study represents a comprehensive examination of the role of stock markets in the global economy. By combining advanced analytical techniques with a nuanced understanding of economic trends, the authors have produced a groundbreaking analysis that has implications for investors, policymakers, and researchers alike.

Keyword(s): Sentiment Analysis; Stock Market; Social media; Twitter, Economic downfall; Machine learning, Pandemic.

References

1. Biswas,S., Ghosh, A. K., Chakraborty, S., Roy, S., and Bose,R .: Scope of Sentiment Analysis on News Articles Regarding Stock Market and GDP in Struggling Economic Condition. International Journal of Emerging Trends in Engineering Research.8(7),3594-3609 (2020)
2. Biswas, S., Ghosh, S.: A Comparative Study into Stock Market Prediction Through Various Sentiment Analysis Algorithms. In: Tavares, J.M.R.S., Dutta, P., Dutta, S., Samanta, D. (eds) Cyber Intelligence and Information Retrieval. Lecture Notes in Networks and Systems, vol 291. Springer, Singapore(2022)https://doi.org/10.1007/978-981-16-4284-5_11

3. Biswas, S., Ghosh, S. , Roy,S.: A Sentiment Analysis on Tweeter Opinion of Drug Usage Increase by TextBlob Algorithm Among Various Countries During Pandemic. Int.J.HIT.TRANSC:ECCN,6(2A) , 1-9 (2020)
4. Biswas, S., Ghosh, S.: Drug Usage Analysis by VADER Sentiment Analysis on Leading Countries. Mapana Journal of Sciences , 21(3) (2022)

The Next Gaming Frontier: Applying Machine Learning (ML) in E-sports

Diptesh Maji[1], Amit Sinha[1], and Bhaskar Roy[1]

[1]Department of Artificial Intelligence & Machine Learning, Asansol Engineering College, Asansol, India
dipteshmaji1@gmail.com,sinha.amit84@gmail.com,bhaskar.ec11@gmail.com

Extended Abstract

Machine learning (ML) is becoming a strong technique with various intriguing applications in the quickly expanding e-sports business. Using ML approaches has many potential benefits for players, teams, developers, and fans as competitive video gaming continues to gain popularity worldwide. Game analytics and strategy optimization are two important areas where machine learning can have an impact. Through the analysis of vast amounts of data produced by professional gaming, including as player inputs, video recordings, and comprehensive telemetry, sophisticated machine learning algorithms can recognize trends, forecast results, and suggest the best courses of action for both individuals and groups. In the cutthroat realm of e-sports competitions, this can offer a major competitive advantage. The greatest professional players' placement, ability utilization, economy management, and other factors that influence their strategic decision-making can be methodically studied by ML models. The tactical strategies that eluded human perception might be revealed by the insights that machine learning algorithms unearth. Additionally, ML seems promising for improved player performance analysis. It provides coaches and teams with an unbiased, data-driven way to assess the advantages and disadvantages of individual and team development. Methods such as motion tracking and computer vision can evaluate player execution mechanics, accuracy, response times, and other aspects. This analytical power can be used to create personalized training plans. ML can identify communication patterns, opportunities for coordination enhancement, and gaps in synergy at the team level. Beyond rivals, ML has the potential to completely change how spectators and supporters experience e-sports. Broadcasts may be overlayed with memorable moments and personalized insights thanks to automated content curation and highlight identification. Situational assessments to improve commentary and live win probabilities could be produced using real-time machine learning prediction models. To generate dynamic gameplay experiences, developers may even investigate using machine learning for procedural content production. However, concerns about bias, privacy, and transparency are brought up by the growing use of ML in e-sports. Unwanted biases favouring play styles or techniques may be displayed by algorithms. Privacy problems are raised by the mandatory collecting of player data. To maintain the interpretability and equity of ML decision-making, governance is required. All things considered, ML applications in e-sports are a promising new field. This technology has the potential to significantly improve watching experiences, optimize performance, and spur growth for this rapidly expanding phenomenon if it is developed properly and under human supervision. Machine learning (ML) has found a home in e-sports, where it can revolutionize the way competitive video games are designed, coached, played, and broadcasted. The state-of-the-art in ML deployment for e-sports is presented in this survey along three major parameters. First, we discuss current advances in deep learning and reinforcement learning on large-scale gameplay datasets for game analytics, strategy mining, and decision assistance. Subsequently, multimodal methods that combine sensor, video, and telemetry data are investigated for thorough player tracking and performance assessment to provide customized coaching insights. We evaluate our unified framework using data from the most recent major e-sports competitions, showing notable improvements over human-driven analysis in terms of fan experience metrics, player improvement suggestions, and strategy evaluation.

Keyword(s): Game analytics and strategy optimization; recognize trends; predict outcomes; Optimal actions; Human perception; Probability predictions.

References

1. Hamari, J., and Sjöblom, M.: Data science in esports: Exploring the role of game data in esport analytics. Communications in Computer and Information Science, 748, 19-30 (2017)
2. Yang, Y. and Mani, S.: Predicting the outcomes of professional Dota 2 matches. In 2018 IEEE Conference on Computational Intelligence and Games (CIG), pp. 1-8, IEEE (2018)
3. Lee, H., Han, S., Lee, J., and Choi, M.: Deep reinforcement learning for esports analytics. In Proceedings of the 2nd International Conference on Big Data and Internet of Things, pp. 83-88, 2019.

4. Ma, Y., Zheng, L., Zhang, X., and Lyu, M. R.: A machine learning approach to hero recommendation in Dota 2. In Proceedings of the 26th International Conference on World Wide Web Companion, pp. 1551-1558.
5. Smith, M.: Predicting Success in Esports: Network Effects on Team Performance. In Proceedings of the 13th International Conference on Social Computing, Behavioural-Cultural Modeling and Prediction and Behaviour Representation in Modeling and Simulation, pp. 39-48.
6. Garcia, R. S., Fernandez, A. M., and Diaz, F. R.: Analysis and prediction of League of Legends player behaviour using machine learning techniques. In IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1077-1084, IEEE (2018)

Revolutionizing Greenhouse Farming: Harnessing Advanced Machine Learning for Fully Automated Crop Cultivation

Subhrodeep Banerjee[1], Amit Sinha[1], and Bhaskar Roy[1]

[1]Department of Artificial Intelligence & Machine Learning, Asansol Engineering College, Asansol, India
subhrodeepbanerjee2002@gmail.com,sinha.amit84@gmail.com,bhaskar.ec11@gmail.com

Extended Abstract

Plant Health Monitoring System Automation is a crucial component that directly affects crop output and quality in contemporary greenhouses. Automating plant health monitoring requires machine learning techniques to enable early problem detection and prompt intervention. Plant health indicators including leaf colour, size, and texture are constantly monitored by these algorithms, which are positioned strategically throughout the greenhouse using an array of sensors and cameras. Plant appearance changes subtly, which could be signs of disease, nutrient deficits, or other stressors. These minute variations can be recognized by algorithms that use cutting-edge computer vision and pattern recognition techniques [1]. An early and focused response to plant health problems is ensured by the system's ability to automatically initiate the proper reactions when anomalies are recognized. Examples of these reactions include modifying the irrigation system's nutrient levels or turning on targeted pesticide spraying systems. Smart irrigation systems are essential for running a greenhouse successfully. Irrigation systems are optimized using machine learning algorithms, which guarantee effective water use and preserve the best growth circumstances possible for each type of plant [2,3]. To establish the ideal irrigation schedule, these algorithms examine real-time sensor data, such as soil moisture content, weather predictions, and plant water needs. The algorithms dynamically modify the frequency, length, and volume of water delivered to various zones inside the greenhouse by taking into account variables including plant growth phases, soil properties, and ambient conditions. This method of precision irrigation reduces the chance of over- or underwatering, encourages healthy plant growth, and conserves water resources. Dynamic climate control ensures optimal plant growth by leveraging machine learning algorithms that continuously monitor temperature, humidity, and CO2 levels from sensor data. These algorithms automatically adjust heating, cooling, and ventilation systems in real time to maintain the precise environmental conditions required for each crop [4]. Assistive maintenance plays a vital role in preventing costly breakdowns and ensuring uninterrupted operations. Machine learning models analyse sensor data across greenhouse systems like HVAC, lighting, and irrigation to detect early signs of potential equipment failures or degradation through advanced pattern recognition and anomaly detection techniques. This enables proactive scheduling of preventive maintenance, minimizing downtime. The harvesting process has been revolutionized through the integration of robotics and computer vision algorithms trained on extensive datasets of fruit and vegetable images. These algorithms can accurately identify ripeness based on visual cues like colour and size, guiding robotic arms to precisely locate and harvest ready-to-pick produce. Planning and rotating crops to maximize yields, preserve soil health, and guarantee long-term sustainability in greenhouse operations, efficient crop planning and rotation techniques are crucial. Machine learning algorithms are essential for optimizing these methods because they analyse soil conditions, market demand, historical crop data, and other environmental aspects Crop rotation techniques are continuously improved through the use of machine learning algorithms that recognize trends in data related to weather, soil, and crop performance [5]. Greenhouse operators benefit from this data-driven approach's capacity to make well-informed decisions, flexibility, and a competitive edge in a market where environmental concerns are paramount. By automating every part of greenhouse management, including climate control, harvesting, maintenance, and health monitoring, machine learning can be integrated to maximize productivity and reduce resource use. Machine learning will play a critical role in sustainable greenhouse management as technology develops, spurring innovation and influencing the direction of agriculture.

Keyword(s): Plant health monitoring; Machine learning algorithms; Computer vision; Pattern recognition; Disease detection; Smart irrigation management.

References

1. Velazquez, G., Roberto, S. et al.: A review on hydroponics and the technologies associated for medium-and small-scale operations. Agriculture vol. 12 no. 5 pp 646 (2022)
2. Joshitha. C., Kanakaraja. P., Kumar, K.S., . kanksha, P.A, Satish, G.: An Eye on Hydroponics: The IoT Initiative. 7th International Conference on Electrical Energy Systems, ISBN: 978 1 7281 7612 3, IEEE, DOI: 10 1109 ICEES51510 (2021)
3. Gour, M.S., Reddy, V.: IoT Based Arming Techniques in Indoor Environment: A Brief Survey. Fifth International Conference on Communication and Electronics Systems (ICCES 2020), IEEE Xplore, ISBN 978 1 7281 5371 (2020)
4. Srivani. P., Devi. C.Y.: A controlled environment agriculture with hydroponics: variants, parameters, methodologies and challenges for smart farming. Fifteenth International Conference on Information Processing, ISBN: 978 1 7281 3807 7, IEEE (2019)
5. Samuel, A.L: Some Studies in Machine Learning Using the Game of Checkers. IBM J. Res. Dev. 44 206 226 (1959)
6. Karaman. S., Yazarel. S., Gokalp. Z.: Automation applications in greenhouse. Vol. 8 Issue 16 pp 154 159 (2019)

Beyond the Visual: Leveraging Multimodal Data for Enhanced Flight Classification

Ayush Das[1], Amit Sinha[2], and Bhaskar Roy[2]

[1]Department of Computer Science & Engineering, Asansol Engineering College, Asansol, India
ayushdas24406@gmail.com
[2]Department of Artificial Intelligence & Machine Learning, Asansol Engineering College, Asansol, India
sinha.amit84@gmail.com,bhaskar.ec11@gmail.com

Extended Abstract

Traditional methods for flight identification often rely on visual data, such as aircraft silhouettes or radar images. However, these approaches are prone to limitations stemming from factors like adverse weather conditions, sensor noise, and variations in aircraft appearance. This paper presents an exploration into the potential of multimodal data fusion coupled with machine learning for enhancing flight classification robustness. We propose a novel framework integrating diverse data sources beyond the visual spectrum. This includes leveraging Automatic Dependent Surveillance-Broadcast (ADS-B) information containing flight plan details, audio recordings capturing engine sounds, and real-time environmental data. By amalgamating these varied features within a unified machine learning architecture, the system aims to transcend the limitations associated with individual data sources. Our experimentation revolves around a custom-built dataset spanning multiple aircraft types and environmental conditions. We conduct comparative analyses, pitting the performance of models trained on individual data modalities (visual, ADS-B, audio, environmental) against a multimodal fusion model. The outcomes underscore the efficacy of the multimodal approach, which consistently outperforms models relying solely on single data sources. Traditional methods for flight identification predominantly hinge on visual data, be it aircraft silhouettes or radar images. Nonetheless, these methods are susceptible to myriad limitations, ranging from inclement weather conditions obstructing clear visual cues to sensor noise introducing inaccuracies in detection, not to mention the inherent variability in aircraft appearances. In response to these challenges, this paper advocates for a paradigm shift towards a more holistic approach to flight classification—one that capitalizes on the complementary nature of diverse data sources. Our proposed framework stands at the intersection of multimodal data fusion and machine learning, presenting a robust solution to enhance flight identification accuracy and resilience. By harnessing a spectrum of data types encompassing visual, ADS-B, audio, and environmental cues, our system strives to build a comprehensive understanding of the flight environment. This holistic perspective enables the system to discern intricate patterns and relationships that might elude conventional unimodal approaches. Through a series of experiments conducted on a meticulously curated dataset, we demonstrate the superiority of the multimodal fusion model over its unimodal counterparts. Leveraging machine learning algorithms adept at handling complex data relationships, our framework exhibits remarkable performance gains, showcasing its potential for real-world deployment in flight identification systems. In conclusion, this research underscores the transformative impact of multimodal data fusion coupled with machine learning in enhancing the robustness and reliability of flight classification systems. By embracing a multifaceted approach that integrates diverse data sources, we pave the way for more resilient and accurate flight identification mechanisms capable of thriving in diverse operational contexts. This paradigm shift holds immense promise for revolutionizing the landscape of flight identification, offering a pathway towards more effective and adaptive systems in the aviation domain. Our findings highlight the pivotal role of multimodal data fusion in addressing the complexities of flight identification, offering a pathway towards more adaptive and effective systems in the aviation domain. This paradigm shift holds immense promise for revolutionizing the landscape of flight identification, ensuring enhanced safety, security, and situational awareness in both civilian and defines contexts alike.

Keyword(s): Flight identification; Multimodal data fusion; Machine learning.

References

1. Stein, G. P.: Multisensor Data Fusion. CRC Press (2005)

2. Jain, A. K., and Kundu, M. K.: Machine Learning Techniques for Multimodal Integration. ACM Computing Surveys, 37(2), 109-139 (2004).
3. Bishop, C. M.: Pattern Recognition and Machine Learning. Springer (2006)
4. Hinton, G., and Salakhutdinov, R. R.: Reducing the Dimensionality of Data with Neural Networks. Science, 313(5786), 504-507 (2006).
5. ICAO.: Manual on Automatic Dependent Surveillance-Broadcast (ADS-B). International Civil Aviation Organization (2019)
6. ITU-R.: Technical characteristics of Automatic Dependent Surveillance-Broadcast (ADS-B) and Multilateration (MLAT) for aeronautical applications. International Telecommunication Union (2019)
7. Schultz, T., and Goussev, V.: Aircraft Sound Analysis and Synthesis Toolbox. Proceedings of the IEEE, 94(9), 1730-1746 (2006).
8. Zheng, Y., and Feng, D.: Aircraft engine noise feature extraction and identification using wavelet packet transform and neural network. Expert Systems with Applications, 38(5), 5756-5763 (2011).
9. Borghys, D., and Poli, R.: A review of machine learning applications in air traffic management. Journal of Air Transport Management, 17(2), 110-116 (2011).
10. Duraisamy, P., and Jayaraman, R.: Airplane Recognition Using Machine Learning Techniques. International Journal of Scientific & Engineering Research, 5(10), 1297-1303 (2014)

Exploring Bengali Image description through the combination of diverse CNN Architectures and Transformer Decoders

Biswajit Patra[1], and Dakshina Ranjan Kisku[2]

[1,2]Department of Computer Science and Engineering, National Institute of Technology Durgapur, Durgapur - 713209, India
{bp.21cs1102@phd.nitdgp.ac.in, drkisku.cse@nitdgp.ac.in}

Extended Abstract

Image description generation using a computer is a complex process as it aims to produce concise and human narrated natural descriptions concerning the visual content of an image. This task leverages its functionalities by employing a computer vision model to understand the image and an NLP model to generate its descriptions. Research works published in this particular domain are mostly based on the English language, where CNN and RNN [1] are used as encoder and decoder models with attention-based [2] approaches as enhancement mechanisms. Currently, Bengali language stands as the sixth most-spoken native language and the seventh most widely spoken language. Despite the growing interest in language research, the Bengali context has received very little attention in comparison to other resource-rich languages like English. The objective of this study is to address the research gap in describing image visuals in Bengali by introducing a novel image captioning approach. The proposed approach utilizes diverse and state-of-the-art Convolutional Neural Network (CNN) architectures as encoder in conjunction with a self-attention-based transformer [3] as decoder in the encoder-decoder framework. The proposed model harnesses the strengths of various CNN architectures, including EfficientNetV2s, ConvNeXt-Small, and InceptionResNetV2 in conjunction with Transformer decoders to produce accurate and contextually relevant captions. Further, the integration of Bengali text-to-speech synthesis in the proposed framework provides the audio description of the relevant image visuals to support visually impaired people in a better manner and offers a convenient alternative for accessing the content to multitasking users in Bengali-speaking regions.

The proposed methodology has emerged to overcome the shortcomings related to computational efficiency and understanding visual context in the image in order to produce viable and most accurate image captions. The proposed methodology employs pre-trained CNN model to extract the distinctive image features from the last pooling layer and feed this features set to the transformer-based decoder with a mapping of dimensions for generating the descriptions. Since the feature map and embedded word are the input in the transformer encoder and transformer decoder, feature embedding and word embedding are done in their input embedding layer. Incorporating positional encoding addresses the challenges related to position tracking encountered with the visual attention model approach. The self-attention mechanism aids in focusing on the contextual information within the image, while the masked self-attention layer filters the essential phrases in the Bengali descriptions. Since many benchmark datasets of image captioning are available in English langauge, however no such widely used datasets are found in Bengali language, it follows the widely used Flickr8k dataset [2]. The study utilizes a Bengali image captioning dataset called Ban-Cap, to train and evaluate the proposed model. It consists of five annotated English and Bengali descriptions for 8091 images, thus forming 40455 image description pairs. Further image and Bengali description pairs are used for training and assessment. The investigation aims to assess the quality of the model-generated description concerning human-annotated ground truth description based on several evaluation metrics. The proposed model attains BLEU-1, BLEU-2, BLEU-3, and BLEU-4 scores of 0.66, 0.59, 0.44 and 0.26 respectively. Thus achieves a competitive advantages while comparing with other state-of-the-art models for Bengali description generation. Moreover, findings revealed the successful contribution of advanced deep learning models and architectures for Bengali image caption generation problems in the context of assisting the Bengali-speaking visually impaired population.

Keyword(s): Bengali description, CNN, Self-attention, Transformer, Bengali text-to-speech synthesis

Diagram of Proposed Methodology

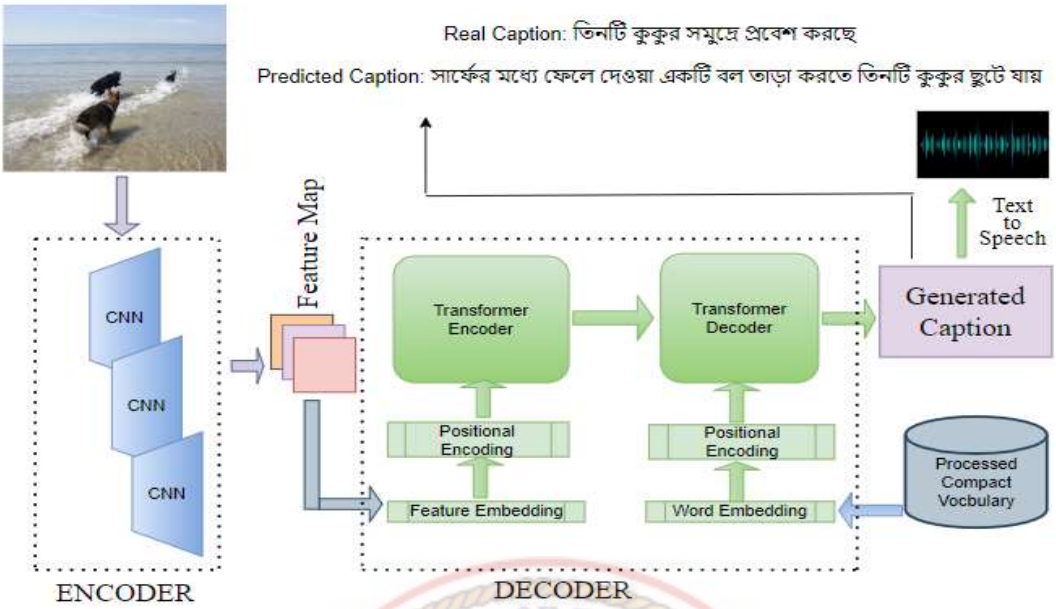


Figure 1: Overview diagram of the proposed methodology.

Table 1: Comparative Analysis with other SOTA models of Bengali description generation on Ban-Cap Dataset

Methodology	BLEU-1	BLEU-2	BLEU-3	BLEU-4
CNN-LSTM Based [1]	0.56	0.35	0.22	0.13
Visual Attention Based [2]	0.58	0.36	0.25	0.14
Proposed Model 1 (EfficientNetV2S)	0.66	0.59	0.44	0.26
Proposed Model 2 (ConvNeXtSmall)	0.65	0.58	0.43	0.25
Proposed Model 3 (InceptionResNetV2)	0.64	0.57	0.42	0.25

References

1. Rahman, M., Mohammed, N., Mansoor, N., & Momen, S. (2019). Chittron: An automatic bangla image captioning system. Procedia Computer Science, 154, 636-642.
2. Ami, A. S., Humaira, M., Jim, M. A. R. K., Paul, S., & Shah, F. M. (2020, December). Bengali image captioning with visual attention. In 2020 23rd International Conference on Computer and Information Technology (ICCIT) (pp. 1-5). IEEE.
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
4. Mishra, S. K., Dhir, R., Saha, S., Bhattacharyya, P., & Singh, A. K. (2021). Image captioning in Hindi language using transformer networks. Computers & Electrical Engineering, 92, 107114.

Auto – PCOS Classification on the Use of the PCOSGen Dataset: An Approach of Moderated Deep Convolutional Neural Network

Chandrika Vengala and Dakshina Ranjan Kisku

Department of Computer Science and Engineering, National Institute of Technology Durgapur, Durgapur – 713209, India
{cv.23cs4124, drkisku.cse@nitdgp.ac.in}

Extended Abstract

PCOS stands for Polycystic Ovary Syndrome. The term 'poly' means 'many' and 'cysts' means 'small follicles, each containing an egg'. Therefore, a Polycystic ovary means 'an ovary with many cysts/follicles'. A follicle is a fluid-filled sac that contains an immature egg. The women with PCOS have a lot of eggs but they never really ovulate and those women are often infertile. Approximately 70% of women worldwide who have PCOS have not received a correct diagnosis, and most of them did not even know that they have this illness. Considering the under diagnoses of PCOS, lack of experts in ultrasound imaging, and overall low patient-physician ratio across globe, there arises a need of modern diagnostic methods to fight PCOS. In challenging circumstances, this computer-aided automated disease detection system helps medical professionals make informed decisions about their patients' conditions. This study proposes a solution to detect Polycystic ovary Syndrome from the ovarian ultrasound images by developing a Convolutional Neural Network (CNN) based image processing method that makes use of moderated Inception modules. With the proposed model the detection accuracy of 77% on the test data has been determined.

There is a limited body of research on this topic, indicating a potential area for further exploration and study. Authors of [1] trained their model and achieved an accuracy of 99.83% on test data and 100% on the training data. But when they trained their model with new unseen dataset their training accuracy went down to 83.33% and testing accuracy to 62.92%. The work discussed in [2], uses a dataset containing 541 samples and it is divided into training and testing subsets containing 364 and 177 samples respectively. The work [3] used a dataset of 80 images with an accuracy of 78.81%. The researches of [4] utilized a dataset comprising only 94 images and mentioned the problems they faced regarding the poor quality of images and limited dataset size. They achieved an accuracy of 70% detection rate.

Based on these studies reported in literature, we can generalize performance of the proposed system that depends on the dataset quality and quantity as well as the evaluation process we follow. So, we used a dataset which is qualitatively and quantitatively efficient. The PCOSGen dataset referenced in [5] is the first of its kind, comprising a good number of ultrasound images collected from various sources, such as YouTube, ultrasoundcases.info, and Kaggle. The dataset contains 3200 images, including both healthy and unhealthy cases, which is bigger than any existing datasets. These ultrasound images have been meticulously annotated with the help of experienced gynaecologists based in New Delhi, India.

We have developed a novel system that leverages the VGG19 architecture as a baseline model, enhancing it with three strategically placed moderated Inception modules integrated between the convolutional layers of our proposed CNN model. Figure 1 illustrates these Inception modules, showcasing the kernel sizes used for convolution and max-pool layers. Our primary rationale for incorporating Inception modules is their ability to include multiple parallel convolutional operations with varying kernel sizes, allowing the network to capture features at various spatial scales. Figure 2 demonstrates the architecture of our proposed system consists of Convolution, Batch Normalization, Dropout, Maxpool, Inception Modules and Dense layers.

We partitioned the dataset by allowing an 80-20 ratio, allocating 80% of the images for model training and reserving the remaining 20% for testing purposes. The experimental results determined an accuracy of 95% on training data and 77% on testing data. Furthermore, there is a scope to improve the results by experimental parameter tuning.

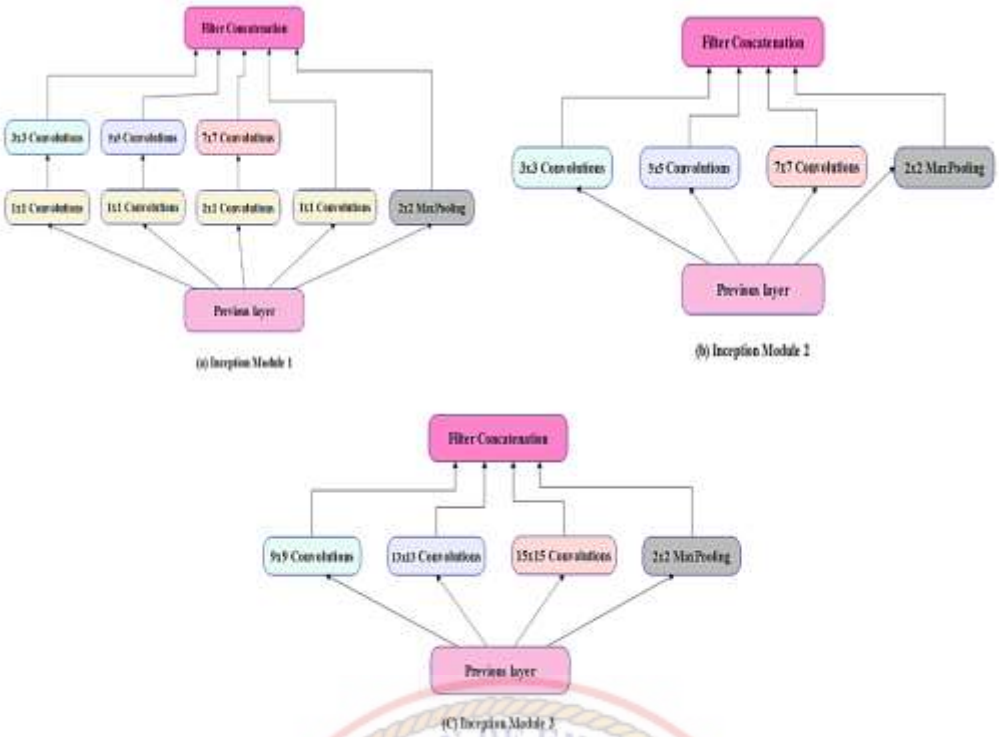


Figure 1: Diagrams of the Moderated Inception Modules used in CNN Model

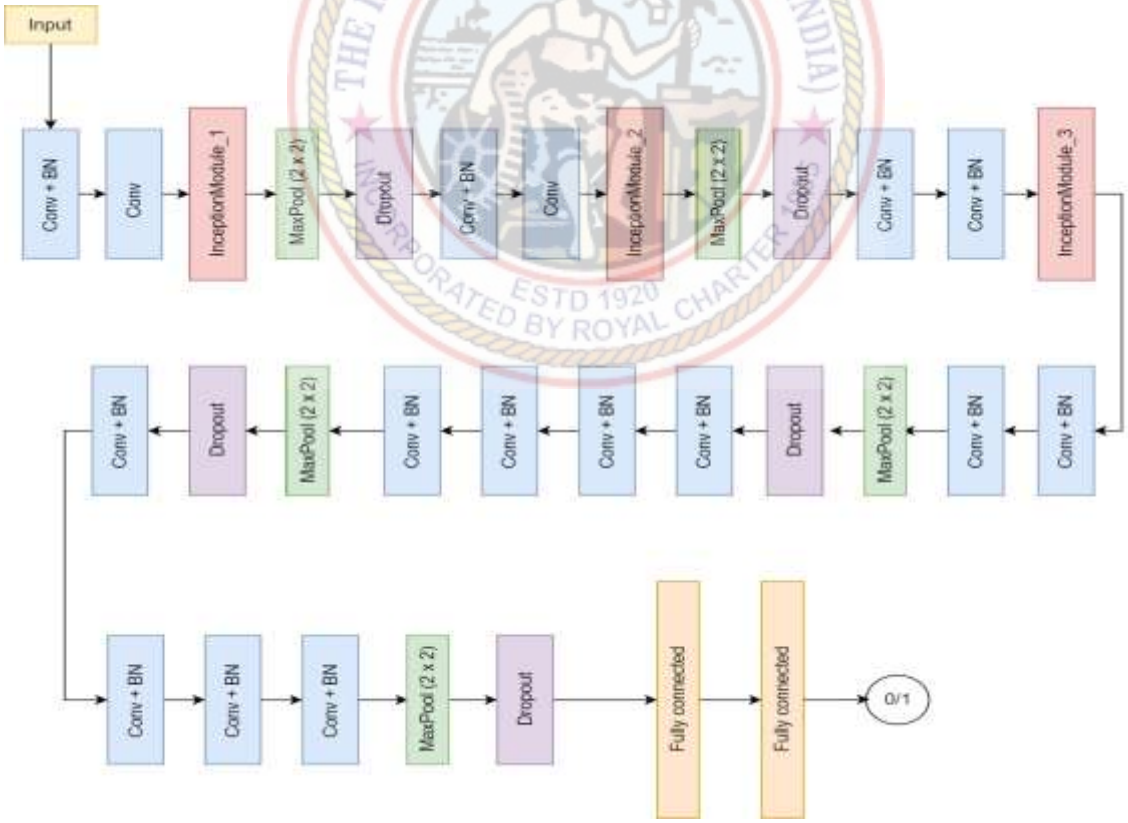


Figure 2: Architecture of the proposed model

Keywords: Computer aided; Convolutional Neural Network (CNN); Image Processing, Inception Modules; Polycystic ovary syndrome; Ultrasound images.

References

1. Alsibai, Mohammed Hayyan, and Seyed Youssef Heydari. "The Effectiveness of CNN in Evaluating Ultrasound Image Datasets: Diagnosing Polycystic Ovary Syndrome (PCOS) as an Example." *Syrian Journal for Science and Innovation* 1.1-Volume 1 (2023).
2. Bhosale, Shubham, Lalit Joshi, and Arun Shivsharanan. "PCOS (polycystic ovarian syndrome) detection using deep learning." *International Research Journal of Modernization in Engineering Technology and Science* 4.01 (2022).
3. Purnama, Bedy, et al. "A classification of polycystic Ovary Syndrome based on follicle detection of ultrasound images." *2015 3rd International conference on information and communication technology (ICoICT)*. IEEE, 2015.
4. Kumari, Sweta. *Classification of pcos/pcod using transfer learning and gan architectures to generate pseudo ultrasound images*. Diss. Dublin, National College of Ireland, 2021.
5. https://zenodo.org/records/10430727

Machine Learning Prediction Model for Multiple Ailments

Ankit Chatterjee[1], Rahul Mandal [1], and Suvobrata Sarkar[1]

[1]Department of Computer Science and Engineering, Dr. B.C. Roy Engineering College, Durgapur, India
{ ankitchatterjee30, rahulmandal1403, suvobrata.sarkar }@gmail.com

Extended Abstract

The development of data-driven diagnostic models to support medical professionals in decision making has raised interest in the detection and evaluation of various ailments. Precise and thoroughly vetted predictions have the potential to significantly enhance personalized care and treatment strategies, ultimately empowering better control over the progression of disease. Machine learning excels in its capacity to uncover challenging-to-discern patterns within extensive, noisy, or intricate datasets. By harnessing the capabilities of various ML algorithms, healthcare practitioners can augment their diagnostic prowess, thereby enhancing patient outcomes and expediting treatment interventions. This paper aims to deliver the most accurate and efficient diagnosis of a given ailment with the aid of several machine learning models viz. Gaussian naive bayes, K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest, and Logistic Regression. Four disease datasets—Diabetes, Parkinson's disease, Breast Cancer, and chronic kidney disease were extracted from Kaggle and were subjected to these ML approaches. However, the outcomes delivered by the ML models do not surpass the insights offered by existing literature, compelling us to integrate ensemble learning into the prediction system. Initially, XGBoost and CatBoost were implemented within ensemble learning, proving quite effective for predicting chronic kidney disease (CKD), but they exhibited comparatively lower performance when employed as standalone models for other diseases. Consequently, an algorithm known as the Voting Classifier was developed to select the optimal model from the ensemble model choices.

The choice of the Voting Classifier was based on its ability to train on an ensemble of multiple models and subsequently utilize the output (class) determined by the models with the highest probability of selecting that class. Hyperparameter tuning was conducted to acquire the optimal values for C, Gamma, and the kernel selection when utilizing the Support Vector Machine model for prediction. The accuracy of each model is confirmed and compared to each other in order to ascertain which prediction model works best. Five evaluation metrics—Accuracy Score, Precision Score, Recall Score, F1 Score, and RMSE—were used to verify the models' performance. The 10-fold cross-validation procedure was used on all models using the same datasets to assess the predictive model accuracy. One benefit of cross validation is that it offers a reliable estimation of the model's generalization performance by employing iterative averaging of results. When assessed separately, chronic kidney disease and breast cancer showed higher cross validation accuracy than other disorders. SVM has an average cross-validation accuracy of 97%, which is higher than other models in the accuracy domain of breast cancer. In contrast, the K-NN model for diabetes disease revealed 83% mean cross-validation accuracy. Parkinson's and chronic kidney disease achieved mean cross validation accuracy of 92% and 100% respectively when performed on ensemble hard voting classifier. The results of this study therefore highlight the enormous potential that data-driven diagnostic models have to transform medical decision-making procedures. Furthermore, Flask-a Python web framework, has been utilized to develop the online interface for the multiple ailment diagnostic system. This interface serves as a user-friendly platform wherein medical professionals and individuals alike can leverage the power of ML algorithms to obtain rapid and accurate diagnoses for a spectrum of ailments.

Keyword(s): Machine Learning Techniques; Parkinson's Disease; Diabetes; Breast Cancer; Chronic Kidney Disease

References

4. Gopisetti L.D., Kummera, S. K, Pattamsetti, S. R, et al. Multiple disease prediction system using machine learning and streamlit. In: Proceedings of the 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 923–931 (2023)
5. Qin, J., Chen, L., Liu, Y., et al. A machine learning methodology for diagnosing chronic kidney disease. IEEE Access. 8, 20991–21002 (2020)

6. Tigga, N. P. and Garg, S. Prediction of type 2 diabetes using machine learning classification methods. Procedia Computer Science. 167, 706–716 (2020)

Improving Heart Disease Predictions: Integrating Linear Regression in Enhanced Medical Analytics

Lipi Mishra [1], Joyjit Patra [1], and Subir Gupta [2]

[1] Department of CSE Dr B. C. Roy Engineering College, Durgapur. West Bengal, India
lipimishra270@gmail.com, joyjitpatra.meet@gmail.com
[2] Department of CSE (AIML), Haldia Institute of Technology, West Bengal, India
subir2276@gmail.com

Extended Abstract

The escalating prevalence of heart disease globally is primarily attributed to lifestyle choices prevalent in contemporary society, such as poor diet, lack of physical activity, and stress. The increasing urbanization and an ageing population further exacerbate this trend. Heart disease, which includes conditions like coronary artery disease, hypertension, and heart failure, remains a leading cause of death and disability, exerting significant pressure on healthcare resources. It necessitates a stronger focus on preventive measures, education, and early detection. Recent developments in automated heart disease prediction underscore a transformative approach to healthcare. Machine learning and data analysis methods actively identify individuals at risk of heart disease early on, leading to prompt medical intervention. These methods use algorithms to find complex data patterns, enhancing diagnostic accuracy, easing healthcare professionals' workloads, and paving the way for customized treatments. Linear regression models, predicting heart disease risks by analyzing variables like age and cholesterol against health records, provide valuable perspectives. However, the simplicity of linear assumptions may overlook the intricate nature of medical data. Performance metrics of classification models, such as precision, recall, and F1 score, reveal the effectiveness of these predictive tools. A model with a precision of 0.79 and a recall of 0.82 indicates a robust predictive capability, further substantiated by an F1 score of 0.83, suggesting a well-balanced model. An overall accuracy rate of 0.82 demonstrates the model's reliability yet indicates room for improvement, particularly in addressing class imbalances. The study shows that automated prediction models are becoming integral to healthcare strategies, offering a means to mitigate costs and enhance the quality of life for individuals predisposed to heart disease. Their implementation represents a significant stride toward a more proactive and personalized medical paradigm.

Keyword(s): Accuracy Plot; Confusion matrix; Count Plot; Precision; Recall;

References

1. Boddu RSK, Karmakar P, Bhaumik A, Nassa VK, Vandana, Bhattacharya S (2021) Analyzing the impact of machine learning and artificial intelligence and its effect on management of lung cancer detection in covid-19 pandemic. Mater Today Proc. https://doi.org/10.1016/j.matpr.2021.11.549
2. Cai A, Zhu Y, Clarkson SA, Feng Y (2021) The Use of Machine Learning for the Care of Hypertension and Heart Failure. JACC Asia 1:162–172. https://doi.org/10.1016/j.jacasi.2021.07.005
3. Dixit J, Prinja S, Jyani G, Bahuguna P, Gupta A, Vijayvergiya R, Kumar R (2023) Evaluating efficiency and equity of prevention and control strategies for rheumatic fever and rheumatic heart disease in India: an extended cost-effectiveness analysis. Lancet Glob Heal 11:e445–e455. https://doi.org/10.1016/S2214-109X(22)00552-6
4. Iyanna S, Kaur P, Ractham P, Talwar S, Najmul Islam AKM (2022) Digital transformation of healthcare sector. What is impeding adoption and continued usage of technology-driven innovations by end-users? J Bus Res 153:150–161. https://doi.org/10.1016/j.jbusres.2022.08.007
5. Jadhav S, Kasar R, Lade N, Patil M, Kolte S (2019) Disease Prediction by Machine Learning Over Big. Int J Sci Res Sci Technol 29–35
6. Khatiwada B, Rajbhandari B, Mistry SK, Parsekar S, Yadav UN (2022) Prevalence of and factors associated with health literacy among people with Noncommunicable diseases (NCDs) in South Asian countries: A systematic review. Clin Epidemiol Glob Heal 18:101174. https://doi.org/10.1016/j.cegh.2022.101174
7. Kumar PM, Devi Gandhi U (2018) A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases. Comput Electr Eng 65:222–235. https://doi.org/10.1016/j.compeleceng.2017.09.001
8. Lattar H, Salem A Ben, Ghezala HH Ben (2020) Does data cleaning improve heart disease prediction? Procedia Comput Sci 176:1131–1140. https://doi.org/10.1016/j.procs.2020.09.109
9. Maji K, Gupta S (2023) Evaluation of Various Loss Functions and Optimization Techniques for MRI Brain Tumor Detection. 2023 Int Conf Distrib Comput Electr Circuits Electron 1–6.

https://doi.org/10.1109/ICDCECE57866.2023.10151232
10. Mohan S, Thirumalai C, Srivastava G (2019) Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. IEEE Access 7:81542–81554. https://doi.org/10.1109/ACCESS.2019.2923707
11. Obinikpo AA, Kantarci B (2017) Big sensed data meets deep learning for smarter health care in smart cities. J Sens Actuator Networks 6. https://doi.org/10.3390/jsan6040026
12. Piryonesi SM, El-Diraby TE (2020) Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems. J Transp Eng Part B Pavements 146:04020022. https://doi.org/10.1061/jpeodx.0000175
13. Sengupta I, Samanta S, Patra J, Gupta S (2023) Impact of Macroeconomic Indicators on the Indian Stock Market: A Study on NSE Nifty. 2023 Int Conf Comput Intell Commun Technol Networking, CICTN 2023 16:275–282. https://doi.org/10.1109/CICTN57981.2023.10140919
14. Tapager I, Olsen KR, Vrangbæk K (2022) Exploring equity in accessing diabetes management treatment: A healthcare gap analysis. Soc Sci Med 292:114550. https://doi.org/10.1016/j.socscimed.2021.114550
15. Wang Y, Zhao Y, Therneau TM, Atkinson EJ, Tafti AP, Zhang N, Amin S, Limper AH, Khosla S, Liu H (2020) Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records. J Biomed Inform 102:103364. https://doi.org/10.1016/j.jbi.2019.103364

Estimating Hemoglobin Levels Non-Invasively Through Analysis of Palm Pallor Using Deep Learning Networks

Sumana Naskar[1], Abhishek Kesarwani [1], Sunanda Das [1], Dakshina Ranjan Kisku[1] and Mamata Dalui[1]

[1]Department of Computer Science and Engineering, NIT Durgapur, Durgapur, India
{sn.21cs1107, ak.18cs1102, sd.19cs1111}@phd.nitdgp.ac.in
{drkisku, mdalui}@cse.nitdgp.ac.in

Extended Abstract

Anaemia is a medical condition characterized by deficiency in number of red blood cells or the amount of haemoglobin in the blood. Haemoglobin is a protein in red blood cells that carries oxygen to tissues throughout the body. Anaemia is a major global health concern. According to World Health Organization (WHO) data, anaemia is prevalent at notable rates, especially in countries like India. However, conventional methods for screening of anaemia are invasive and expensive, presenting challenges in resource-constrained environment. Apart from lack of infrastructural support and availability of convenient test facilities, lack of awareness and other social taboos associated with invasive test procedures create the major hindrance for performing diagnostic tests for anaemia. Hence, there is a growing demand for cost-effective, reliable, and portable non-invasive solutions. This work targets to merge cutting-age computational methodologies with the age-old practise of visual estimation of blood haemoglobin level by monitoring pallor in palm of hand, thereby realizing a non-invasive anaemia detection system. The proposed system functions by inducing pallor in palm through controlled pressure application and release, subsequently measuring the rate of colour changes and conducting time-domain analysis to establish correlation with blood haemoglobin concentration. Recorded video footage capturing the colour changes in the palm, induced by a custom-designed device, is acquired using a smartphone camera. This footage is then subjected to a series of image processing and analysis techniques for thorough examination. From the captured video of 1 minute 40 seconds, frames are extracted at the rate of 2fps. From the extracted frames, Region of Interest (ROI) is selected and then image resizing process is applied to meet the deep learning model requirement. Four key frames corresponding to video recording starting point (initial frame), colour fading point with pressure application, maximum colour fading point, normal palm colour restoration point with pressure release, from palm video photography are crucial for primary feature extraction, as they contain maximum information related to palm colour changes. Therefore, these four key frames are selected for automatic feature extraction. Those critical frames are identified through polynomial curve fitting, a process applied to the extracted correlation values, such as the Pearson correlation coefficient values. The extracted features serve as inputs to the deep neural networks. The hybrid architectures combine the strengths of Convolutional Neural Networks (CNNs) in spatial feature extraction with recurrent networks in temporal feature. Although CNNs can capture some temporal features but they are not so effective as recurrent networks. To get spatio-temporal features in data such as videos, where each frame represents a snapshot in time, recurrent networks are more effective. Therefore, we consider three different recurrent networks such as recurrent neural network (RNN), long short-term memory (LSTM), and bidirectional LSTM (Bi-LSTM) models along with CNN to predict haemoglobin level as shown in Figure 1. During experimentation, we consider cross validation with 80:20 and 70:30 train-test-splits (Table 1). The first prediction model using RNN, ensures a mean RMSE value of 0.69 g/dL, classification accuracy of 83.33%, sensitivity of 83% and specificity of 83% respectively in case of 80:20 train-test-split and second prediction model using LSTM, ensures a mean RMSE value of 0.55 g/dL, classification accuracy of 87.5%, sensitivity 91% and specificity 83% respectively in 80:20 train-test-split case. The third proposed model with Bi-LSTM prediction approach outperforms the previous two by ensuring a mean RMSE value of 0.52 g/dL, classification accuracy of 91.66%, sensitivity of 92% and specificity of 90% respectively in 80:20 train-test-split considering the clinically tested haemoglobin levels of 120 individuals. 70:30 ratio of train-test samples on the other hand ensures consistent performance for all three recurrent networks and gives a competitive performance while comparing with 80:20 ratio of train and test samples.

Keyword(s): Anaemia; Haemoglobin Estimation; RNN; LSTM; Bi-LSTM

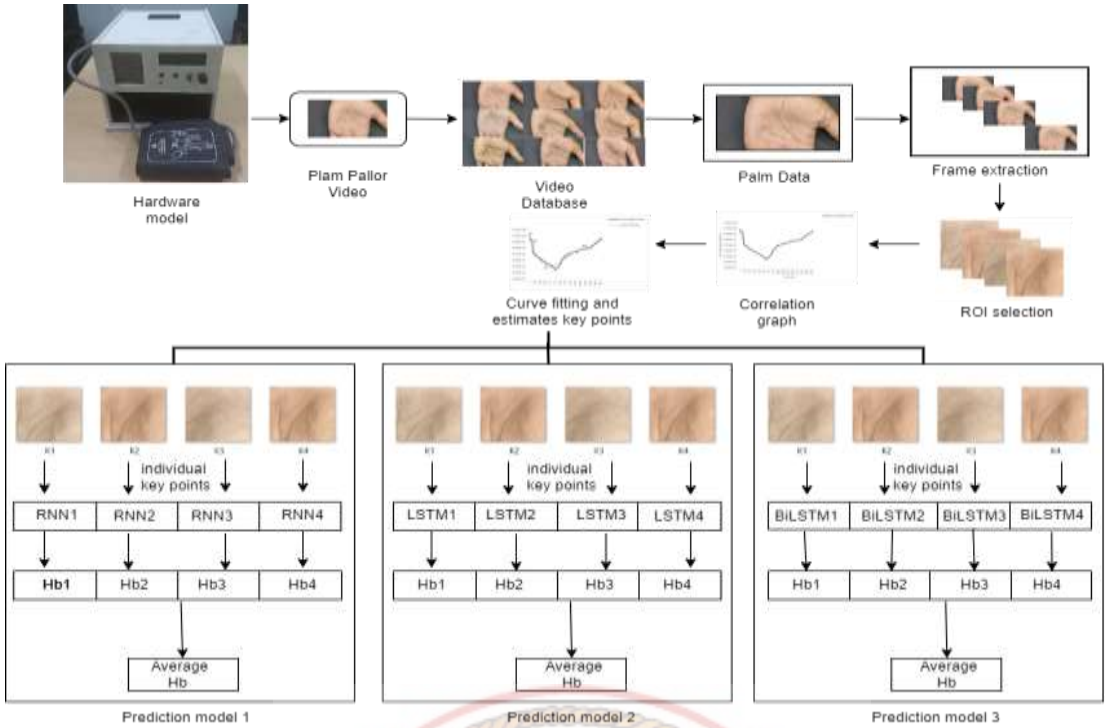


Figure 1: Architecture of the prediction models

Table 1: Comparative analysis of the proposed models with SOTA models

Methods		RMSE	MSE	Accuracy	Specificity	Sensitivity
A. Kesarwani et al. [1]		0.69	0.70	85.49%	61.29%	93%
S. Das et al. [2]		0.58	-	86.25%	75.92%	93.5%
Proposed model 1	70:30	0.62	0.50	83.33%	78%	86%
	80:20	0.69	0.48	83.33%	83%	83%
Proposed model 2	70:30	0.59	0.42	86.11%	80%	90%
	80:20	0.55	0.46	87.5%	84%	91%
Proposed model 3	**70:30**	**0.57**	**0.39**	**88.88%**	**84%**	**91%**
	80:20	**0.52**	**0.38**	**91.66%**	**90%**	**92%**

References

1. Das, S., Kesarwani, A., Kisku, D. R., Dalui, M.: Non-invasive haemoglobin estimation by observing nail color: A pca based approach. In: 2022 IEEE 6th Conference on Information and Communication Technology (CICT), pp. 1-5. IEEE (2022)
2. Kesarwani, A., Das, S., Dalui, M., Kisku, D. R., Sen, B., Roy, S., Basu, A.: Non-invasive anaemia detection by examining palm pallor: A smartphone-based approach. Biomedical Signal Processing and Control. 79, 104045 (2023)
3. Woźniak, M., Wieczorek, M., Siłka, J.: BiLSTM deep neural network model for imbalanced medical data of IoT systems. Future Generation Computer Systems. 141, 489-499 (2023)

Energy Efficient Routing for Improving Lifetime in MWSN: A Clustering Approach

Ranadeep Dey[1] and Parag Kumar Guha Thakurta[1]

[1]Department of Computer Science & Engineering
National Institute of Technology, Durgapur, India, PIN - 713209
{1rd.18cs1505@phd.nitdgp.ac.in, 2pk.guhathakurta.cse@nitdgp.ac.in}

Extended Abstract

A mobile wireless sensor networks (MWSNs) consists of a number of mobile sensor nodes, which can able to move within a network. It is more flexible than a static WSN [1], as the deployed nodes can survive against the rapid topological transformations due to mobility. These nodes having limited battery power, are used to collect specific data like temperature, humidity, pressure, light etc., from various environments [2,3]. The collected raw data are sent to a static sink node of the network [4], where the data is stored for future use. Hence, the design of MWSNs requires efficient connectivity among the sensor nodes and sink node. Under these circumstances of change in network topology, the data routing plays a significant role [5] as the network needs to be partitioned in such a way, so that it can avoid frequent disconnections and subsequently, can avail more reliable data transmissions. A hierarchical cluster based network [2] is mostly suitable one for data transmission in MWSNs over alternative possibilities, as it can adapt efficiently without affecting other associated factors. It is noteworthy to mention here that an efficient routing in MWSNs is possible to accomplish in reduced time [4]. The energy dissipation by the sensor nodes is involved with the data transmission related to routing. As the energy level of a node is limited, so it motivates to propose an energy efficient routing in the hierarchical MWSNs which can improve the lifetime of the network.

An energy efficient routing protocol for MWSNs is proposed in this paper to improve network lifetime by reducing energy consumption of the nodes in the network. In order to obey this objective aimed here, a random number of sensor nodes are initially considered with their current locations in the network. Hence, K-means clustering algorithm [6] is used to find a predefined number of clusters with their initial cluster heads (CHcs) and the centroid location of these clusters is also determined. The role of these CHcs is to elect our "Distance to sink and cluster centroid with Battery Level Aware Cluster Head" (DDBLACH) nodes from each cluster, by sending and receiving intra-cluster messages among other member sensor nodes. The proposed DDBLACH in a cluster is determined by the value of a cost function which associates three factors, namely the distances between all the intra-cluster sensor nodes with its centroid, the distance between mobile sensor node and the sink node and the battery level of that particular node. The node having maximum value of this cost function is considered as elected DDBLACH, which is responsible for data transmission from intra-cluster sensor nodes to sink node of MWSN using a hierarchical routing of tree arrangement [7]. Meanwhile, the topological changes occur in the network due to node mobility and therefore, it reconstructs the clusters by K-means algorithm and follows as per workflow shown in Figure 1.

Under such a scenario, an energy efficient routing is discussed in an iterative approach, so that network lifetime is increased. The simulation results indicate the superiority of our proposed routing scheme over some other existing methods [8, 9, 10] in terms of improved number of alive nodes and the average network lifetime.

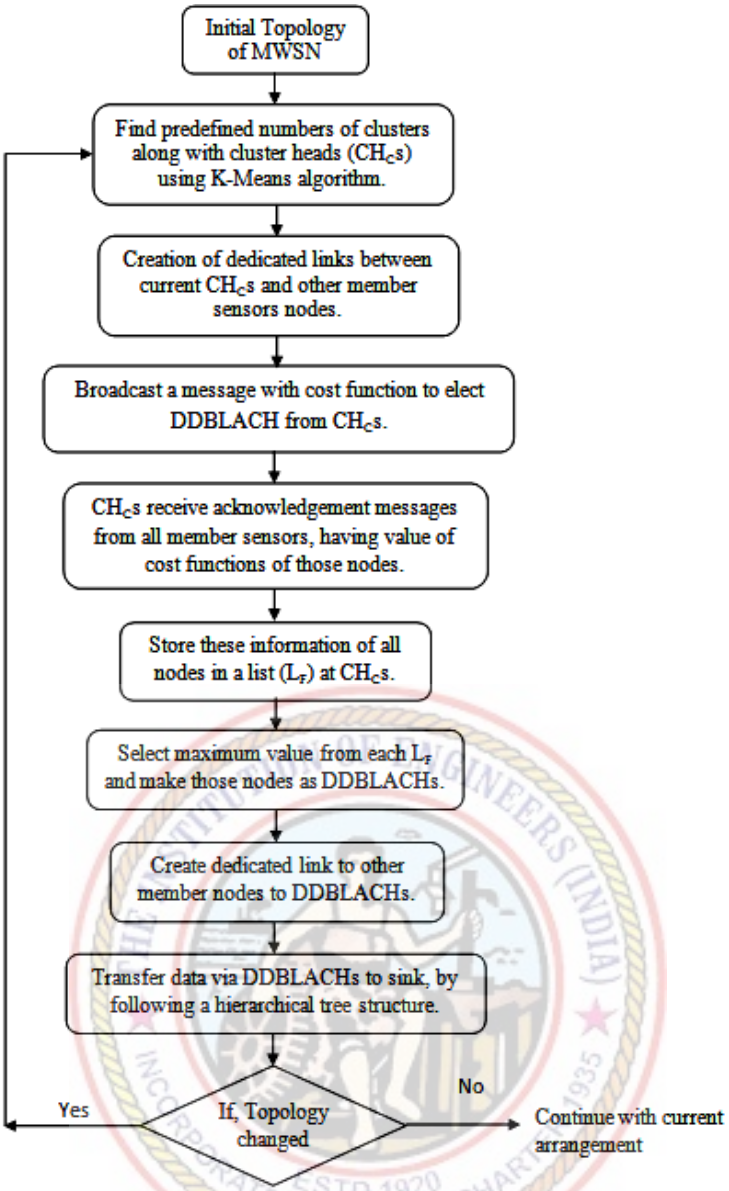


Figure 1: Workflow diagram of our proposed work

Keyword(s):MWSN, clustering, cluster head, routing, energy, lifetime.

References

1. Luo, R. C., Chen, O.: Mobile Sensor Node Deployment and Asynchronous Power Management forWireless Sensor Networks. IEEE Transactions on Industrial Electronics,vol. 59, 5, pp. 2377- 2385 (2012)
2. Ramasamy, V.: Mobile Wireless Sensor Networks: An Overview. Wireless Sensor Networks-InsightsandInnovations,InTech,(2017)
3. Mohamed, S. M., Hamza H. S., Saroit, I. A.: Coverage in mobile wireless sensor networks (M-WSN): A survey. Computer Communications, vol. 110, pp. 133-150, (2017)
4. .Liu, X.: A typical Hierarchical Routing Protocols for Wireless Sensor Networks: A Review. IEEESensorsJournal,vol. 15,10,pp. 5372-5383, (2015)
5. Sabor, N., Sasaki, S., Abo-Zahhad, M., Ahmed, S. M. : A Comprehensive Survey on Hierarchical-Based Routing Protocols for Mobile
6. Wireless Sensor Networks: Review, Taxonomy, and Future Directions. Wireless Communications and Mobile Computing, Hindawi, Article ID 2818542, (2017)

7. Sasikumar, P., Khara, S.:K-Means Clustering in Wireless Sensor Networks. IEEE 2012 Fourth International Conference on Computational Intelligence and Communication Networks, Mathura, India, pp. 140-144, (2012)
8. Rady, A., El-Rabaie,E.L.S.M.,Shokair, M., Abdel-Salam, N.: Comprehensive survey of routing protocols for Mobile Wireless Sensor Networks. International Journal of Communication Systems 34, 15, (2021)
9. Nguyen, L. T., Defago, X., Beuran, R., Shinoda, Y.:An energy efficient routing scheme for mobilewireless sensor networks. IEEE international Symposium on Wireless Communication Systems,Reykjavik,Iceland,pp.568-572, (2008)
10. Anitha, R. U., Kamalakkannan, P.: Enhanced cluster based routing protocol for mobile nodes inwireless sensor network. International Conference on Pattern Recognition, Informatics and MobileEngineering,Salem,India,pp.187-193, (2013)
11. 10. Jambli, M. N., Zen, K., Lenando, H. , Tully, A.: Performance evaluation of AODV routing protocol for mobile wireless sensor network. 7th International Conference on Information Technology in Asia, Sarawak, Malaysia, pp. 1-6, (2011)

Advances of Machine Learning for diagnosis, monitoring and treatment of different Neuro-degenerative Diseases: A comprehensive review

Sneha De[1], K. Srikanth[2], Bimal Das[3], Deepshikha Datta[4]

[1]Assistant Professor, Department of Food & Nutrition, Brainware University, Kolkata
de.sneha1995@gmail.com

[2]Assistant Professor, Department of Hospital Management, Brainware University, Kolkata
dks.hm@brainwareuniversity.ac.in

[3]Department of Chemical Engineering, National Institute of Technology Durgapur, West Bengal, India
bimal.das@che.nitdgp.ac.in

[4]Department of Chemistry, Brainware University, Barasat, Kolkata, West Bengal, India
deepa.datta81@gmail.com

Extended Abstract

Artificial Intelligence (AI) refers to the simulation of human intelligence processes by machines, typically computer systems. These procedures involve gathering data, creating rules for utilizing that data and reasoning which is applying rules to arrive at a conclusion and self-correction. Artificial Intelligence (AI) methods include a variety of approaches, including Machine Learning (use algorithms to analyse data, draw conclusions, and gain knowledge from it), computer vision (allowing machines to interpret and understand the visual world), natural language processing (allowing machines to understand and generate human language), and robotics (building robots that can perform tasks autonomously). These days, Machine Learning (ML) techniques are highly helpful in both diagnosing and monitoring disorders such as neuro-degenerative diseases. A class of illnesses known as neuro-degenerative diseases are defined by a steady deterioration of the nervous system's structure and functionality. Numerous symptoms, such as mobility problems, cognitive impairment, and behavioural and emotional disturbances, are brought on by this degeneration. Amyotrophic Lateral Sclerosis (ALS), Parkinson's disease, Alzheimer's disease and Huntington's disease are a few examples of neuro-degenerative disorders. This study aims to investigate the significant influence of artificial intelligence and its diverse approaches on transforming the diagnosis and monitoring of neuro-degenerative illnesses. The various bioinformatics and AI approaches to finding potential biomarkers, early diagnosis for neuro-degenerative diseases like Parkinson's and Alzheimer's and the application of ML and Deep Learning (DL) methods to better understand disease progression, treatment planning and patient outcome prediction are all covered in this review paper.

Keyword(s): Artificial Intelligence, Machine Learning, Neuro-degenerative diseases, Bioinformatics, Monitoring

References

1. Choi, B. K., Madusanka, N., Choi, H. K., So, J. H., Kim, C. H., Park, H. G., Bhattacharjee, S., & Prakash, D. (2020). Convolutional neural network-based MRI image analysis for Alzheimer's disease classification. Current Medical Imaging, 16(1), 27-35.
2. Solana-Lavalle, G. (2021). Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation. Biomedical Signal Processing and Control. Elsevier.
3. Mei, J., Desrosiers, C., & Frasnelli, J. (2021). Machine Learning for the Diagnosis of Parkinson's Disease. Frontiers in Aging Neuroscience.
4. Taut, A.-M., Ionescu, B., & Santarnecchi, E. (2021). Artificial intelligence in neurodegenerative diseases: A review of available tools with a focus on machine learning techniques. Artificial Intelligence in Medicine. Advance online publication.

Utilizing Machine Learning for Anticipating Complications in Pregnancy and Child Birth

Suvojit Deb[1], Amit Sinha[2], and Bhaskar Roy[2]

[1]Department of Computer Science & Engineering, Asansol Engineering College, Asansol, India
debsuvojit12@gmail.com
[2]Department of Artificial Intelligence & Machine Learning, Asansol Engineering College, Asansol, India
sinha.amit84@gmail.com,bhaskar.ec11@gmail.com

Extended Abstract

In order to forecast different diseases, artificial intelligence-including machine learning and deep learning is crucial to the medical sector. Machine learning (ML) has been applied widely to identify potential maternal risks during pregnancy or after childbirth disorders, as well as to forecast the mode of labour. Every year, a large number of newborns in our nation and other nations are disabled and pass away from diseases passed down from their parents. The main goals are to identify potential issues during childbirth, such as congenital disorders, congenital heart diseases, genetic abnormalities, clubfoot, etc., and to forecast the best way to give birth. Pregnancy-related illnesses may be identified early and treated to improve the health of both the mother and the unborn child. When risk factors are identified early in pregnancy and managed, prevented, mitigated, and adhered to, adverse perinatal outcomes and challenges for mother and child can be significantly reduced. Given the workload that medical professionals currently have, clinical decision support systems, or CDSSs, can aid in risk management. However, these systems do require high-quality, clinically interpretable decision support models that are based on validated medical data. In order to develop models for predicting the risks of childbirth and the dates of delivery, we retrospectively examined electronic health records from the perinatal centre of the Almazov Specialized Medical centre in Saint-Petersburg, Russia. Congenital heart disease (FHD) is a frequent and serious congenital defect in children. In Asia, the frequency of FHD birth abnormalities has risen to 9.3%. In order to identify birth defects and death early on, echocardiography remains the most effective method of screening for prenatal heart abnormalities. Moreover, the pathophysiological alterations that transpire across different phases of pregnancy demand a high degree of proficiency to recognize and evaluate the progression of an illness. Thus, research on automated FHD screening is required. Machine learning (ML) technology can only identify blood cancer and congenital heart defects in utero, as well as chromosomal abnormalities. In the end, the findings of this review study influenced the development of a conceptual framework that would improve the machine learning-based maternal healthcare system. All things considered, this research will provide a state-of-the-art framework for machine learning (ML) in the context of maternal healthcare, supporting clinical decision-making, medical diagnosis and treatment, and the forecasting of pregnancy problems and delivery alternatives. Furthermore, machine learning (ML) can analyse DNA before a child is born and eliminate genetic diseases (such thalassemia) in the embryo. In the end, the findings of this review study influenced the development of a conceptual framework that would improve the machine learning-based maternal healthcare system. All things considered, this research will provide a state-of-the-art framework for machine learning (ML) in the context of maternal healthcare, supporting clinical decision-making, medical diagnosis and treatment, and the forecasting of pregnancy problems and delivery alternatives. Furthermore, machine learning (ML) can analyse DNA before a child is born and eliminate genetic diseases (such thalassemia) in the embryo.

Keyword(s): Machine learning (ML); Fetal, Child; Newborn; DNA; Diseases.

References

1. Fatima, M., Pasha, M.: Survey of Machine Learning Algorithms for Disease Diagnostic. J Intell Learn Syst Appl, 9, 1-16 (2017). (https://doi.org/10.4236/jilsa.2017.91001)
2. Workowski, K.A., Berman, S.: Sexually transmitted diseases treatment guidelines. Morbidity and Mortality Weekly Report, 59, pp. 107–113 (2010)
3. WHO fact sheet on congenital anomalies, updated September 2016, Available at www.who.int/mediacentre/factsheets/fs370/en/

Stock Market Analysis Using Supervised Machine Learning

Arkaprava Mukherjee[1], Amit Sinha[2], and Bhaskar Roy[2]

[1]Department of Computer Science & Engineering, Asansol Engineering College, Asansol, India
arkapravam148@gmail.com
[2]Department of Artificial Intelligence & Machine Learning, Asansol Engineering College, Asansol, India
sinha.amit84@gmail.com,bhaskar.ec11@gmail.com

Extended Abstract

Exploring the realm of stock market prediction poses various challenges and has garnered recent attention. This study employs machine learning algorithms and historical data from financial indices to forecast future stock prices. Specifically, It evaluates the prediction power of Long Short-Term Memory (LSTM) and Regression models using important characteristics such as volume, Initiate, Conclude, Minimal, and Maximum. The stock market functions as a dynamic environment with diverse stakeholders, including financial institutions, brokerage firms, and individual investors, all seeking to optimize returns and manage risks. Despite inherent complexities, this investigation aims to enhance stock price prediction through the adept application of machine learning techniques. Ultimately, the goal is to democratize access to advanced forecasting methods for broader utilization. Understanding and predicting stock market movements present significant challenges, fuelled by the complexity and volatility inherent in financial markets. Recently, there has been a surge of interest in leveraging machine learning methodologies to enhance forecasting accuracy. This study contributes to this burgeoning field by examining the effectiveness of Long Short-Term Memory (LSTM) and Regression models in predicting future stock prices. The stock market operates within a dynamic ecosystem, influenced by various factors and stakeholders. Financial institutions, brokerage firms, and individual investors all play vital roles in shaping market dynamics, aiming to maximize returns while managing risks. Despite the challenges posed by this dynamic environment, the study endeavors to enhance stock price prediction through sophisticated machine learning techniques. To conduct this investigation, historical data from financial indices are utilized, incorporating key metrics such as volume, opening, closing, minimum, and maximum values. By analyzing these factors, the study aims to ascertain the predictive capabilities of the selected machine-learning algorithms. LSTM, known for its ability to capture long-term dependencies, is particularly promising in capturing the intricate patterns inherent in stock market data. The stock market operates within a dynamic ecosystem, influenced by various factors and stakeholders. Financial institutions, brokerage firms, and individual investors all play vital roles in shaping market dynamics, aiming to maximize returns while managing risks. Despite the challenges posed by this dynamic environment, the study endeavours to enhance stock price prediction through sophisticated machine-learning techniques. The ultimate goal of this research is to democratize access to advanced forecasting methods, making them accessible to a broader audience. By leveraging machine learning algorithms and historical data, the study seeks to provide actionable insights for investors and analysts alike. Furthermore, by shedding light on the predictive capabilities of LSTM and Regression models, this research contributes to the ongoing discourse surrounding the application of machine learning in finance. The ultimate goal of this research is to democratize access to advanced forecasting methods, making them accessible to a broader audience. By leveraging machine learning algorithms and historical data, the study seeks to provide actionable insights for investors and analysts alike. Furthermore, by shedding light on the predictive capabilities of LSTM and Regression models, this research contributes to the ongoing discourse surrounding the application of machine learning in finance. In summary, this study represents a significant step towards improving stock market prediction through the integration of machine learning methodologies. By addressing key challenges and leveraging historical data, the research aims to provide valuable insights for stakeholders navigating the complexities of financial markets.

Keyword(s): Closing price; High value; Low value; LSTM architecture; Opening price; Regression analysis

References

1. McCallum A., Nigam K., Rennie J. and Seymore K.: A Machine learning approach to Building domain-specific Search engine,
2. IJCAI'99: Proceedings of the 16th international joint conference on Artificial intelligence. Volume 2, pp- 662–667 (1999)

3. Yadav S.: STOCK MARKET VOLATILITY - ASTUDY OF INDIAN STOCK MARKET. Global Journal for Research Analysis, vol-6, no. 4 (2017)
4. Montgomery D.C., Peck E.A. and Vining G.G.: Introduction to linear regression analysis. John Wiley & Sons, Vol. 821, (2012)

A Review on the Aircraft Detection System using Machine Learning and Radar System

Snehangshu Pandey[1], Amit Sinha[2], and Bhaskar Roy[2]

[1]Department of Computer Science & Engineering, Asansol Engineering College, Asansol, India
snehangshupandey141@yahoo.com
[2]Department of Artificial Intelligence & Machine Learning, Asansol Engineering College, Asansol, India
sinha.amit84@gmail.com,bhaskar.ec11@gmail.com

Extended Abstract

The study delves into the application of supervised machine learning algorithms for the crucial task of distinguishing aircraft as allies or adversaries, with a specific focus on its implications for various sectors including airspace security, public safety, resource management, and military operations. By harnessing machine learning techniques, particularly through the analysis of flight test data, the study aims to gain insights into flight movement patterns. The developed model exhibits the capability to discern information from six different aircraft, thereby addressing the complexity inherent in aircraft identification. Key strategies employed in this research involve the development of a manoeuvre extraction algorithm to dissect and understand flight dynamics. This algorithm plays a pivotal role in analysing the intricate movements exhibited by aircraft during flight, shedding light on their behaviour and enabling more accurate classification. Additionally, the study utilizes statistical measures to represent time-series data effectively, facilitating a deeper analysis of flight dynamics and movement patterns over time. Furthermore, the study tackles the challenge of an imbalanced dataset, a common issue in real-world scenarios where certain classes may be underrepresented. By implementing a modified oversampling method, the researchers ensure more robust model training by synthetically increasing the number of instances for underrepresented classes, thereby mitigating potential biases and improving overall model performance. A variety of machine learning algorithms are explored within the study's framework, ranging from traditional techniques like Logistic Regression to more advanced approaches such as Support Vector Machines (SVMs) with both linear and nonlinear kernels, as well as Artificial Neural Networks (ANNs). The selection of hyperparameters for each algorithm is a crucial step, performed through rigorous model estimation and optimization processes guided by performance metrics. This meticulous approach ensures that the models are tuned to their optimal configuration, enhancing their predictive accuracy and generalization capabilities. Ultimately, this study showcases the potential of machine learning in enhancing aircraft identification processes, with implications extending to critical domains such as airspace security and military operations. Through meticulous algorithmic development and model optimization, the research contributes towards improving the efficiency and accuracy of aircraft classification systems, thereby bolstering safety and security measures in various contexts. During the course of the study, in-flight tests revealed challenges with linear models in handling complex classes and manoeuvre extraction. Non-linear classifiers, on the other hand, emerged as more suitable alternatives, demonstrating robust performance in handling the intricacies of aircraft manoeuvres and movement patterns. This finding underscores the importance of exploring and leveraging advanced machine learning techniques to tackle complex real-world problems effectively. In an intriguing deviation from the primary focus on aircraft identification, the study also highlights the significance of understanding Unidentified Falling Objects (UFOs) in the context of Large Hadron Collider (LHC) operations. These objects, if left undetected, can potentially impact machine availability and safety within the LHC complex. To address this challenge, the researchers employ innovative approaches such as wire scanners and Diamond Beam Loss Monitors (BLMs). These technologies offer promising avenues for analyzing calibrated UFO dynamics and validating quick detectors for movement plane determination, ultimately enhancing the overall safety and operational efficiency of the LHC. However, the study also encountered obstacles in the form of connectivity issues with beam 2 dB LMs and low signal-to-noise ratios in low-energy measurements. These challenges underscore the complexity of UFO research and the need for robust detection and protection measures in high-energy physics experiments like the LHC. Overcoming such hurdles will be crucial in ensuring the reliable identification and mitigation of potential threats posed by UFOs, safeguarding the integrity of sensitive scientific equipment and experiments.

Keyword(s): Machine learning; Artificial Intelligence; Radar System; UFO.

References

1. Torralba, A., Oliva, A., Castelhano, M. S., & Henderson J. M.: Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. Psychological review, 113(4), 766-786, (2006)
2. Shi, C., Wang, F., Tian, J., & Xu, F.: Aircraft recognition in remote sensing images based on transfer learning and ensemble learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 5358-5371 (2021)
3. Samal, A., and Iyengar P. A.: Automatic recognition and analysis of human faces and facial expressions: A survey. Pattern recognition. 25(1), 65-77, (1992)

Quantifying the Influence of Random Variables on Structural Failure Probabilities: Global Sensitivity Analysis Using Kriging-Based Active Learning Approach

Pijus Rajak[1], and Pronab Roy[2]
[1]Department of Civil Engineering, National Institute of Technology Durgapur, Durgapur-713209, India
pr.19ce1105@phd.nitdgp.ac.in
[2]Department of Civil Engineering, National Institute of Technology Durgapur, Durgapur-713209, India
proy.ce@nitdgp.ac.in

Extended Abstract

Quantifying the contribution of random variables to the output variance is essential for enhancing model precision and understanding the impact of uncertainty on structural failure probabilities. Sobol' indices play a crucial role in quantifying the effects of these variables on model outcomes through global sensitivity analysis. Kriging surrogate models have been widely used in reliability analysis and sensitivity studies due to their accuracy and capability to handle complex limit state functions. However, standard Kriging approximations may result in significant uncertainties, especially when prior information about the problem is limited. Active learning techniques can refine surrogate models through iterative sampling to balance exploration and exploitation. This enhances prediction accuracy while controlling computational costs. In this research, presented a computational framework that incorporates an active learning-based Kriging model approach for facilitating a variance-based global sensitivity assessment using a manageable number of model evaluations.

The Kriging model construction involves two main steps. In the first step, a small set of training samples is selected from the initial Monte Carlo samples (MCS) using K-means clustering. The number of clusters is determined using the elbow method. The variables are normalized before calculating distances to minimize the influence of variables with large range. Then, two samples from each cluster are selected based on certain criteria to represent the respective cluster. Additionally, the mean point of the random variables is included as the centre point. A Kriging model is then built using the first training set. In the second step, samples in the concerned domain are identified from the MCS using the Kriging model. K-means clustering is applied again to divide these samples into clusters. Two samples from each cluster are selected to represent the concerned domain, comprising a second training set. To refine the approximation of the limit state function (LSF), five additional samples are sequentially selected using an active learning algorithm. The surrogate model is optimized by enriching the concerned domain using samples near the approximated LSF and space-filling design in both the safe and failure domains. Initially, two samples are selected sequentially near the updated LSF to minimize misclassification from samples close to the LSF. Then a well-known space-filling design, maximin criterion is used to reduce the risk of misclassification caused by large variance by selecting two samples sequentially. The midpoint of two training samples has been added where the sample is at maximum distance from the previous training samples. This improves the Kriging model and achieves the desired accuracy while avoiding misclassification. The failure probability is estimated at each iteration. A stopping criterion ends the iterative enrichment process when the approximation of LSF reaches a sufficient accuracy level.

This study presents a framework for global sensitivity analysis considering first-order, and total-order Sobol' indices. Sobol' indices are used to quantify the contributions of individual input parameters and their interactions to the output variability of a system. In this approach, the output of a system is approximated as the weighted sum of its input parameters, where the weights are based on their importance. Saltelli sampler is used to compute Sobol' indices, which quantify the contribution of individual and combined input variables to the overall output parameter. Fig. 1 shows the flowchart summarizing the overall approach to estimate first and total-order Sobol' indices. The applicability of the proposed framework is demonstrated through two examples, the Ishigami function and a 23-bar truss structure.

Keyword(s): Global sensitivity analysis; Sobol' indices; Kriging-based active learning; Monte Carlo method; Failure probability.

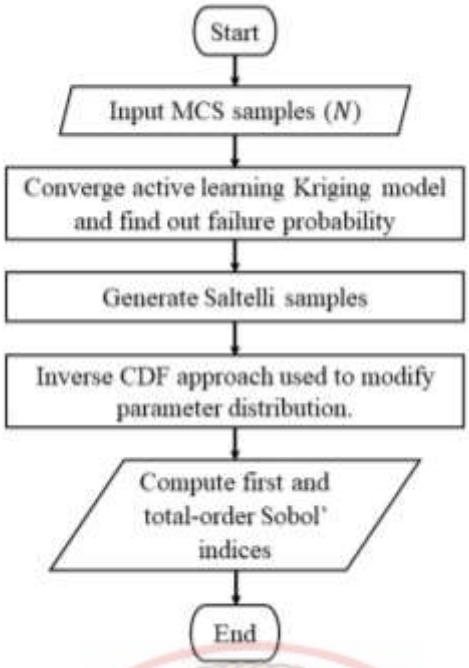


Figure 1: Flowchart for estimating first and total-order Sobol' indices

References

1. Papaioannou, I., Straub, D.: Variance-Based Reliability Sensitivity Analysis and the FORM α-Factors. Reliability Engineering & System Safety, 210, 107496 (2021)
2. Sudret, B.: Global Sensitivity Analysis Using Polynomial Chaos Expansions. Reliability Engineering & System Safety, 93(7), 964-979 (2008)
3. Guo, H., Dong, Y., Bastidas-Arteaga, E., Gu, X. L.: Probabilistic Failure Analysis, Performance Assessment, and Sensitivity Analysis of Corroded Reinforced Concrete Structures. Engineering Failure Analysis, 124, 105328 (2021)

Machine Learning Approaches for Predicting Soil Liquefaction Potential in Non-Plastic Soil

Titas Chakraborty[1], Sushovan Dutta[2], Sanku Konai[2], Gilbert Hinge[2]

[1]Department of Civil Engineering, National Institute of Technology Durgapur, Durgapur – 713209, West Bengal, India
tc.20u10845@btech.nitdgp.ac.in

[2]Department of Civil Engineering, National Institute of Technology Durgapur, Durgapur – 713209, West Bengal, India
{sdutta.ce,skonai.ce,ghinge.ce}@nitdgp.ac.in

Extended Abstract

Soil liquefaction poses a significant threat to infrastructure integrity, especially in regions prone to seismic activity. Traditional methods for assessing liquefaction potential are often laborious and time-consuming, prompting the need for alternative, more efficient prediction approaches (Whitman, 1971; Young-Su & Byung-Tak, 2006). This study introduced a methodology utilizing various machine learning models, including Artificial Neural Networks, Support Vector Regression, and Random Forest Regression, to accurately predict liquefaction potential, reducing time, effort, and associated costs. The models were trained using datasets obtained from Sezer et al. (2019) on non-plastic soil deposits, which included multiple cyclic Triaxial tests across varying Density Index ranges. Seven parameters, including confining pressure, axial load, cyclic deviatoric stress ratio, axial strain, number of cycles, effective mean principal stress, and relative density, were used as inputs, with excess pore water pressure ratio as the output. Data cleaning procedures were carried out to eliminate redundancies and erroneous values. Model performance was assessed using standard metrics such as R^2 score and Mean Squared Error (MSE). Results indicated the superior performance of Random Forest Regression, achieving an R^2 score of 0.75 and an MSE value of 0.032, indicating its effectiveness in liquefaction potential assessment. Additionally, the integration of Genetic Programming revealed significant correlations between soil parameters, loading conditions, and liquefaction susceptibility (Fig. 1). These findings highlighted the potential of machine learning models, particularly Random Forest Regression, in accurately predicting soil liquefaction potential, emphasizing the importance of using such advanced techniques in enhancing infrastructure resilience. However, to improve the accuracy of the model, other parameters like coefficient of uniformity, grain size distribution of soil can be used in future studies.

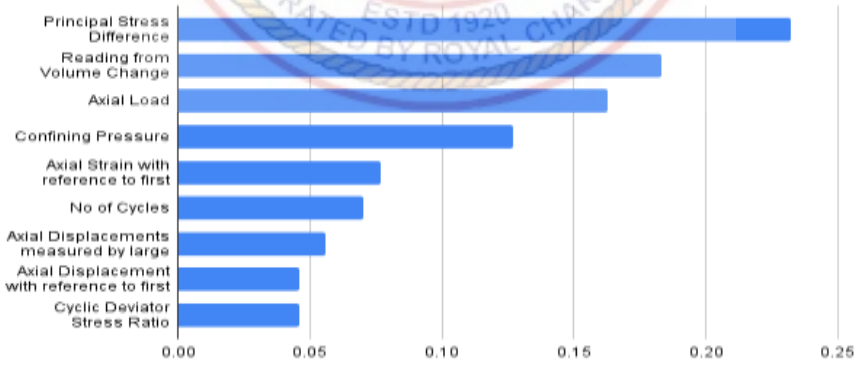


Figure 1: Relevance of the input parameters

Keyword(s): Soil Liquefaction; Machine Learning; Random Forest Regression; Artificial Neural Networks; Support Vector Regression; Cyclic Triaxial Tests.

References

1. Sezer, A., Tanrınıan, N., Karakan, E.: Data for: Cyclic undrained behavior and post liquefaction settlement of a nonplastic silt, Mendeley Data, V1, doi: 10.17632/9t8bpsrznm.1 (2019)
2. Young-Su, K., & Byung-Tak, K.: Use of artificial neural networks in the prediction of liquefaction resistance of sands. Journal of Geotechnical and Geoenvironmental Engineering, 132(11), pp. 1502-1504 (2006)
3. Whitman, R. V.: Resistance of soil to liquefaction and settlement. Soils and Foundations, 11(4), pp. 59-68 (1971)

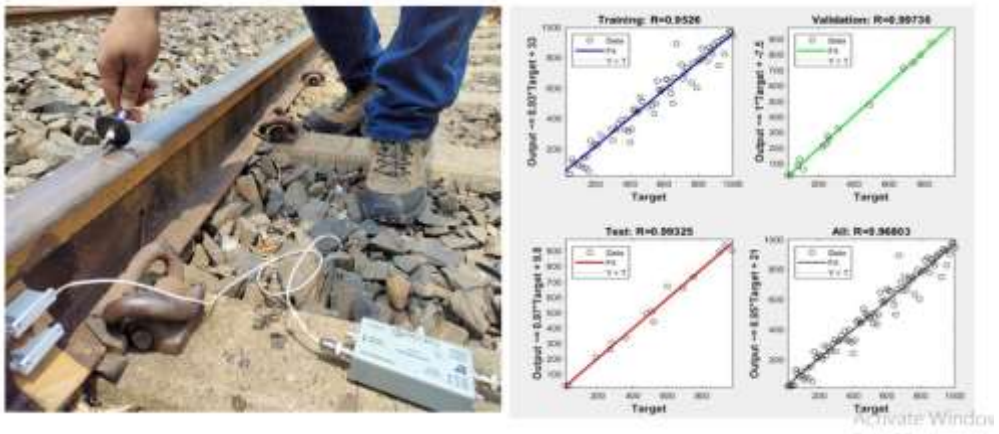
Application of AI in Health Monitoring of Railway Track

Apurba Pal[1] and Aloke Kumar Datta[2]

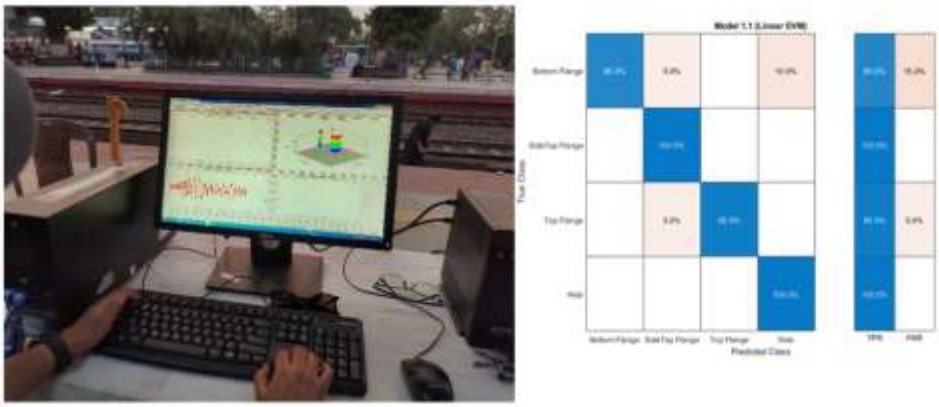
[1]Research Scholar, Civil Engineering Department, National Institute of Technology, Durgapur, India
apurbapal1992@gmail.com
[2]Associate Professor, Civil Engineering Department, National Institute of Technology, Durgapur, India
akdatta.ce@nitdgp.ac.in

Extended Abstract

Railway tracks are integral components of transportation infrastructure, facilitating the movement of goods and passengers across vast distances. However, ensuring the structural integrity and safety of railway tracks presents a formidable challenge for railway operators worldwide. Traditional methods of health monitoring of railway tracks, typically relying on manual inspections conducted at periodic intervals, are often labour-intensive, time-consuming, and prone to human error. The emergence of Artificial Intelligence (AI) technologies with real-time sensor data offers a promising solution to revolutionize health monitoring practices in railway tracks by enabling smart, real-time, automated detection and diagnosis of faults. Existing literature in this context indicate that there is a growing trend of online real-time fault detection using Acoustic Emission (AE) technique allied with AI models. This research endeavours to develop an efficient, smart fault detection approach using laboratory and field-tested AE sensor data. Acoustic Emission (AE) sensors are strategically deployed along the track to capture and analyse signals emitted by various sources of defects. These AE sensors record real-time data which are considered for AI model and help in real-time fault detection of rail [1]. The AE parameters such as amplitude, frequency, RMS, counts etc., serve as the input to the AI models trained to classify and localize faults. Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are among the machine learning algorithms employed in this study, owing to their proven efficacy in fault localization and classification tasks [2-3]. Real-world experiments are conducted on operational railway tracks to validate the performance and reliability of the developed AI model with laboratory data. The experimental setup is depicted in figures 1(a-b), while figure 1(c) displays the regression curve generated by the developed Artificial Neural Network (ANN) model, exhibiting the R value of 0.96803. Additionally, figure 1(d) illustrates the Support Vector Machine (SVM) classification results. The results demonstrate significant improvements in fault detection accuracy and efficiency compared to traditional manual inspection methods. Moreover, the system's ability to prioritize and schedule maintenance activities based on the severity and location of detected faults contributes to optimizing maintenance operations and reducing overall maintenance costs. In conclusion, the application of AI in health monitoring for railway tracks represents a transformative paradigm shift towards proactive and predictive maintenance strategies. By harnessing the power of AI technologies, railway operators can ensure the continued safety and reliability of their infrastructure, thereby facilitating uninterrupted railway operations and enhancing overall transportation efficiency.



(a) **(b)**



(b) **(d)**

Figure 1: (a) Data collection in Field, (b) Acoustic Emission (AE) System i.e., Micro-II Express, (c) ANN Regression curve, (d) Confusion Matrixes for test validation

Keyword(s): Rail Section, Artificial Intelligence, Machine Learning, Artificial Neural Network, Support Vector Machine, Acoustic Emission

References

7. Kostryzhev, A. G., Davis, C. L., and Roberts, C.: Detection of crack growth in rail steel using acoustic emission. Ironmaking & Steelmaking. 40(2), 98-102 (2013).
8. Zhang, X., Hao, Q., Wang, K., Wang, Y., Shen, Y., & Hu, H.: An investigation on acoustic emission detection of rail crack in actual application by chaos theory with improved feature detection method. Journal of Sound and Vibration, 436, 165-182 (2018).
9. Samanta, B. I. S. W. A. J. I. T., Al-Balushi, K. R., & Al-Araimi, S. A.: Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. Engineering applications of artificial intelligence. 16(7-8), 657-665 (2003).

Enhancing Landslide Susceptibility Mapping in Sikkim and Darjeeling, India using Machine Learning Models

Amandeep Kumar[1], Luwang Clinton Nongmaithem[2], Gilbert Hinge[3]

[1]Department of Civil Engineering, National Institute of Technology Durgapur, Durgapur – 713209, West Bengal, India
ak.20u10037@btech.nitdgp.ac.in

[2]Department of Civil Engineering, National Institute of Technology Durgapur, Durgapur – 713209, West Bengal, India
clinton67.cn@gmail.com

[3]Department of Civil Engineering, National Institute of Technology Durgapur, Durgapur – 713209, West Bengal, India
ghinge.ce@nitdgp.ac.in

Extended Abstract

Landslides pose an enormous threat to communities and infrastructure in mountainous regions, predominantly in areas defined by steep slopes, fragile geological formations, and heavy rainfall (Poonam & Sharma, 2022). Sikkim and Darjeeling in India exemplify such regions, where landslides are not only frequent but also have devastating consequences (Rajinder et al., 2002). Accurate Landslide Susceptibility Mapping (LSM) is therefore vital for identifying vulnerable areas and implementing effective mitigation measures. Traditional LSM approaches often rely on empirical models or expert knowledge, which may lack the precision and scalability needed to address the complexities of landslide dynamics in these regions (Bui et al., 2013). In response, this study explores the application of machine learning techniques to enhance LSM in Sikkim and Darjeeling, India. The study leverages historical landslide inventories sourced from Indian government entities and the NASA Landslide catalog, alongside various conditional variables. These datasets serve as inputs for developing and optimizing multiple machine learning models. Bayesian Optimization is employed to fine-tune model parameters and enhance performance. The models are rigorously evaluated using standard metrics to assess their accuracy and reliability in predicting landslide susceptibility. Among the machine learning models tested, the ANN model was found to be the best model, exhibiting superior accuracy scores of 0.94 during training and 0.90 during testing phases. The ANN's ability to capture complex nonlinear relationships between input variables and landslide occurrence makes it well-suited for LSM in these heterogeneous landscapes. The results of this study have imperative inferences for landslide risk management and disaster preparedness in Sikkim and Darjeeling, India. Accurate LSM enables authorities to identify high-risk areas, prioritize resource allocation, and implement targeted mitigation strategies. By incorporating machine learning techniques, the proposed methodology offers a scalable and adaptable approach to landslide susceptibility mapping, capable of accommodating diverse environmental and anthropogenic factors. Furthermore, the use of publicly available datasets and open-source machine learning libraries enhances the accessibility and reproducibility of the methodology, facilitating its adoption by researchers and practitioners worldwide. Thus, the results of this study exhibit the potential of ML techniques, notably ANN, to advance LSM in landslide-prone regions like Sikkim and Darjeeling, India. The methodology used in this study shows promise for accurate landslide susceptibility mapping, offering valuable insights that can be utilized for enhancing preparedness and resilience in this study area

Keywords: Landslide susceptibility mapping; Sikkim; Darjeeling; Machine learning, Neural Network.

References

1. Poonam,K., & Sharma,N. 2022.: Hybrid Machine Learning Approach for Landslide Prediction, Uttarakhand, India." Scientific Reports 12(1): 1–23. doi:10.1038/s41598-022-22814-9. (2022)
2. Rajinder,B., Grimstad,E., Otto,J., Dhawan,A., Singh,R., Verma,S., and Venkatachalam,K.: Landslide Hazards and Mitigation Measures at Gangtok , Sikkim Himalaya." 64: 351–68 (2022)
3. Bui,T., Pradhan,B., Lofman,O., Revhaug,I., and Dick.O.: Regional Prediction of Landslide Hazard Using Probability Analysis of Intense Rainfall in the Hoa Binh Province, Vietnam." Natural Hazards 66(2): 707–30. doi:10.1007/s11069-012-0510-0 (2013)

Double Diffusive Natural Convection within a Partially Saturated Porous Cavity Separated by a Wavy Interface: ALocal Thermal Non-Equilibrium Approach

Debayan Bhowmick[1*], Amit Vishwakarma[2], Bimal Das[2] and Deepshikha Datta[3]

[1]Department of Mechanical EngineeringBrainware University
398, Ramkrishnapur Road, Barasat, Kolkata - 125
Corresponding author's email: db.me@brainwareuniversity.ac.in

[2]Department of Chemical Engineering
National Institute of Technology Durgapur, West Bengal, India

[3]Department of ChemistryBrainware University
398, Ramkrishnapur Road, Barasat, Kolkata - 125

Extended Abstract

The present numerical study investigated the influence of porous permeability under local thermal non-equilibrium state within a square cavity partially saturated porous media. The two-dimensional steady flow condition has been maintained along with non- uniform temperature and concentration distribution at the left vertical wall. Moreover, the porous substance has been separated from the free flow region by a wavy interface. To capture the flow physics within the cavity Darcy-Brinkman-Forchhiemer modelhas been incorporated. The Galerkin Finite Element method is implemented to solve the governing differential equations. The parameters regulating the study are varied as: Darcy number ($10^{-2} \leq Da \leq 10^{-3}$); Rayleigh number ($10^3 \leq Ra \leq 10^6$); Thermal conductivity ratio ($0.01 \leq \square \leq 100$); Lewis number ($1 \leq Le \leq 100$) and buoyancy ratio ($-3 \leq N \leq 3$). The outcome of our result indicates that the effect of Le is more significant upon average mass transfer rate (*Sh*) whereas, $\square$ and *N* play dominating role uponthe average solid and fluid heat transfer rate. Overall, the impact of LTNE condition is prominent upon the transport phenomena of fluid and thermal. However, it has less impact upon solute distribution.

Keywords Double diffusive convection; Local thermal non-equilibrium; Partially porous cavity; Darcy-Brinkman-Forchheinermodel; Galerkin Finite Element Method.

References

1. P. Bera, S. Pippal, A. K. Sharma, "A thermal non-equilibrium approach on double-diffusive natural convection in a square porous-medium cavity," Int. J. Heat Mass Transf., **78**, 1080 – 1094 (2014).
2. O. R. Alomar, N. M. Basher, A. A. Yousif, "Natural convection heat transfer from a bank of orthogonal heated plates embedded in a porous medium using LTNE model: A comparison between in-line and staggeredarrangements," Int. J. Therm. Sci., **160**, 106692 (2021).
3. D. Bhowmick, P. R. Randive, S. Pati, H. Agrawal, A. Kumar, P. K. Srivastava, "Natural convection heat transfer and entropy generation from a heated cylinder of different geometry in an enclosure with non-uniform temperature distribution on the walls," J. Therm. Anal. Calorim., **141**, 2405–2427 (2020).

Fake Product Review Detection

Jitendra Musale [1], Samruddhi Shinde [2], Shruti Chaugule [3], Rutuja Thigale [4], Srushti Chavan [5] and Dewendra Bharambe [6]

[1]Computer department of Anantrao pawar college of engineering and research,parvati,pune.
jitendra.musale@abmspcoerpune.org
Samruddhisshinde17@gmail.com

Extended Abstract

Online reviews have become integral to consumer decision-making, yet the prevalence of fake product reviews poses significant challenges to the reliability of these platforms. This paper investigates the detection of fake product reviews, employing a semi-supervised machine learning approach to enhance algorithmic performance. By analyzing parameters such as timing, frequency, and content of reviews, the study aims to improve the accuracy of fake review detection algorithms.

The research explores the evolution of deceptive review tactics and the escalating sophistication of perpetrators. Initially, simplistic patterns such as repetitive language hinted at potential deception, but as perpetrators adapted, detection methods evolved. Today, the challenge requires a multifaceted approach integrating advanced artificial intelligence and machine learning techniques. Various studies cited underscore the difficulty humans face in distinguishing fake from genuine reviews, with heuristic methods often falling short. Hence, the paper advocates for the integration of human observation with algorithmic analysis to improve detection accuracy.

The proposed research methodology employs the Iterative Classification Framework (ICF) method alongside Natural Language Processing (NLP) techniques. The ICF method iteratively refines parameters to enhance detection accuracy, while NLP analyzes linguistic patterns to identify potential fraud. The study introduces a novel concept of using mobile phones for counterfeit product detection, empowering consumers to verify product authenticity through image capture and machine learning analysis. This approach promises widespread accessibility and cost-effectiveness compared to existing anti-counterfeiting technologies.

Furthermore, the paper explores a range of machine learning algorithms including LSTM, Agglomerative Clustering, Support Vector Classification (SVC), K Nearest Neighbors (KNN), Logistic Regression, and LGBM. Each algorithm offers unique advantages in detecting fake reviews, from capturing sequential dynamics to clustering data and establishing decision boundaries. The proposed model offers several advantages over existing systems, including improved trust, fair competition, better consumer decisions, reduced reputation damage, enhanced platform credibility, legal compliance, and increased customer satisfaction. However, ethical and legal considerations are paramount in addressing fake reviews.

In conclusion, the research highlights the significance of combating fake reviews in maintaining the integrity of online platforms. It underscores the potential of machine learning-based approaches, coupled with human oversight, to effectively detect and mitigate fraudulent activities. The paper concludes with a discussion on future research directions, emphasizing the importance of data availability, algorithm efficiency, and ethical considerations in the ongoing fight against fake reviews.

Keyword(s): Machine Learning, Fake Product Review, Poor Reviews, Testing

References

10. Barbados, R., Araque, O., and Iglesias, C. (2019). A framework for fake review detection in a online consumer electronics retailers.and Inf. Process. Manage. 56, 1234–1244. doi: 10.1016/j.ipm.2019.03.002
11. Cha, C., Kim, K., and Myung, J. (2021). "10,000 fake reviews a month… "I"m tired of fighting, should I try to manipulate it", Maeil Business Newspaper.is Available online at: https://www.mk.co.kr/news/society/9844317 (accessed June 10, 2022).
12. Chengzhang, J., Jinhong, Z., and Kang, D. (2014). Survey on fake review detection of e-commerce sites. J. Korea Inst. Inf. Commun. Eng. 1, 79–81.

13. Hyun, J., Yoo, S., and Lee, S. (2019). How to improve the accuracy of recommendation systems: and its combining ratings and review texts sentiment scores. J. Intell. Inf. Syst. 25, 219–239.
14. Invesp (2022). The State of Fake Reviews—Statistics and Trends. are Available online at: https://www.invespcro.com/blog/fake-reviews-statistics

Optimizing Steelmaking Slag Chemistry Using Machine Learning and Genetic Algorithm Integration

Narottam Behera[1], Deep Laha[2], Sandip Kumar Lahiri[2]

[1]Emirates steel Arkan, Abu Dhabi, UAE
Narottam.Behera@Emiratessteel.com

[2]National Institute of Technology Durgapur, India.
sandipkumar.lahiri@che.nitdgp.ac.in

Extended Abstract

In the steel manufacturing sector, achieving precise control over slag chemistry is essential for enhancing production efficiency and minimizing refractory erosion. Predicting iron oxide content (FeO) in slag of Electric Arc Furnaces (EAF) using phenomenological models poses significant challenges due to the complex nature of the process (Ref [1], Ref [2]). Data-driven machine learning models offer a viable alternative, leveraging the abundance of industry data to achieve more accurate and reliable predictions. This study introduces an innovative methodology to forecast steelmaking slag chemistry in Electric Arc Furnaces (EAF) by focusing on reducing FeO% (iron oxide content). The research integrates machine learning techniques with genetic algorithm (GA) optimization to optimize FeO and yield loss in slag effectively.

A comprehensive dataset was compiled from a Direct reduced iron-based EAF from commercial plant, comprising a wide range of input parameters such as raw material composition, operational conditions, and laboratory analyses of slag chemistry. After rigorous data cleaning, a refined dataset of 743 sets of 21 input variables were obtained for model development. Laboratory analysis of FeO % in molten slag is kept as output. Advanced Feature engineering techniques were applied to extract relevant features for slag chemistry prediction. Subsequently, several machine learning algorithms including Linear Regression, Random Forest Regression, Gaussian Process Regression, Support Vector Regression, and Neural Networks were evaluated to identify the most accurate model.

Performance evaluation of the machine learning models was conducted using standard metrics such as Root Mean Square Error (RMSE), Mean Square Error (MSE), R-squared (R2), and Mean Absolute Error (MAE) through 5-fold cross-validation. The results demonstrated exceptional accuracy, with R2 values approaching 0.99 on both training and test datasets by Gaussian Process Regression, indicating minimal discrepancies between predicted and actual FeO % values.

To robustly optimize slag chemistry parameters, a genetic algorithm was employed. The optimization phase focused on minimizing Slag FeO% through genetic algorithm optimization. Five controllable input parameters were identified for optimization, while others were considered as external factors. The genetic algorithm systematically explored these parameters within predefined ranges to minimize FeO% using the established machine learning model. The optimization process led to a significant reduction in Slag FeO%, with the average FeO% decreasing from approximately 26% to 22%. This resulted in a substantial decrease in iron loss per heat and an increase in metallic yield. This combined approach of machine learning and genetic algorithms equips the steel industry with a powerful tool for enhancing process control, reducing production costs, and improving product quality.

In conclusion, this study presents a comprehensive approach to optimizing steelmaking slag chemistry by integrating machine learning and genetic algorithms. The proposed methodology demonstrates significant potential for enhancing process efficiency and profitability in the steel manufacturing industry. By synergistically leveraging the capabilities of machine learning and genetic algorithms, this innovative approach facilitates precise and efficient management of slag chemistry, ultimately leading to higher yields and increased profitability in steel production. Future directions of this research could explore real-time implementation of the developed machine learning and genetic algorithm-based model within steel production facilities for on-site slag chemistry optimization. Additionally, extending the methodology to incorporate advanced sensor technologies and adaptive learning techniques could further enhance the model's predictive capabilities and overall efficiency in the steelmaking process.

Keyword (s): Electric arc furnace, Slag, Machine learning and genetic algorithms

References

1. Zhang, W., et al. (2017). Influence of FeO content and basicity on the viscosity and surface tension of EAF slags. Journal of the Minerals, Metals & Materials Society, 69(8), 1451-1459.
2. Yildirim, I., & Paksoy, T. (2014). Prediction of electric arc furnace slag viscosity and basicity based on charge materials and operating parameters. Metallurgical and Materials Transactions B, 45(5),

Reconditioning Used Lubricant: A Sustainable Approach through Environmentally Friendly Extraction-flocculation Technology with RSM Modelling – A Comparative Analysis

Sayantan Sarkar[1], Deepshikha Datta[2], Gokulan Ravindiran[3] Bimal Das[1*]

[1]Department of Chemical Engineering, National Institute of Technology Durgapur-713209
Sayantansarkar717@gmail.com
[2]Department of Chemistry, Brainware University, Barasat, Kolkata -700125
deepa.datta81@gmail.com
[3]Department of Civil Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hydrabad
*bdas.che@nitdgp.ac.in

Extended Abstract

The prolonged use of lubricating oil generates waste that poses threats to the environment, health, and economy. Proper disposal methods and treatment are imperative for fostering sustainable environmental development. This study focuses on reconditioning of used lubricant through a sustainable approach of environmentally friendly extraction-flocculation technology. Further, the study encompasses a comparative assessment and parametric optimization study using two different flocculants: (i) an inorganic flocculant (KOH) and (ii) a bio-flocculant (Sodium alginate), employing the central composite design approach to achieve maximum percentage recovery of refined oil. The optimized parameters evaluated were extraction time (80 minutes), extraction temperature (50°C and 60°C), solvent to waste oil ratio (7:1 and 3:1), and flocculant dosages (2 g/kg of solvent and 1 g/kg of solvent for 1-butanol-KOH and 1-butanol-sodium alginate solvent-flocculant systems respectively). The yield of refined base oil with 1-butanol-sodium alginate was found to be higher (91.28%) than with 1-butanol-KOH (85.91%) due to its lower tendency for eradicating oxidative products from regenerated oil.

Keyword(s): Waste lubricating oil, Response Surface Methodology, Central Composite Design, Extraction-Flocculation, Yield

References

1. Mohammed, R.R., Ibrahim, I.A.R., Taha, A.H., McKay, G.: Waste lubricating oil treatment by extraction and adsorption. Chem. Eng. J. 220, 343–351 (2013).
2. Sarkar, S., Datta, D., Saha, S., Das, B.: Assessing the efficacy of three bio- based flocculants in the reclamation of spent lubricating oil. Can. J. Chem. Eng. 1–15 (2024).
3. Parekh, K., Shahabuddin, S., Gaur, R., Dave, N.: Prospects of conducting polymer as an adsorbent for used lubricant oil reclamation. Mater. Today Proc. 62, 7053–7056 (2022).
4. Zzeyani, S., Mikou, M., Naja, J., Elachhab, A.: Spectroscopic analysis of synthetic lubricating oil. Tribol. Int. 114, 27–32 (2017).

Exploration of Antibiotic Sensitivity in Breeder Chickens and Antimicrobial Potential of Poultry-associated Microorganisms: A Statistical Comparative Study of Six Different Drugs

Anindya Singha Thakur[1], Sanjukta Dasgupta[1], Bimal Das[2*] and Deepshikha Datta[3]

[1]Department of Biotechnology, Brainware University, Barasat, Kolkata, West Bengal-700125
[2]Department of Chemical Engineering, National Institute of Technology Durgapur, West Bengal, India - 713209
[3]Department of Chemistry, Brainware University, Barasat, Kolkata, West Bengal-700125
*bdas.che@nitdgp.ac.in

Extended Abstract

The poultry industry plays a pivotal role in meeting global food demands, but it faces formidable challenges due to infectious diseases caused by a myriad of pathogens. One of the most concerning issues is the rampant emergence of antibiotic-resistant bacteria, largely attributed to the excessive use of antibiotics in poultry production (1). This study delves into the effects of six different drugs–amoxicillin, gentamicin, cephalexin, co-trimoxazole, oxytetracycline, and ciprofloxacin on breeder chickens, with a particular focus on understanding their impact and the ensuing implications for combating antibiotic resistance in poultry farming.

Through statistical analysis using GraphPad Prism 7 software, it was discerned that breeder chickens exhibited the highest sensitivity to amoxicillin among the tested drugs (2). Further elucidation of amoxicillin's properties using the PubChem tool revealed critical characteristics such as its molecular weight of 365 g/mol and heavy atom count of 24. This comprehensive understanding enhances our insight into the drug's efficacy and its potential implications for poultry health, thereby contributing to informed decision-making regarding antibiotic usage in poultry production.

Moreover, this study embarks on the isolation and identification of diverse microorganisms sourced from various niches within poultry birds, including the gut, skin, and litter. These isolates underwent rigorous screening to assess their ability to inhibit the growth of select poultry pathogens such as Salmonella spp., Escherichia coli, and Campylobacter jejuni. Employing agar diffusion or broth microdilution assays, researchers evaluated the antimicrobial potential of these isolates, with a primary objective of identifying novel agents capable of combating antibiotic-resistant bacteria.

Statistical analyses generated compelling insights into the antimicrobial activity of the isolated microorganisms, with a notable emphasis on their efficacy against Escherichia coli. This discovery holds significant promise for the development of targeted interventions aimed at addressing prevalent bacterial pathogens in poultry production systems. By unraveling the antimicrobial capabilities of poultry-associated microorganisms, this study paves the way for innovative strategies to counter antibiotic resistance and promote sustainable practices within the poultry industry.

The implications of these findings transcend the confines of poultry farming, offering a paradigm shift in antimicrobial discovery and intervention strategies. By harnessing the diverse microbial communities inherent in poultry birds, researchers can potentially tap into a vast reservoir of untapped resources for combating antibiotic resistance. This approach aligns with the overarching imperative of developing forward-thinking solutions to confront the escalating threat posed by multidrug-resistant bacteria. Furthermore, this research underscores the imperative of adopting holistic approaches to safeguard both animal and public health. By elucidating the intricate interplay between antibiotic use, microbial ecology, and pathogen dynamics within poultry production systems, this study furnishes invaluable insights into the multifaceted challenges confronting modern agriculture. Through interdisciplinary collaboration and concerted efforts, stakeholders can strive towards implementing effective strategies to mitigate antibiotic resistance while ensuring the sustainability and resilience of poultry farming practices.

In conclusion, this study represents a significant milestone in advancing our understanding of antibiotic resistance dynamics within the poultry industry. By unraveling the efficacy of various antibiotics and exploring novel antimicrobial agents derived from poultry-associated microorganisms, researchers have laid a solid foundation for future interventions aimed at mitigating the threat of antibiotic-resistant bacteria. These findings underscore the imperative of adopting a multifaceted approach to address the complex challenges facing poultry production, with the ultimate goal of safeguarding public health and fostering sustainability in agriculture.

Keyword(s): Poultry industry, statistical analysis, microorganisms, computational biology

References

1. Bobkov M, Zbinden P. Occurrence of Veterinary Drug Residues in Poultry and Products Thereof. A Review. Chimia (Aarau). 2018 Oct 31;72(10):707-712. doi: 10.2533/chimia.2018.707. PMID: 30376920.

2. Mitteer DR, Greer BD. Using GraphPad Prism's Heat Maps for Efficient, Fine-Grained Analyses of Single-Case Data. Behav Anal Pract. 2022 Jan 3;15(2):505-514. doi: 10.1007/s40617-021-00664-7. Erratum in: Behav Anal Pract. 2022 Jan 18;15(2):515. PMID: 35692516; PMCID: PMC9120324.

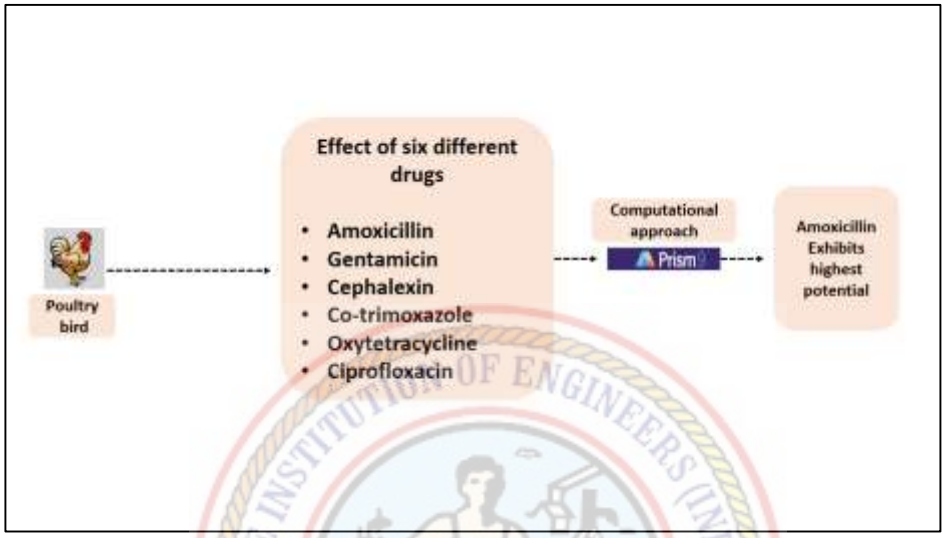


Figure 1: Workflow of the Proposed Study.