Shashank Awasthi · Goutam Sanyal · Carlos M. Travieso-Gonzalez · Pramod Kumar Srivastava · Dinesh Kumar Singh · Rama Kant *Editors*

Sustainable Computing

Transforming Industry 4.0 to Society 5.0

Springer

Sustainable Computing

Shashank Awasthi • Goutam Sanyal •
Carlos M. Travieso-Gonzalez • Pramod Kumar
Srivastava • Dinesh Kumar Singh • Rama Kant
Editors

Sustainable Computing

Transforming Industry 4.0 to Society 5.0

Springer

Editors
Shashank Awasthi
Department of Computer Science and Engineering
G L Bajaj Institute of Technology and Management
Greater Noida, India

Goutam Sanyal
Department of Computer Science and Engineering
National Institute of Technology (NIT)
Durgapur, West Bengal, India

Carlos M. Travieso-Gonzalez
University of Las Palmas de Gran Canaria
Las Palmas de Gran Canaria, Las Palmas, Spain

Pramod Kumar Srivastava
Department of Mathematics
Rajkiya Engineering College
Azamgarh, Uttar Pradesh, India

Dinesh Kumar Singh
Department of Electronics and Communication Engineering
G L Bajaj Institute of Technology and Management
Greater Noida, India

Rama Kant
Department of Computer Science and Engineering
G L Bajaj Group of Institutions Mathura
Mathura, Uttar Pradesh, India

ISBN 978-3-031-13576-7 ISBN 978-3-031-13577-4 (eBook)
https://doi.org/10.1007/978-3-031-13577-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Dedicated to Covid-19 Warriors

Preface

Industry 4.0 focuses on a number of technologies, including artificial intelligence, cloud computing, and the Industrial Internet of Things, which have revolutionised industrial technology by providing benefits such as lower costs, higher-quality goods, and faster data processing. The beauty of these technologies is that they can provide long-term solutions to the problems that our environment faces. These technologies' long-term solutions might be immensely useful to society, and businesses must take advantage of them. Now we're going to Society 5.0, which is based on Industry 4.0 technologies and aims to revolutionise industries such as healthcare, manufacturing, and others. Society 5.0, a word that originated in Japan, will have a favourable influence on some sectors of functioning such as government, business, and politics. The key to Society 5.0 is interconnected data, not financial capital. It will help resolve long-standing social issues like population ageing, social divisiveness, societal collapse, and energy conservation constraints. Indeed, as the Covid-19 pandemic has increased the world's reliance on technology for the better, the evolution of these technologies will speed up. However, the most pressing issue is the long-term viability of the environment and resources. Ecological sustainability and economic growth must coexist. In Society 5.0, striking a balance between technological evolution and sustainability can be easily achieved as sustainability computing spreads its wings in areas like intelligent agent–based systems, social and cloud computing, Internet of Things, sustainable cyber-physical systems, and so on. This book contains valuable research topics that aid in the creation of a clear understanding of the existing and sustainability problems of Industry 4.0, as well as their solutions. Academics, researchers, and professionals working in the fields of Industry 4.0 and Society 5.0 will find this book to be an invaluable resource. It will also contribute to a better understanding of the delicate balance between technological advancement and long-term sustainability for the benefit of society as a whole.

Greater Noida, UP, India — Shashank Awasthi
Durgapur, WB, India — Goutam Sanyal
Las Palmas de Gran Canaria, Spain — Carlos M. Travieso-Gonzalez
Azamgarh, UP, India — Pramod Kumar Srivastava
Greater Noida, UP, India — Dinesh Kumar Singh
Mathura, UP, India — Rama Kant

Acknowledgement

First and foremost, we would like to praise and thank God, the Almighty, for the countless blessings, knowledge, and opportunities that have enabled us to complete the book.

It takes proper guidance, direction, and support to write a thoroughly researched book, and we would like to express our heartfelt gratitude to the intellectual community for their unwavering support.

We are grateful for our parents' blessings, without whom any task would remain a dream. We appreciate their consideration and patience throughout the process.

Prof. Anurag Gupta's tireless efforts and contributions to the completion of this book are greatly appreciated. Our appreciation extends to the entire team for their tireless efforts in completing the technicalities with ease. We sincerely thank you all for your efforts.

Contents

About the Editors

Shashank Awasthi holds a PhD in computer science and engineering. He obtained his MTech in computer science and engineering from Dr APJ Kalam Technical University, Lucknow, and MCA from Dr BR Ambedkar University, Agra. Dr Awasthi is presently working at G. L. Bajaj Institute of Technology & Management, Gautam Buddha Nagar (India), as a professor in CSE Department. His area of interest is WSN and computer networks. He has more than 18 years of teaching and research experience. Dr Awasthi was invited as plenary speaker in the 4th International Conference on Research in Intelligent and Computing in Engineering 2019 (RICE 2019), August 8–9, 2019, at Hanoi University of Industry, Vietnam (HaUI). He has been attending and presenting his research papers at international conferences. He has published approximately 30 papers in international journals/conferences of repute. Dr Awasthi holds membership of IEEE and International Association of Engineers, Hong Kong. He has delivered lectures and chaired sessions in international conferences. He is a member of the editorial boards of various reputed international journals. Dr Awasthi has visited more than eight countries for research purpose. He is certified by Aps Germany, Center for Mechatronics, for Industrial robotics under Indo Euro Synchronization.

Goutam Sanyal received his BTech and MTech degrees from the National Institute of Technology (NIT), Durgapur, and PhD (Engg.) from Jadavpur University, Kolkata, in robot manipulator path planning. He possesses an experience of more than 35 years in the field of teaching, research, and administration. He has published more than 180 papers in international journals and conferences. He has guided 19 PhD scholars in computer science and engineering in the area of steganography, wireless sensor network, computer vision, and natural language processing. He has guided more than 40 PG and 300 UG theses. Dr Sanyal has written a book on computer graphics and multimedia and eight book chapters. He is a reviewer for reputed journals (IET, ELSIEVER, SPRINGER, INDER SCIENCE, and Transactions) and conferences. Dr Sanyal has filed one patent and delivered lectures and chaired sessions in international conferences. He is teaching the following subjects to UG and PG students: computer graphics, computer vision, image processing, VLSI, and computer architecture. He went abroad for research collaboration at the Universities of New Castle and Cranfield in the UK and NUS in Singapore. He is serving as a PhD thesis examiner at VTU, JNTU, OSMANIA University, Anna University, RPTU Bhopal, and NITs. He has served as expert member of faculty selection at NITs and other Universities. Dr Sanyal has wide experience in working with different accreditation bodies. He has served as dean of AA & RG, dean of students' welfare, and dean of faculty welfare at NIT Durgapur for nearly 10 years and is now working as a professor and head of the Department of Computer Science and Engineering. He is a regular member of IEEE, life member of CSI, and a fellow of IEI. Dr Sanyal received national awards, and his biography has been selected for inclusion in Marquis Who's Who in the World 2016, 2017, and 2018 editions.

Carlos M. Travieso-Gonzalez received her MSc degree in 1997 in telecommunication engineering from Polytechnic University of Catalonia (UPC), Spain, and PhD degree in 2002 from the University of Las Palmas de Gran Canaria (ULPGC-Spain). He is Full Professor of Signal Processing and Pattern Recognition and head of Signals and Communications Department at ULPGC; teaching since 2001 subjects on signal processing and learning theory. His research lines are biometrics, biomedical signals and images, data mining, classification system, signal and image processing, machine learning, and environmental intelligence. He carried out researched in 51 international and Spanish research projects, some of them as head researcher. He is co-author of 4 books, co-editor of 24 proceedings books, guest editor for 8 JCR-ISI international journals, and author of 24 book chapters. He has over 440 papers published in international journals and conferences (72 of them indexed on JCR – ISI - Web of Science). He has published seven patents at the Spanish Patent and Trademark Office. He has been supervisor on 8 PhD theses (12 more are under supervision) and 130 master's thesis. He is founder of the IEEE IWOBI conference series and president of its steering committee, of the InnoEducaTIC conference series, and of the APPIS conference series. He is evaluator of project proposals for the European Union (H2020), Medical Research Council (MRC – UK), Spanish Government (ANECA – Spain), Research National Agency (ANR - France), DAAD (Germany), Argentinian Government, and Colombian Institutions. He has been a reviewer in different indexed international journals (<70) and conferences (<250) since 2001. He was vice dean from 2004 to 2010 in the Higher Technical School of Telecommunication Engineers at ULPGC and Vice-Dean of Graduate and Postgraduate Studies from March 2013 to November 2017. He won 'Catedra Telefonica' Awards in Modality of Knowledge Transfer in the 2017, 2018, and 2019 editions.

Pramod Kumar Srivastava is working as dean (student welfare) at Rajkiya Engineering College, Azamgarh (Affiliated to Dr APJ Abdul Kalam Technical University, Lucknow), India. He has a doctorate in mathematics from the University of Allahabad, Allahabad, India. He has several publications in various reputed international journals and international conferences. Dr Srivastava has also published multiple books. His research interests include wireless sensor networks, mathematical modelling, and social networks. He has more than 15 years of teaching and research experience along with PhD guidance. He has successfully guided one PhD student and is presently guiding a few more. Excellence in learning has been his motto since he was a student and the proof of this was the gold medal awarded to him for securing the highest marks in his postgraduation. He has applied this motto in his teaching too, and being well known in his genre, he has been invited to give guest lectures at reputed institutes as well as international conferences. He has been a member of organising committees as well as technical committees of many international conferences. Dr Srivastava is renowned not only for the intellectual treasure he possesses but also for his benevolence and altruism.

Dinesh Kumar Singh received his BE in electronics and communication engineering from Kumaon Engineering College, Dwarahat, Almora, in 2003. He has completed MTech in digital communication from RGPV University, Bhopal, India. Dr Singh has obtained his PhD from the Indian Institute of Technology (ISM), Dhanbad, Jharkhand, India. His area of interest is microwave engineering. He is currently working as an associate professor in the Electronics and Communication Engineering Department, G L Bajaj Institute of Technology and Management, Greater Noida, UP, India. His research interests include design of high-gain, compact, reconfigurable, fractal-shape, and circularly polarised microstrip antennas, substrate integrated wave-guide (SIW,) and magneto-electric (ME) dipole antenna for modern communication system. He has published more than 20 papers in various reputed international journals and conferences. He is also reviewer of the *AEU-International Journal of Electronics and Communication* and *Electronics Letter*.

Rama Kant received his BTech in computer science and engineering from Uttar Pradesh Technical University in 2004, Lucknow, India. He has completed MTech in computer science and engineering at the Institute of Engineering and Technology, Lucknow, India, in 2012. Dr Kant has been awarded PhD by Dr APJ Kalam Technical University, Lucknow, in 2018. His area of interest is theoretical computer science and machine learning. He is currently working as an associate professor in the Computer Science and Engineering Department, G L BAJAJ Group of Institutions, Mathura, UP, India. His research interests include formal methods and verification of large systems, semantics of concurrent programming languages, machine learning, and deep learning. He has published more than 10 papers in various reputed international journals and conferences. He is also reviewer of many international conferences and journals.

A Review on Renewable Energy Sources, Potential and Policy in India

Arvind Kumar Tiwari

Abbreviations

ADB	Asian Development Bank
CASE	Commission for Additional Sources of Energy
CCGTs	Combined Cycle Gas Turbines
CPSUs	Central Public Sector Undertakings
DST	Department of Science and Technology
EPS	Electric Power Survey
GHG	Green House Gas
IAEA	International Atomic Energy Agency
IEPR	Integrated Energy Policy Report
IPPs	Independent Power Producers
MNERS	Ministry of New and Renewable Energy Sources
MNRE	Ministry of New and Renewable Energy
NPCIL	Nuclear Power Corporation India Limited
NSM	National Solar Mission
OWSC	Offshore Wind Steering Committee
PPP	Public-Private partnerships
RES	Renewable Energy Sources
SECI	Solar Energy Corporation of India
SHP	Small Hydro Power
UMWS	Union Ministry of Water Resources

A. K. Tiwari (✉)
Rajkiya Engineering College, Sonbhadra, India

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_1

1 Introduction

In India, the majority of the population lives in rural areas. As per the 2011 census, 68.84% of the Indian population lives in rural areas. India has nearly 0.638 million rural communities; thus, a proposal for electrification of the rural communities based on renewable energy such as biomass will be a powerful option. The "Ministry of New and Renewable Energy Sources" (MNERS) has a proposal to grasp a power of 4324.22 MW through biomass power [1–5, 7]. The MNRE has taken initiatives such as providing economic support and incentives for encouraging the optimal usage of RES [1]. The consumption of biomass as RES is in application in India since ancient eras as animal dung, wood, husk, and several existing usual feed stocks. In current scenario, global warming, decrease in conventional energy resources, and other global problems have directed to the growth of RES. Several nations have encouraged RES power generation through definite plans and financial incentives. Most of the Indian states are facing 3% to 21% energy shortages and 10.3% at the national level, while peak demand scarcity is 15.4% [2, 3]. In 1982, the Indian government set up a "Commission for Additional Sources of Energy" (CASE). The CASE is governed by the DST. In this review paper, various mode-wise RES potentials in India are discussed. The paper discusses the current position of conventional and RES progress and strategies to encourage RES. Forthcoming understanding of strategy on RES is also emphasized. The several types of RES and their potential in India are also discussed. Program regarding providing the grants on RES-founded power generation in India is stated. The remarkable share of this review paper is the policy adopted by the Indian government which is attentive to the development of RES-based power generation regions with tactical strategy and program. The basic reason for energy deficit is the decline in global resources of energy and the exponential growth in the rate of energy requirement [4, 5]. In India, electricity deficit is very common where most of the population has no contact to recent energy facilities. On a usual, for the next 25 years, the demand for electricity will probably see an upswing of 7.4% annually. To achieve the energy demand with a sustainable approach is the major issue and a suitable available option is RES in the present scenario. Hence, it is essential to strengthen the renewable energy program. With renewable power generation, we can achieve enriched energy security state, diminish the need to import, resolve fuel price instability problem, etc. By using 1 GW of renewable power, CO emission can be reduced up to 3.3 million tons annually; consequently, it will support to diminish the contrary effects of environment alteration [1, 6, 7]. The solar and wind renewable power are irregular in nature because of their dependency on solar radiation (or intensity of light) and wind speed, respectively, and also are hardly probable resources. Whereas, tidal wave power is also of intermittent nature but simply probable and trusts on the similar facts of wind turbine. Energy extraction from mixed resources is a challenging task. Under low solar emission circumstances, the photovoltaic panel cannot guarantee the essential solar power generation. Likewise, wind turbine will not operate unless the wind speed is equal to or greater than its cut-in value. Accordingly, to confirm an

Table 1 The common characteristics of RES [2]

Resource	Dispatchable generation	Variability	Predictability
Biofuel	High	Low	High
Biomass	High	Low	High
Geothermal	Medium	Low	High
Hydroelectricity	Medium	Medium	High
Solar	Low	High	Medium
Tidal	Low	High	High
Wave	Low	Medium	Medium
Wind	Low	High	Low

efficient, stable system operation and to avoid energy shortage in the grid, loads are the core aim of energy management. The common characteristics of various RES such as dispatchable generation, variability, and predictability have been shown in Table 1 [2].

The paper is organized as follows: Section 2 briefly overviews the available RES in. Section 3 describes the energy scenario during the 12th Five Year Plan. Extensively, a review on Added Capacity Adding Attained for the duration of the 12th Plan is mentioned in Sect. 4. Section 5 deals with comprehensive studies of Peak Demand and Energy Requisite Prediction. Section 6 discusses the Potential of Renewable Energy Sources in India. Section 7 deals with Capacity Addition from RES during the 12th Plan and Sect. 8 describes the Indian Policy on RES. Section 9 presents the Recent Renewable Energy Initiatives provided by the Indian Government. Section 10 deals with the other Programs for RES, while Sect. 11 focuses on Future Potential of RES in India. Section 12 mentions the Projection of RE generation in India. Finally, Sect. 13 mentions the conclusions of the paper.

2 Available Conventional Energy in India: Overview

India is a growing country in the world, and its population is growing much higher. Consequently, due to high population growth, economical and industrial growth is also high, which mandates high energy demand. The main sources that provide the energy supplies in India are natural fuels, i.e., oil and coal. The common fuel alternatives accessible for electricity generation are coal and lignite, hydro, nuclear, and natural gas as traditional sources, whereas solar, wind, biomass, small hydro, tidal, geothermal, waste, hydrogen/fuel cells are available as renewable energy sources. In India, the installed capacity, as on Sep 30, 2018, was 344,718.61 MW involving 221,802.59 MW thermal (which included 196,097.5 MW from Coal, 24,867.46 MW from Gas, 837.63 MW Diesel), 6780 MW from Nuclear, 45,487.42 MW hydro, while the installed capacity of renewable was 70,648.61 MW as on June 30, 2018, as illustrated in Fig. 1 [3–7]. India has a huge potential for RES

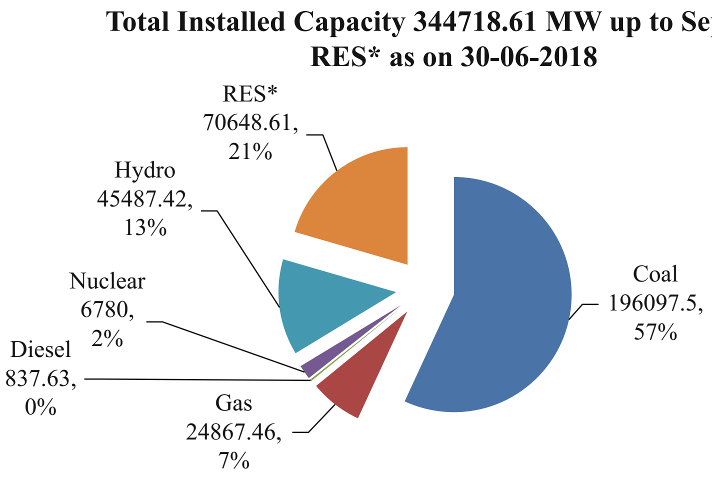


Fig. 1 All-India electrical power installed capacity as on Sep 30, 2018 [3–7]

for electricity generation and, after thermal power, India has achieved 2nd position in renewable power [3–7].

2.1 Generation From Traditional Sources in India

2.1.1 Coal/Lignite

In India, a prominent source for electricity generation is coal, and since low CO_2 emission approach has to be adopted, supplementary sources of electricity essential to be attached in optimum case which are other than coal. In the Indian power sector, coal-based generation of electricity is the pillar and will remain to govern generation of electricity. Supercritical method of processing the coal have been executed in view of environmental concerns. During the 12th Plan, based on supercritical approaches, a total volume of nearly 35,230 MW has been commissioned. In India, lignite is present at a limited number of places such as Neyveli, Surat, and Akrimota and Barsingsar, Palana, and Bithnok. The coal-based installed capacity of electricity generation was 196,097.50 MW as on Sep 30, 2018, approximately 57% of the overall installed capacity in India [1, 3]. The all-India annual coal consumption for electricity generation since 2004–2005 to 2017–18 is exhibited in Fig. 2 [1, 4–7]. This shows that all-India annual coal expenditure for electricity production is increasing exponentially and has doubled within the last decade. It means more coal is required per year due to greater demand for electricity, which can be fulfilled by the expansion of the existing capacity or the installation of new coal-based power plants.

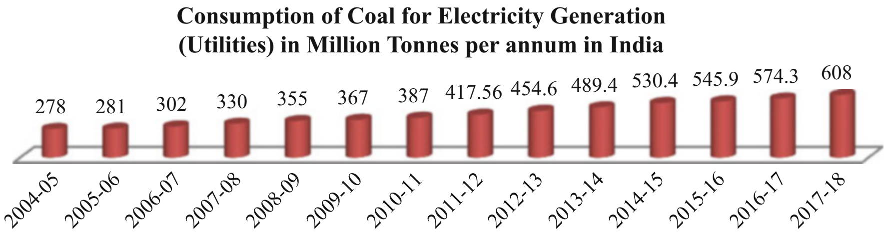


Fig. 2 All-India annual coal consumption for electricity generation [1]

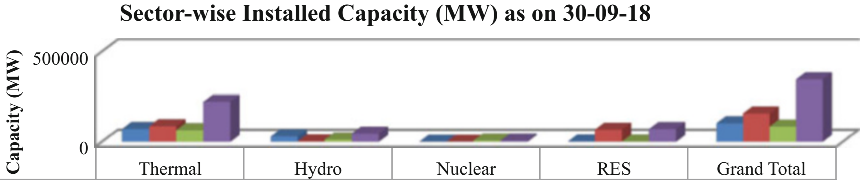
In 2017–18, the coal production was 675.40 million tons, with a maturation of 2.66% ended the preceding year. Although, production of lignite was 46.26 million metric tons in the same year with a growth of 2.27% ended in the preceding year. In world coal production, India is in 2nd position. In India, Andhra Pradesh, Uttar Pradesh, Bihar, Madhya Pradesh, Maharashtra, Orissa, Jharkhand, and West Bengal are the largest coal-producing states [1].

2.1.2 Hydro

The installed capacity for hydro-based electricity generation was 45,487.42 MW as on Sep 30, 2018, which was approximately 13% of the overall installed capacity in India [3, 6, 7]. Hydro-based power plant is utilized to meet its capacity for obtaining heavy loads and entirely fresh plans need to be considered bearing this goal in mind. However, the complete growth of hydro-electric potential, though strictly realistic, encounters many problems, with issues regarding water privileges, relocation of project-affected population, and environmental concerns, etc.

2.1.3 Nuclear

In September 1987, the Indian government formed Nuclear Power Corporation India Limited (NPCIL) under the Companies Act 1956, the Atomic Energy Act 1962 was set up with the objective of electricity generation from atomic power stations [1, 6, 7]. All nuclear power plants functioning through NPCIL are ISO-14001 approved (Environment Management System). At global level, International Atomic Energy Agency (IAEA) was established as a self-governing organization on July 29, 1957 [8]. To encourage peaceful usage of nuclear energy is the core goal of IAEA, and to avoid any military purpose use, composed of nuclear weapons. As per NPCIL, currently, 21 reactors of 6680 MW capacity are installed in India [1, 4–6]. Currently 13 out of 21 reactors are installed for electricity generation of 4280 MW capacities under the supervision of IAEA, which are based on imported oil. However, eight reactors of installed capacity of 2400 MW use native fuel for electricity generation.



	Thermal	Hydro	Nuclear	RES	Grand Total
State	71009.38	29942	0	2004.87	102956.25
Private	86600.3	3394	0	67141.44	157135.74
Central	64192.91	12151.42	6780	1502.3	84626.63
All India	221802.59	45487.42	6780	70648.61	344718.61

Fig. 3 Sector-wise installed capacity as on Sep 30, 2018 [3]

The nuclear-based electricity generation installed capacity was 6780 MW as on Sep 30, 2018, which was almost 2.1% of the overall installed capacity in India [3, 5, 6].

2.1.4 Gas

India has insufficient gas-based power plants and is facing huge generation loss due to lack of appropriate gas supply. Currently, surviving gas power plants are operational at a much small plant load factor (PLF) of approximately 23%, whereas some gas power plants are running inactive due to unobtainability of domiciliary natural gas. The gas-based electricity generation installed capacity was 24867.46 MW as on Sep 30, 2018, which was almost 7% of the overall installed capacity in India [3, 5, 6]. For less CO2 release and with the ability of gas to its fast increase, and decrease nature related problems, gas-based electricity generation is vital. In vision of bulky incorporation of renewable energy sources the benefit of fast ramping ability added significantly. Modern CCGTs have high efficiency of nearly 55% compared to gross efficiency of 40% of coal power plants. The sector-wise installed capacity as on Sep 30, 2018, is shown in Fig. 3 [3]. From this figure, we can observe that the contribution of private sector is more than central as well as state sector, while there is no contribution of private sector in nuclear-based power plant. The contribution of private sector in renewable areas is also more than that compared to the central as well as state sectors.

Figure 4 [1] exhibits annual per capita consumption of electricity since 2005–2006 to 2017–18. This shows that consumption of electricity increased exponentially and doubled within last one decade. Thus, due to industrialization, urbanization, villages' electrification, and development in other fields which require electricity, more power is required day by day and so, expansion of existing capacity or installation of new power plants is needed. Figure 5 shows the growth of rural electrification and pump sets' energization since the 9th Plan to the end of the 12th Plan [1]. From Fig. 5, it is observed that there are wonderful changes in pump sets' energizations from the 11th to the end of the 12th Plan.

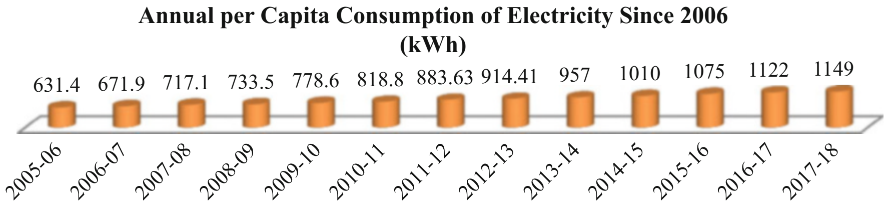


Fig. 4 Annual per capita consumption of electricity since 2005–2006 to 2017–18 [1]

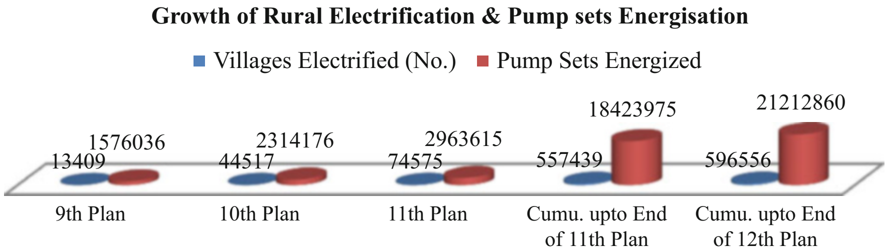
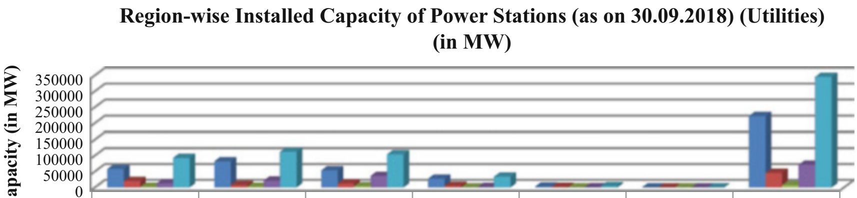


Fig. 5 Growth of rural electrification and pump sets energizations [1]

The region-wise installed capacity of power station as on Sep 30, 2018 is exhibited in Fig. 6 [3–5]. More installed capacity of power (111,527.99 MW) is positioned in the western region than in the southern (103,710.78 MW) and northern regions (92,107.11 MW). From Fig. 6, it is perceived that most of the power stations are based on thermal (221,802.59 MW), and RESs-(70,648.61 MW) based installed capacity in second position out of 3, 44,718.61 MW at all-India status. The hydro-based installed capacity is more in the northern region (19,707.77 MW), while 11,838.03 MW is in the southern region and total capacity at all-India level is 45,487.42 MW. The huge amount of installed capacity of power based on RES is positioned in the southern region (35,535.49 MW). The RES-based installed capacity of power is 20,725.38 MW in western, 13,012.88 MW in northern, 1075.85 MW eastern and 286.46 MW, 12.56 MW in the northeastern region and islands, respectively.

3 Energy Scenario During 12th Five-Year Plan (2012–2017) [1]

The National Electricity Policy, 2005 was decided with the supreme goal to completely attain the entitlement of electricity up to 2012, but the goal achieved successfully by the year 2017. For the 12th plan capacity of 88,537 MW was decided as addition target. Particulars of capacity addition targets and achievements as per sector and mode-wise in the 12th Plans are specified in Table 2 [1, 4, 5].



	Northern	Western	Southern	Eastern	North-Eastern	Islands	All India
Thermal	57766.46	81415.11	53017.26	27301.64	2262.07	40.05	221802.59
Hydro	19707.77	7547.5	11838.03	4942.12	1452	0	45487.42
Nuclear	1620	1840	3320	0	0	0	6780
RES	13012.88	20725.38	35535.49	1075.85	286.46	12.56	70648.61
Grand Total	92107.11	111527.99	103710.78	33319.61	4000.53	52.61	344718.61

Fig. 6 Region-wise installed capacity of power stations as on Sep 30, 2018 [3]

Table 2 Targets and achievements of capacity addition in the 12th Plan (in MW) [1]

	Thermal		Hydro		Nuclear		Total	
Sector	Target	Ach.	Target	Ach.	Target	Ach.	Target	Ach.
Central	14,878	15,869	6004	2584	5300	2000	26,182	**20,453**
State	13,922	22,201	1608	2276	0	0	15,530	**20,453**
Private	43,540	53,661	3285	619	0	0	46,825	**54,280**
All India	**72,340**	**91,730**	**10,897**	**5479**	**5300**	**2000**	**88,537**	**99,209**

Table 3 Target and achievement of capacity addition for 2017–2018 (in MW) [1, 3]

	Thermal		Hydro		Nuclear		Total		%
Sector	Target	Ach.	Target	Ach.	Target	Ach.	Target	Ach.	
Central	4880	3170	800	390	500	0	6180	**3560**	**57.61**
State	3546	1760	300	200	0	0	3846	**1960**	**50.96**
Private	2940	3780	205	205	0	0	3145	**3985**	**126.71**
Total	**11,366**	**8710**	**1305**	**795**	**500**	**0**	**13,171**	**9505**	**72.17**

Table 2 shows the capacity addition achievement was 99,209 MW in the 12th Five-year Plan against the target set of 88,537 MW capacities from traditional energy sources. The private sector's achievement was 54,280 MW against the target set by 46,825 MW, which is more than that of central and state sectors' achievement, i.e., 26,182 MW and 15,530 MW, respectively, against the target set of 26,182 MW and 15,530 MW, respectively. For the thermal, capacity, the additional target set was 72,340 MW, while for hydro and nuclear, it was 10,897 MW and 5300 MW, respectively. The target and achievement of capacity addition for 2017–2018 and capacity addition targets/achievements during Sep-2018 from conventional energy sources for different sectors are exhibited in Table 3 [1, 3] and Table 4, respectively [3].

Table 4 Targets/achievements of capacity addition during Sept. 2018 (in MW) [3]

Schemes	Sector	Target 2018–2019	Sep-18		April 2018– Sept 2018		Deviation
			Target	Ach.	Target	Ach.	(+) / (−)
Thermal	Central	2760	0	0	0	0	0
	State	4506.15	0	0	0	0	0
	Private	0	0	0	0	0	0
	Total	**7266.15**	**0**	**0**	**0**	**0**	**0**
Hydro	Central	710	0	0	110	110	0
	State	130	63.33	30	129.99	30	−99.99
	Private	0	0	0	0	0	0
	Total	**840**	**63.33**	**30**	**239.99**	**140**	**−99.99**
Nuclear	Central	**0**	**0**	**0**	**0**	**0**	**0**
All India	Central	3470	0	0	110	110	0
	State	4636.15	63.33	30	129.99	30	−99.99
	Private	0	0	0	0	0	0
	Total	**8106.15**	**63.33**	**30**	**239.99**	**140**	**−99.99**

Table 5 Generation of electricity for Sept. 2018 (in BU) [3]

Type	Achievement Sept. 2017	Targets Sept. 2018	Achievement Sept. 2018	% Change w.r.t. 2017
Thermal	84.87	88.21	87.18	2.72
Nuclear	2.67	2.617	2.56	−4.16
Hydro	14.10	15.741	17.42	23.54
Bhutan Import	0.855	0.674	0.884	3.39
All India	102.494	107.239	108.044	5.41

Table 6 Generation of electricity for the duration of April 2017–Sept. 2017 and April 2018–Sept. 2018 (in BU) [3]

Type	April 2017– Sept 2017	April 2018– Sept. 2018	% Change w.r.t April 2017– Sept 2017
Thermal	509.00	529.43	4.01
Nuclear	17.06	18.911	10.85
Hydro	81.36	83.370	2.44
Bhutan Import	3.805	3.805	0.00
All India	611.25	635.517	3.97

The attainment of electricity generation was 108.044 BU against the target set of 107.239 BU from conventional energy sources for September 2018 as shown in Table 5 [3].The achievement of electricity generation for September 2017 was 102.494 BU and the percentage change for September 2018 was 5.41 with respect to 2017. Table 6 [3] exhibits the generation of electricity for the duration of April 2017–Sept. 2017 and April 2018–Sept. 2018, which exhibits that the percentage change for April 2018–Sept. 2018 was 3.97 with respect to April 2017–Sept. 2017.

Table 7 Generation capacity addition for Sept. 2018 (in MW) [3]

Type	Achievement	Targets	Achievement	% Change w.r.t. Sept. 2017
	Sept. 2017	Sept. 2018		
Thermal	510	0	0	0
Hydro	100	63.33	30	47.37
Nuclear	0	0	0	0
All India	**510**	**63.33**	**30**	**47.37**

Table 8 Producing capacity addition for the duration of April 2017– Sept. 2017 and April 2018–Sept. 2018 (in MW) [3]

Type	April 2017–Sept. 2017	April 2018–Sept. 2018	% Change
Thermal	4300	0	−100
Hydro	278	140	−50
Nuclear	0	0	0
All India	**4578**	**140**	**−97**

The target and achievement of generation capacity addition for Sept. 2018 for the conventional energy sources is presented in Table 7 [3]. During this time, only hydro power was targeted for 63.33 MW and achievement was 30 MW. The achievement of generating capacity addition of hydro for Sept. 2017 was 100 MW and the percentage change for Sept. 2018 was 47.37 with respect to Sept. 2017. Table 8 [3] exhibits producing capacity addition for the duration of April 2017–Sept. 2017 and April 2018–Sept. 2018, which exhibits that the percentage change for April 2018– Sept 2018 was −97% with respect to April 2017– Sept. 2017.

Table 9 presents the region-wise power supply status (Energy & Peak) for Sept. 2017 and Sept. 2018 [3]. The availability of energy for Sept. 2018 was 109,099 MU against the requirement of energy of 109,640 MU, i.e., a deficit of 0.5%., while for Sept. 2017, the availability was 101,561 MU against the requirement of energy of 102,465 MU, i.e., a deficit of 0.9%. The peak met of power for Sept. 2018 was 175,528 MW against the peak demand of power of 176,538 MW, i.e., a deficit of 0.6%., while for Sept. 2017, the peak supply of power was 158,550 MW against peak demand of power of 162,452 MW, i.e., a deficit of 2.4%.

4 Added Capacity Attained for the Duration of 12th Plan (Not Involved in the Targeted Capacity of 88,537 MW) [1, 3]

For effortlessly, the establishment of coal-based power plants has been relicensed according to Electricity Act, 2003. Consequently, a capacity of 35,296.6 MW not included in the 12th Plan goal has returned profits in the 12th Plan. Table 10 [1,

Table 9 Status of power supply (energy & peak) in Sept. 2018 [3]

Region	Energy (MU)				Surplus(+)/Deficit (−) in %	
	Requirement		Availability			
	Sep-2017	Sept. 2018	Sept. 2017	Sept. 2018	Sept. 2017	Sept. 2018
Northern	34,454	33,357	33,824	32,957	−1.8	−1.2
Western	29,344	32,764	29,232	32,745	−0.4	−0.1
Southern	24,727	28,691	24,685	28,631	−0.2	−0.2
Eastern	12,414	13,291	12,345	13,278	−0.6	−0.1
North-Eastern	1526	1536	1475	1489	−3.3	−3.1
All India	**102,465**	**109,640**	**101,561**	**109,099**	**−0.9**	**−0.5**
Region	**Power (MW)**				Surplus(+)/Deficit (−) in %	
	Peak demand		Peak met			
	Sept. 2017	Sept. 2018	Sept. 2017	Sept. 2018	Sept. 2017	Sept. 2018
Northern	57,203	56,409	54,649	55,650	−4.5	−1.3
Western	46,382	52,933	45,710	52,895	−1.4	−0.1
Southern	41,071	45,587	40,852	45,428	−0.5	−0.3
Eastern	20,274	21,781	20,208	21,781	−0.3	0.0
North-Eastern	2629	2921	2520	2850	−4.1	−2.4
All India	**162,452**	**1,76,538**	**158,550**	**1,75,528**	**−2.4**	**−0.6**

Table 10 Added capacity attained for the duration of 12th Plan (in MW) [1, 3]

Sector	Hydro	Thermal			Nuclear	Total
		Coal	Gas/LNG	Total		
State	0.0	3920.0	51	3971	0.0	**3971.0**
Private	9.0	8520.0	397.1	8917.1	0.0	**8926.1**
Central	24.0	18,445.0	3930.5	22,375.5	0.0	**22,399.5**
Total	**33.0**	**30,885.0**	**4378.6**	**35,263.6**	**0.0**	**35,296.6**

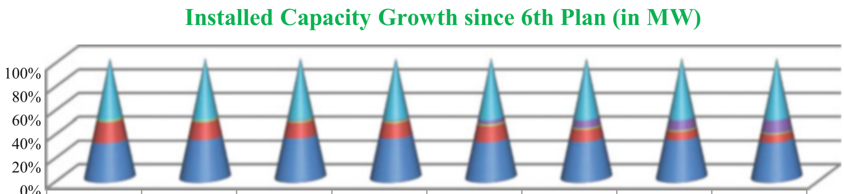
3] presents the sectorwise and modewise added capacity attained for the duration of the 12th Plan. This comprises involvements from the private sector of 22,399.5 MW, around 63.5% of overall capacity addition, which is exterior to the goal.

4.1 *A Cooperative Analysis of 12th Plan with Earlier Five-Year Plans*

Table 11 [1, 3] exhibits the target and achievement capacity addition for the duration of earlier five-year plans. It is observed that, during the 9th Plan capacity addition, achievement was less than the 8th plan. It also exhibited that for earlier five-year plans, capacity addition achievement was little of the target; however, the capacity addition achievement was 112% of the target during 12th Plan.

Table 11 Target and achievement of capacity addition in previous five-year Plans (in MW) [1, 3]

Plan/sector	8th plan		9th plan		10th plan		11th plan		12th plan	
	Target	Actual	Target	Actual	Target	Actual	Target	Actual	Target	Actual
State	14,870	6835	10,748	9353	11,157	6245	26,783	16,732	15,530	23,277.4
Private	2810	1430	17,589	5262	7121	1930	15,043	23,012	46,825	55,479.5
Central	12,858	8157	11,909	4504	22,832	13,005	36,874	15,220	26,182	20,452.6
Total	30,538	16,423	40,245	19,119	41,110	21,180	78,700	54,964	88,537	99,209.5
% Achievement	**53.7**		**47.5**		**51.5**		**69.84**		**112.1**	



	End of 6th Plan	End of 7th Plan	End of 8th Plan	End of 9th Plan	End of 10th Plan	End of 11th Plan	End of 12th Plan	Up to Sep-2018
Total	42584.72	63636.34	85795.37	105045.96	132329.21	199877.03	319606.3	344718.61
RES	0	18.14	902.01	1628.39	7760.6	24503.45	50018	70648.61
Nuclear	1095	1565	2225	2720	3900	4780	6780	6780
Hydro	14460.02	18307.63	21658.08	26268.76	34653.77	38990.4	44478.42	45487.42
Thermal	27029.7	43745.57	61010.28	74428.81	86014.84	131603.18	218329.88	221802.59

Fig. 7 Installed capacity growth since the 6th Plan (in MW) [1]

Figure 7 exhibits the growth of installed capacity during the previous eight-year plans, i.e., since the 6th Plan [1]. It is observed that installed capacity of thermal, hydro, and nuclear increases in each five-year plan, but the installed capacity of RES increases very fast in comparison to conventional sources.

4.2 Slipped Projects and Causes of Slippage from 12th Plan Capacity Addition Target [1]

From the 12th Plan, a total of 24,613.8 MW capacities (29% of 88,537 MW, i.e., the targeted capacity addition) have slipped. Table 12 exhibits the sectorwise and modewise particulars of capacity slipped [1]. It is observed that the maximum projects slipped from private sector (56%), while lowest in state sector (5%) and 39% from central sector. The modewise maximum slipped projects were in thermal sector (65%), whereas the lowest was in the nuclear sector (13%) and 22% from the hydro sector. The detail of statewise capacity slipped and the power projects slipped from 12th Plan capacity addition target is given in [1].

Table 12 During the 12th Plan, capacity slipped from capacity target of 88,537 MW (in MW) [1]

Sector	Hydro	Thermal			Nuclear	Total
		Coal	Gas/LNG	Total		
State	541.0	600.0	37.8	637.8	0.0	**1178.8 (5%)**
Private	1490.0	12,245.0	0.0	12,245.0	0.0	**13,735.0 (56%)**
Central	3420.0	2980.0	0.0	2980.0	3300.0	**9700.0 (39%)**
Total	**5451.0(22%)**	**15,825.0**	**37.8**	**15,862.8(65%)**	**3300.0(13%)**	**24,613.8**

[a]Excludes 10 MW downward capacity revision in respect of Hinduja TPP [1]

Table 13 Regionwise electrical energy requirement (utilities) (MU) [1, 3]

Region	2016–17	2021–22	2026–27
Northern	356,521	468,196	616,345
Western	352,304	481,501	627,624
Southern	307,047	420,753	550,992
Eastern	128,300	171,228	217,468
North- Eastern	15,876	23,809	34,301
Islands	381	537	705
All-India (Electrical Energy Requirement)	**1160,429**	**1566,023**	**2047,434**

The sluggish growth of civil works, poor geology, and unfavorable weather conditions are the main causes of slippage for hydro projects. Some other issues are concerned with law and order, funds' limitations, agreement, environmental, local difficulties, resettlement and rehabilitation that contributed to capacity slippage during the 12th plan. The main problems which concern with slippage of thermal projects are different from those in hydro projects, such as problems in land acquisition, which is required for establishing of plants, transmission lines, etc. The other causes also responsible for slippage of the plants such as state policies; availability of startup power at site, lack of natural gas, issues with rate invades on account of interruption in well-timed end of projects, problems in fund from banks and financial institutions, poor performance of main contractor, contractual disputes, natural calamities, and extreme weather conditions, etc.

5 Peak Demand and Energy Requisite Prediction

Based on the 19th report of Electric Power Survey (EPS), assessment of electricity demand has been carried out by generation expansion planning studies. The regionwise expected necessity of peak demand and energy are given in Tables 13 and Table 14, respectively [1, 3].

Table 14 Regionwise peak demand of electricity (utilities) (MW) [1, 3]

Region	2016–17	2021–22	2026–27
Northern	55,596	73,770	97,182
Western	50,141	71,020	94,825
Southern	44,782	62,975	83,652
Eastern	20,883	28,046	35,674
North-Eastern	2810	4499	6710
Islands	77	108	142
All-India (Electrical energy requirement)	**161,834**	**225,751**	**298,774**

6 Renewable Energy Sources: Potential in India

Renewable sources that are frequently restocked by nature are sun, wind, hydro, biomass, geothermal, idal energy, etc. This comprises heat and electricity produced from the above-mentioned resources. India is the fourth leading electricity user country at global level after the USA, China, and Russia. At present, India has more programs for RESs in the world. For research and development expertise and for manpower advancement in renewable energy sector, the MNRE has been loyal. Huge potential of RES is available in different forms such as solar, SHP, wind, biomass, etc. An overall capacity of 310 GW has been set up, which involves 69.4%, 13.9%, 14.8%, and 1.9% power generation from thermal, hydro, renewable, and nuclear plants, respectively [1, 5, 6]. After thermal power, India has attained 2nd position in renewable power and renewable power scattering speedily [1]. The Indian government has decided on a goal for renewable power of 175 GW capacity. This comprises 100 GW solar power, 60 GW wind power, 10 GW bio-power, and 5 GW small hydro power targeted for 2022 [1, 5, 6]. The MNRE is implementing extensive schemes with monetary and economic backing and open-minded programs to attain this goal. The Ministry has taken quite a lot of steps to achieve visualization of clean energy. The principal goals of usage of renewable energy in India have been electricity unavailability, energy access, climate variation, energy security, etc.

6.1 Solar Energy

India has a huge potential of solar power owing to its promising solar region from 400 S to 400 N positions in world map. Normally, most regions of India receive an average solar radiation of 4–7 kWh/m^2/day, since there are nearly 300 strong sunny days in a year. Among all accessible different RES in India, solar energy has the maximum potential. Solar powers with a total capacity of 70,648.61 MW were installed as on Sep 30, 2018 against probable potential of about 100,000 MW in India [3, 9]. When the sun hits the atmosphere, it is at 1017 W, and on the earth's

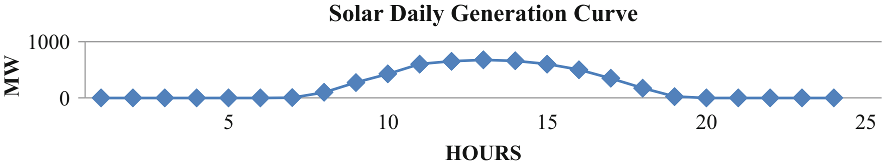


Fig. 8 Typical daily profile of a solar power generation [1]

surface, it is at 1016 W, then solar power can be easily harnessed. For all needs of civilization, a total demand for power is 1013 W at global prominence. However, sun provides 1000 times more power than total demand of civilization in the world. If only 5% of solar energy is used, then 50 times world energy requirement will be fulfilled. In a year, the corresponding energy potential is around 6000 million GWh of energy. National Solar Mission (NSM) has goal for grid solar power at 20,000 MW and off-grid capacity 2000 MW for which solar thermal collector areas of 20 million square meter and solar lighting of 20 million is under execution by 2022 [1, 9]. In January 2010, NSM has given an excessive enhancement to solar power development in the country. As on March 31, 2017, a capacity of 12,288.83 MW solar power was installed in India, which is around 21.5% of total capacity installed of renewable energy [1, 3]. This installed capacity of solar power is mostly spread in the following states in descending order of installed capacity such as: Andhra Pradesh (1867.23 MW), Rajasthan (1812.93 MW), Tamil Nadu (1691.83 MW), Telengana (1286.98 MW), Gujarat (1249.37 MW), Madhya Pradesh (857.04 MW), Punjab (793.95 MW), etc. [1, 9]. Figure 8 exhibits the typical daily profile of solar power generation [1]. This exhibited that the solar power generation slowly rises once begin, extents an extreme position approximate in noon and then slowly falls down in evening.

6.2 Wind Energy

In India, wind-based power schemes are mostly spread in regions such as southern, western, and northern states, whereas there is no grid link of wind power plant in the eastern and northeastern states. The electricity generation based on wind in India is extremely affected by the monsoon. In the context of Indian monsoon, strong monsoon starts in the month of May–June from the southwest, and moist air travels in the direction of the land when cool, whereas a weaker monsoon starts in the month of October from the northeast and dry air travels in the direction of the ocean when cool [10–12, 14]. Wind speed is comparatively weak in the month of November to March. Wind power generation is at the mercy of wind speed which is influenced by daily and seasonal weather patterns. Figure 9 exhibits the profile of a typical daily wind power generation [1]. In India, wind-based power generation program was initiated since 2010–11, i.e., finish of the 6th five yearly

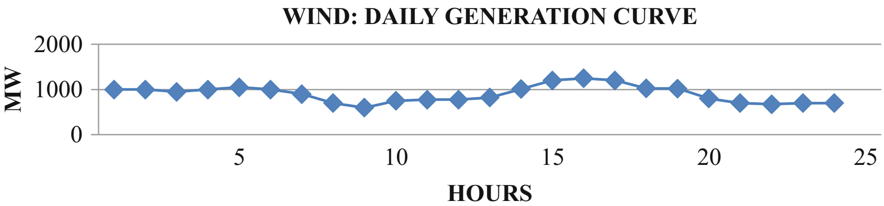


Fig. 9 Profile of typical daily wind power generation [1]

plans and substantial growth can be seen in the last few years. The main objectives of program were to commercialization of wind power generation, supporting for research and development, afford assistance on wind projects and also to create attentiveness in citizens. In connection with this program, MNRE has done several amendments concerning policies, schemes, and incentives towards the development of wind energy [13, 15]. Compared to Denmark or the USA, India is a beginner in the wind energy sector. Due to interest of the Indian government in wind energy sector, it is rated fifth with leading wind power installed capacity. Presently, after the USA, China, Spain, and Germany, India has fifth rank in installed wind power capacity of 34,293.48 MW as on June 30, 2018 [3]. In India, wind power plants are mostly positioned in the southern, western, and northern regions and the installed capacity was 32,279.77 MW as on March 31, 2017, which is around 56.4% of total installed capacity of renewable [1, 3]. In India the wind power installed capacity is mostly spread in following states in descending order of installed capacity such as: Tamil Nadu (8631 MW), Gujarat (6044 MW), Maharashtra (4789 MW), Rajasthan (4300 MW), Karnataka (4544 MW), Andhra Pradesh (4007 MW), Madhya Pradesh (2520 MW), Telangana (128 MW), Kerala (53 MW) [1]. Muppandal wind farm located in Tamil Nadu is the biggest in the country with a total capacity of 1500 MW [1]. Under MNRE, Indian Renewable Energy Development Agency (IREDA) offers several incentives in order to rapidly grow wind power plants. In National Electricity Plan (2012), the target of total power generation capacity of about 350–360 GW by the year 2022 was stated by the Central Electricity Authority. In August 2012, MNRE established Offshore Wind Steering Committee (OWSC) and, in 2013, OWSC released a draft on National Offshore Wind Energy Policy. As on March 31, 2017, grid-connected installed capacity of wind power was 32,279.77 MW [1].

6.3 Biomass Energy

Other sources of RESs are the usage of living and lifeless forms of biological material for energy generation as biomass. Biomass can be used as energy either as heat, which is produced directly via combustion, or after transforming it to different forms of biofuel indirectly. Presently, direct-fired approach-based biomass plants are mostly used, which are similar to most fossil-fuel-fired power plants. India is a country with an agriculture-based economy and has massive potential for biomass. The main objective of biomass power/cogeneration program implementation is encouraging expertise for optimal use of wastes, cotton stalk, soya husk, rice husk, bagasse, groundnut/coconut shells, saw dust, de-oiled cakes, straw, etc., as biomass resources. Presently, the contribution of biomass is about 14% of the total energy amount globally, and 38% of this energy is spent in developing nations mostly in the rural and traditional areas of the economy [1, 5, 16]. In India, from agriculture, forestry, plantations, and agro-industrial residues, around 540 million tons yearly is the assessed availability of biomass. With current existing expertise assessment, the grid class power of over 16,000 MW can be produced by using surplus agricultural residues [1, 5]. In addition, it is the assessment that if entirely 550 sugar mills shift over to recent procedures of co-generation, then almost 5000 MW of power can be produced. Consequently, it is reflected that India has about 21,000 MW power potential through biomass. Ever since the mid-1990s, the Indian government has been executing programs on biomass power/cogeneration. In India, through 167 projects on biomass powers and cogeneration of capacity, 1650 MW have been fitted for serving power to the grid [1, 5]. Additionally, a capacity of 1850 MW electric power is under various stages of execution through nearly 171 projects based on biomass and cogeneration. From sugar mills, which comprise 82 projects based on cogeneration projects with fitted capacity adding to 690 MW. Additional 107 projects of 1280 MW capacity are under execution [1, 5, 16]. In India, some states are at top position in cogeneration projects implementation such as Andhra Pradesh, Tamil Nadu, Karnataka, and Uttar Pradesh. However, for biomass-based power projects, the topmost states are Andhra Pradesh, Karnataka, Chhattisgarh, Maharashtra, and Tamil Nadu. As on June 30, 2018, a capacity of 8839.10 MW biomass power was installed, this is almost 14.5% of whole fitted capacity of renewable, whereas installed capacity of biomass for grid-connected was 8295.78 MW as on March 31, 2017 [1, 3, 16]. Bagasse is known as waste from sugar mills which is used in bagasse cogeneration [1, 16]. Consistent with MNRE, via biomass and bagasse cogeneration it is projected that 73,000 MW power will be produced by 2032 [1, 16].

6.4 Small Hydro Power (SHP)

Hydro-based power plants below 25 MW capacities are characterized as small hydro power plants (SHP). In India, the assessed potential is approximately 20,000 MW for power generation from such plants, out of which about 2.5 GW has been developed [1, 5, 17]. According to the potential available river-based projects are situated in Himalayan states and irrigation waterways based situated in other states. In these plants, the water flow is utilized to rotate the turbine blades which drive the mechanically coupled rotor of generator for generating electrical power. Generated electricity through hydro-based system is the function of turbine size and the amount

Table 15 RESs installed capacity as on June 30, 2018 (in MW) [1, 3, 5]

Small hydro power	Wind power	Bio-power		Solar power	Total capacity
		BM power/Cogen.	Waste to energy		
4493.20	34293.48	8700.80	138.30	23022.83	70648.61

of water flowing through the turbine. In remote regions, SHP plants are mostly used as separate power systems. In India, installed capacity of SHP was 4493.20 MW as on June 30, 2018, which is about 7.17% of overall installed capacity of renewable [1, 3, 5, 16]. Ambition of country for grid interactive power generation capacity is that the 2% of total power should come from SHP. The Indian government has currently fixed a goal of 60,000 MW from SHP projects by 2022 [1, 3, 16]. For existing SHP projects under the public sector, only the government has provided grant for its overhaul and modernization. Grid-connected installed capacity of SHP plants was 4379.86 MW as on March 31, 2017 [1, 5, 16].

6.5 Ocean Surface Waves and Tidal Power

There is a method to harness energy of ocean surface waves as wave power and energy of tides as tidal power. These are two types of hydro power with future potential. Nevertheless, they are not broadly working commercially [1, 5].

6.6 Installed RES Potential in India

Table 15 exhibits RESs installed capacity of 70,648.61 MW as on June 30, 2018 [1, 3, 5, 16]. Figure 10 represents the growth of RES since the 6th Plan [1, 3, 5, 16]. It is observed that at the end of the 10th Plan, installed capacity of RES was 7760.6 MW, whereas at the end of the 11th Plan, what was achieved was 24,503.45 MW. At the end of the 12th Plan, it was 50,018 MW, and up to Sept. 2018, total installed capacity achieved was 70,648.61 MW [3, 5].

Generation of electricity through conventional and renewable sources in Sept. 2017 and Sept. 2018 is given in Fig. 11 [3]. Total growth of electricity generation is achieved by 8.24%, while 46.9% in renewable generation and 5.41% in conventional generation has been achieved.

All India power generation from RESs achieved 62,659.401 MU from April 2018 to August 2018 as shown in Fig. 12 [3]. Out of which, wind power contributes 39,314.548 MU, solar power 14,702.788 MU, biomass power 1205.59 MU,

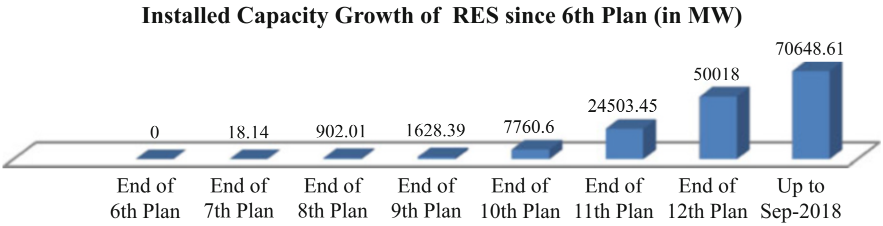


Fig. 10 Installed capacity growth of RES since the 6th Plan (in MW) [1, 3, 5]

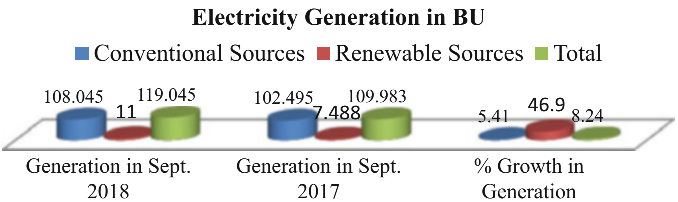
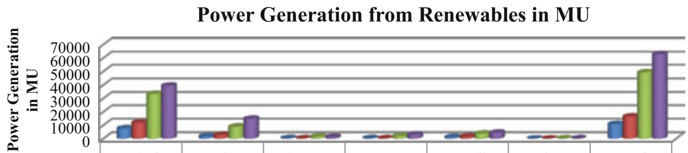


Fig. 11 Electricity generation in BU through conventional and renewable sources in Sept. 2017 and Sept. 2018 [3]



	Wind	Solar	Biomass	Bagasse	Small Hydro	Others	Total
Aug-17	7512.157	1660.26	264.182	281.647	1070.064	28.614	10816.924
Aug-18	11771.655	2635.793	185.61	218.277	1531.777	32.532	16375.645
Cumulative for April-2017-Aug-2017	33172.505	8808.937	1468.707	1814.848	3705.692	149.672	49120.361
Cumulative for April-2018-Aug-2018	39314.548	14702.788	1205.59	2923.271	4335.633	177.571	62659.401

Fig. 12 All India power generation from RESs [3]

Table 16 Grid-connected installed capacity of renewable power plants (as on Mar 31, 2017) [1]

Renewable energy source	Installed capacity (MW)
Solar power	12,288.83
Wind power	32,279.77
Bio-power & waste power	8295.78
Small hydro power	4379.86
Total	**57,244.24**

cogeneration bagasse 2923.271 MU, small hydro power 4335.633 MU, and others 177.571 MU. Table 16 exhibits grid-connected installed capacity of renewable power plants in India [1]. State-wise grid-connected installed capacity of renewable power plants as on Mar 31, 2017, has been given in [1].

Table 17 Capacity addition target of RES in 12th plan [1]

Source	Capacity (MW)
Solar	10,000
Wind	15,000
Other RES	5000
Total	**30,000**

Table 18 Installed capacity of RESs as on March 31, 2017 [1, 3]

Source	Capacity (MW)
Solar	12,288.83
Wind	32,279.77
Bio-power and waste power	8295.78
Small hydro	4379.86
Total	**57,244.24**

Table 19 Source-wise capacity addition reached from RESs during the 12th Plan [1]

Source	Capacity (MW)
Solar	11,347.53
Wind	15,383.17
Bio-power and waste power	5040.78
Small hydro	969.36
Total	**32,740.84**

7 Capacity Addition from RES in 12th Plan

During the 11th Plan, a capacity addition of 16,744 MW was achieved from RESs, although at the end of the 11th Plan (2007–12), the installed capacity of RESs was 24,503 MW. For the duration of the 12th Plan, capacity addition of 30,000 MW was targeted from RES; details are provided in Table 17 [1, 3, 5, 16].

Although, by 2021–22, the goal of installed capacity of RESs has been reviewed to 175 GW [1, 3]. Table 18 exhibits the installed capacity of RESs as 57,244.23 MW as on March 31, 2017 in India [1, 3]. From Table 18, present scenario of installed capacity of RES displays that although wind plants are leading nowadays but in forthcoming installed capacity of solar plants shall have more comparison to wind plants.

As on March 31, 2017, a capacity addition of 32,740.84 MW has been reached from RESs during the 12th Plan. Table 19 exhibits details of source-wise capacity added during the 12th Plan [1].

8 RES Policy of India

In current scenario of India, due to fast population and high economic growth, energy consumption has been growing very rapidly. Expected demand of electricity has increased due to fast development and improved human living status of millions of Indian families. Consequently, Government of India is creating several plans and strategies in the energy sector at present. Subsequently, viable progress is now the significant goal of the world and RES are considered as the most favorable option for power generation. MNRE has implemented several schemes, policies, and strategies in the RESs' sector and is also encouraging adoption of these approaches by offering many grants and incentives [1, 5, 16]. MNRE is providing many kinds of grants for both sectors, private and the government [1, 5, 16].The renewable power sector has appeared, for instance, to play the major role in the grid connected power generation [16]. The Indian government has taken many steps for the development through RES such as the concept of solar parks, organizing RE-Invest 2015-a global investor's' meet, initiation on grid connected rooftop solar scheme, concept of Green Energy Corridor reserving amount of Rs.38,000crore for it, scheme of mounting 100,000 solar pumps and solar installations under the Surya Mitra scheme training 50,000 people, for solar and wind power free inter-state transmission charges etc. [1, 5, 16]. The Energy and Resources Institute (TERI) report on Pricing of power from Non-Conventional Sources is described in [18]. The Statewise Track of Power Scenario in India for 2018 is discussed in [19]. Annual report of 2017 reported by World Institute of Sustainable Energy, Pune is given in [20]. The Electricity Act 2003 and other guidelines for consumers are mentioned in detail in [21–28].

9 Recent Renewable Energy Initiatives

9.1 Solar Parks

In India, for quick growth of solar-based projects, the concept of solar parks model has been introduced. The Indian Government has sanctioned a total of 34 solar parks for 21 states with a capacity of 20,000 MW [1, 16]. In order to minimize the risks associated with the solar projects, the state governments or EPC developers have taken the initiatives for establishing their solar projects (SP). Charanka solar park is the first established solar park in the country in Gujarat. The cost of big size solar power projects hassles, therefore, capacity sizes of 500 MW (Ultra-Mega projects) or above are planned. For solar park growth, big portions of land are presented in a few states. The Indian Government has set a goal of mounting solar power plants of 100 GW by 2022. Table 20 shows the state-wise approved capacity of solar parks [1, 16].

9.2 Solar Cities

In India, development, economic growth, and increasing population are the main causes of rapid increase in energy demand, which boosted Green House Gas (GHG)

Table 20 State-wise approved capacity of 34 solar parks for 21 states [1, 16]

States	Approved capacity of solar parks (in MW)
Jammu Kashmir, Arunachal	100 (one solar park in each state)
Haryana, Chhattisgarh, West Bengal, Telengana, Tamil Nadu	500 (one solar park in each state)
Rajasthan	680 + 1000 + 500 + 750 + 321 (5 solar parks)
Gujarat	700
Madhya Pradesh	750 + 500 + 500 + 500 + 500 (5 solar parks)
Maharashtra	500 + 500 + 500 (3 solar parks)
Karnataka	2000
Kerala	200
Himachal Pradesh, Orissa	1000 (one solar park in each state)
Uttarakhand	50
Uttar Pradesh	600
Assam	69
Nagaland	60
Meghalaya	20
Andhra Pradesh	1500 + 1000 + 1000 + 500 (4 solar parks)

emissions. For encouraging use of RES and decreasing GHG emissions, several countries set a goal and also announced polices for developing solar cities like Australia and the USA. In India, quite a lot of cities and towns are sense fast growth in the peak demand of electricity [1, 3, 5–7]. At the local level, government bodies or electricity utilities are unable to handle fast growth during peak demand of electricity, thus most of the cities or towns are facing shortage in electricity supply. MNRE has taken initiation for the advancement of cities by means of solar power with certain objective as offer a structure and assistance to make a principal plan with valuation of existing energy position, forthcoming claim and action plans. Other objective of solar cities is to supervise the execution of viable energy opportunities through public–private partnerships (PPP). MNRE has set the goal of 60 solar city projects out of which 56 solar city projects are approved for development [5, 16]. Currently, for urban areas, MNRE has commenced several programs for encouraging solar water-heating systems in offices, industry, hospitals, hotels, and hostels.

9.3 Solar Pump

In 2016, a Water Aid report categorized India as the nastiest country for most people without safe water at global level. In India, more than 76 million people are away from the safe water supply and their living standard is more serious. As per expectations of the Asian Development Bank (ADB), India will have a water shortage of 50% by 2030. However, as per expectation of Union Ministry of Water Resources (UMWS) country's present water necessities per year are estimated to be about 1100 billion cubic meters, although for the year 2025 and 2050 to be about 1200 and 1447 billion cubic meters, respectively [1, 5, 16]. With the increase in population, the claim for fresh water is also likely to increase, but the diminishing supply will make it difficult to meet the requirements. The population of India is approximately 18% of the world's population, but access to usable water sources is 4% only. As per 'Composite Water Management Index' report released by National Institution for Transforming India (NITI Ayog) in June 2018, Delhi and other 21 cities across the country will run out of groundwater by 2020 [1, 5, 16]. In urban areas, about 70.1% of the households require safe drinking water. Almost 18.7% of rural populations obtain drinking water through systematized pipe supply and others have to depend on unprocessed ground and surface water. The Indian Government has taken numerous initiatives for the irrigation and drinking water supply [1, 3–7, 16]. For irrigation and drinking water supply purpose, the Government of India has executed a system to set up 1 lakh solar pumps with funding of state nodal agencies and NABARD. The main purpose of these solar pumps would be to provide benefit to farmers for agriculture productions, revenue, and facilitating drinking water [5].

10 Other Programs/Policies/Schemes

The Indian Government has launched other useful programs, policies, and schemes based on RES such as (i) National Solar Mission for Solar projects; (ii) Wind and Solar Resources Atlas 2015 of India; (iii) Offshore Wind Energy Policy; (iv) Remote Village Electrification Program (RVEP); (v) Rural and Decentralized Renewable Power Schemes; (vi) Renewable Energy Park Scheme; (vii) Aswayuja Shops policy; (viii) Hybrid Vehicles schemes; (ix) Alternative Fuels schemes; (x) Policies on Geothermal & Tidal Energy. Details about above programs, policies, and schemes are mentioned in [16].

11 Future Potential of Renewable Energy in India

11.1 Renewable Energy Target by 2022

In India, RES is evolving to a major role in the electrical power system with and without grid. It is predictable that RES has to show an abundant greater part in reaching energy security also to be an essential fragment of the energy scheduling. In India, due to huge potential of RES, as India has reorganized a goal to an installed capacity of 175 GW by 2022 [1, 5–7, 16]. Table 21 shows that India has currently fixed a goal to an installed capacity of 175 GW from RESs by 2022, which

Table 21 Installed capacity target of from RESs by 2022 [1, 3, 5, 16]

Renewable energy sources	Installed capacity target
Solar	1,00,000 MW
Wind	60,000 MW
Biomass	10,000 MW
Small hydro	5000 MW
Total	**1,75,000 MW**

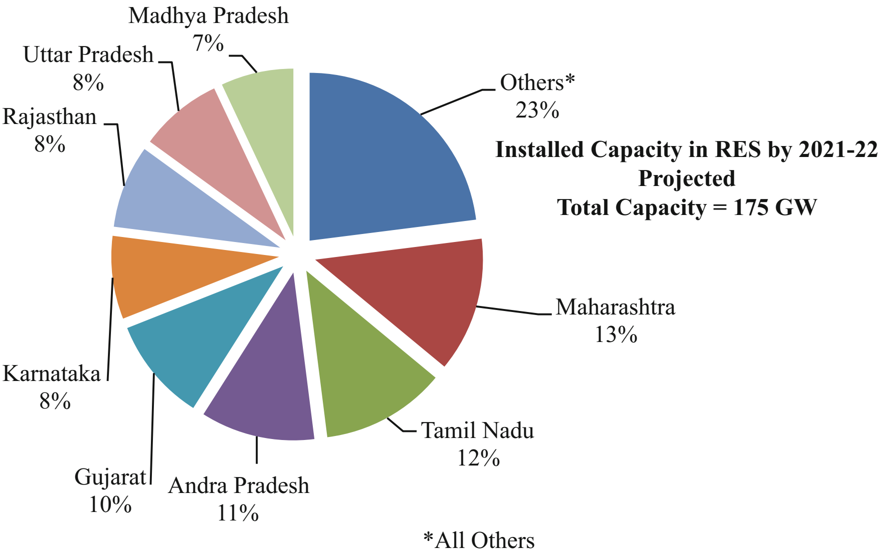


Fig. 13 Statewise target to installed renewable power of 175 GW capacities by 2022 [1, 3]

comprises 100 GW from solar, 60 GW from wind, 10 GW from biomass and 5 GW from SHP [1, 5–7, 16]. The goal of solar energy of capacity 100 GW is set under the National Solar Mission (NSM), out of which 40 GW is from solar roof tops and 60 GW from grid-connected projects of large and medium scale category. The above projects are implemented through the state Governments as well as other authorized bodies such as Solar Energy Corporation of India (SECI), Independent Power Producers (IPPs), Central Public Sector Undertakings (CPSUs), etc. [5]. The extensive higher capacity goal of renewable energy will guarantee superior energy security, upgraded energy access, and boosted employment chances.

Figure 13 exhibits the state-wise goal to fitted renewable power of 175 GW capacities by 2022 [1, 3, 5, 16]. From mentioned data, we have observed that about 13% of target will be installed in Maharashtra, after that about 12% in Tamil Nadu and about 11% in Andhra Pradesh. It also shows that these regions are most favorable for RESs. From Fig. 13, we have observed that eight states in India shall give approximately 77% of the renewable installed capacity by 2022. From Fig. 14, if we have seen region-wise, we observe that about 34% (100 GW) of total capacity will be installed in the southern region, after that, about 31% in the western region,

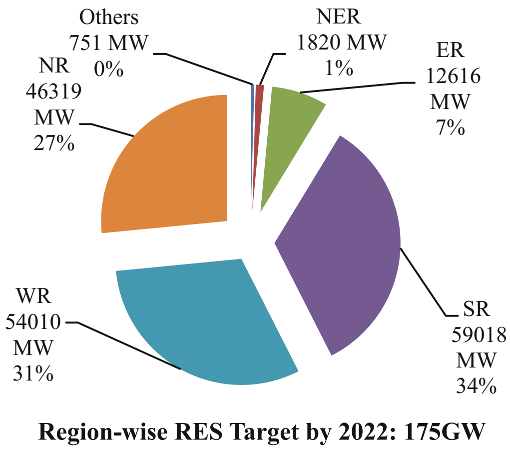


Fig. 14 Regionwise target to installed renewable power of 175 GW capacities by 2022 [1, 3]

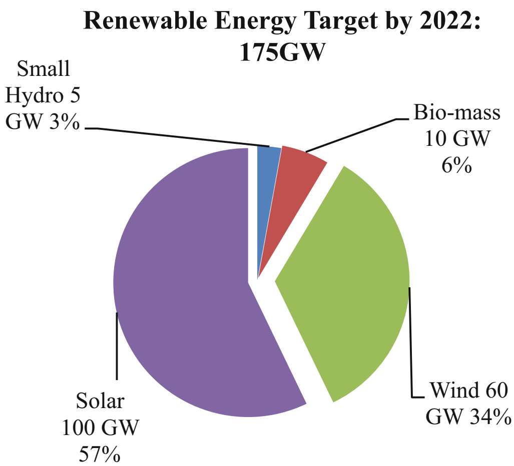


Fig. 15 Modewise target to installed renewable power of 175 GW capacities by 2022 [1, 3]

and about 27% in the northern region, while about 1% in northeastern region [1, 5, 16]. Figure 15 shows modewise goal to be fitted renewable power of 175 GW capacity by 2022 [1, 5, 16]. From this figure, we have seen that target to installed capacity from RESs are as 57% (100 GW) from solar, 34% (60 GW) from wind, 6% (10 GW) from Bio-mass, and 3% (5 GW) from SHP. From mentioned data, we observe that eight states shall give approximately 77% of entire installed capacity of renewable by 2022 [1, 5–7, 16].

Table 22 represents the year-wise goals fixed by MNRE for installation of renewables to attain the objective of 175 GW capacities by 2022 [1, 3]. From the table, we have seen that total solar target is 18,711 MW which comprises 10,311 MW from rooftop solar, 8400 MW from ground-mounted solar by 2022. The targets for other RES are 5720 MW wind power, 304 MW Biomass power, and 220 MW power from SHP by 2022.

Table 22 Targets of RES (2017–22) [1, 3]

Category	Capacity addition[a] (in MW)				
	2017–18	2018–19	2019–20	2020–21	2021–22
Rooftop solar	5000	6000	9000	9000	10,311
Ground-mounted solar	10,000	10,000	10,000	10,000	8400
Total solar	15,000	16,000	19,000	19,000	18,711
Wind	4700	5300	6000	6000	5720
Biomass	350	350	350	350	304
Small hydro power	100	100	100	100	220
TOTAL	**20,150**	**21,750**	**25,450**	**25,450**	**24,955**

[a]The capacity has been used to reach at entire RES capacity of 175 GW by 2021–22

Table 23 Projected electricity generation from RES for the years of 2021–22 and 2026–27 [1, 3, 5, 16]

Year	Installed capacity of RES (GW)	Expected generation in (BU)					Contribution of RES to total energy demand (%)
		Solar	Wind	Biomass	SHP	Total	
2021–22	175	162	112	37	15	326	20.1%
2026–27	275	243	188	63	24	518	24.4%

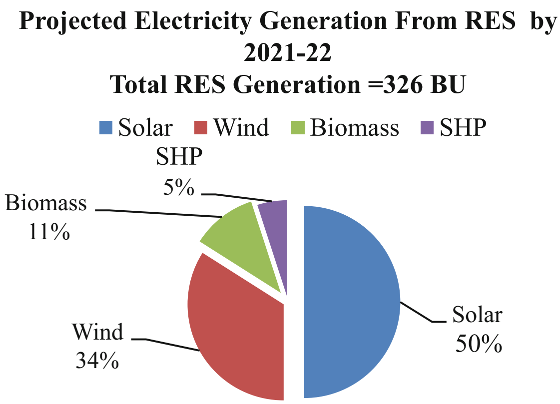


Fig. 16 Projected electricity generation from RES by 2021–22 [1, 3, 5, 16]

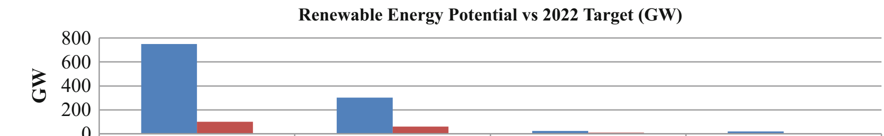
12 Projection of Renewable Energy Generation

According to MNRE, as per capacity addition targets predictions of RES by the year 2021–22 and in view of an RES capacity summation of 100,000 MW for the duration of 2022–27, predicted electricity generation from several RESs has been assessed as mentioned in Table 23 and Fig. 16 [1, 3, 5, 16]. From Table 23, we have found that participation of RES will be approximately 21% of the overall energy for the duration of 2021–22, whereas 24% by 2026–27 in India.

India has planned to achieve energy required by increasing renewable energy sources. The total domestic energy production is estimated to achieve 844 MTOE by 2021–22. This will catch about 69% of probable energy, as it is expected to be almost 375.6 MTOE by 2021–22, and the rest to be gained from imports. The Wind Energy Department is scheduled to reseat 10–15% in order to catch the power demand. With the help of weighty and significant incentives, solar power is also scheduled to catch energy capacity of 20 GW by 2022 [5]. As, by 2017–18, in India, the private sector has played a significant role in the production of electricity and the credit goes to the private sector for more than 35% of electricity production. In order to brand India as independent in electricity, a huge number of power projects based on solar are to be set up. RESs-based power projects are also to help indirectly diminish CO_2 in the atmosphere as well as decrease pollution.

12.1 Perspective Plan 2022

The Indian Planning commission has prepared an Integrated Energy Policy Report (IEPR) and point out that RES remains essential to the energy sector in India. RESs would not be out of place to note that RES may possibly be a major competitor achieving energy freedom for the lengthy time. Regarding plan 2022 as per planning commission, the Indian Government has decided viable aim for the country. Figure 17 shows the renewable energy potential vs. 2022 target [5, 6, 16]. Figure 18 exhibits all-India load duration curve for 2021–22 [1].



	Solar	Wind	Bio Energy	Small Hydro
Potential	750	302	25	20
Target	100	60	10	5

Fig. 17 Renewable energy potential vs. 2022 target (GW) [5, 6, 16]. (Source: WISE, 2017 (compiled from MNRE Annual Report 2017))

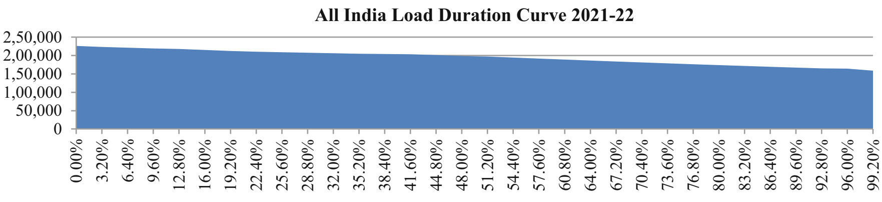


Fig. 18 All-India load duration curve for 2021–22 [1]

12.2 Perspective Plan for Grid-Interactive Renewable Power

India has also planned for Grid-interactive Renewable Power based on solar energy with the ambition to set up grid-interactive rooftop solar plant of capacity 40 GW by 2022 [5, 16]. The rooftop solar plants would be installed in several sites such as government and state-owned sector buildings, institutional, commercial, industrial, and residential buildings. The rooftop solar capacity of 656 MW was installed in India on March 31, 2017. Indian Government provided subsidies or incentives on rooftop projects that is 30% in institutional buildings, residential and social, whereas 15–25% for projects in government and state-owned enterprises. However, there are no subsidies or incentives provided in the industrial and commercial sectors for setting up of rooftop solar projects. Up to the 13th Plan, capacity of renewable power is projected to catch 54,000 MW, consisting of wind power of 40,000 MW, SHP of 6500 MW, and biomass-power of 7500 MW. According to MNRE, statewise and year-wise capacity addition target to attain the cumulative capacity of 40 GW by 2022 is presented in [5, 16].

12.3 Perspective Plan for Renewable Power for Urban, Industrial, and Commercial Applications

The main objectives comprise solar collector area of 50 million square meter for thermal uses, mostly solar water heating and from bio-source grid/captive power of 7500 MW, together with wastes of urban and industrial. As per the position in 2017, there is a gigantic capacity which needs solid attention across all. The symbolic aims for 2022 and 2032 are given in Table VII of [5, 7, 16].

12.4 Medium Term (2032) Deployment Goals

The Indian government has targets of 15% grid-connected renewable power by 2032 and substitute fuels such as hydrogen, bio fuels, and synthetic fuels replacement up to 10% by 2032 [6]. The Indian government has also made plans in various sectors by 2032 [5, 16] such as (a) energy recovery using suitable waste in 423 cities from municipal together with 107 municipal corporations; (b) solar water heating systems; (c) solar sensors-based controlled street lighting in all cities to 100% exposure; (d) energy retrieval from suitable industrial wastes across the country; (e) cogeneration with 100% recovery of potential of sugar and biomass-based industries; and (f) expansion of motive power and cooking in electrified villages. The status of renewables at global level is reported in [29, 30].

13 Conclusion

A vital study of the existing and upcoming scenario of energy resources and its potential in India has been highlighted. Several energy resources in widespread diversity and different forms of RES are obtainable in India. In this paper, the installed capacity of conventional and non-conventional power and future plans for RES as per MNRE is also described in detail. Discussions also reflect the huge potential of RESs for electricity generation in different regions in India. As per GOI plans, policies, and programs on utilization of RES to achieve the increasing energy requirements, fulfilling of energy scarcity and energy security has been the focus in this paper. As per MNRE dreams, the present status and future demand of electricity reflect that although right now the involvement of RES is less but upcoming growth will influence brand RES technique more modern to shift conventional sources of energy. The approach for realizing superior aims regarding huge application of RES will mostly be influenced by the active contribution of all performers such as government agencies, NGOs, manufacturers, research and development institutions, financial institutions, developers among the new breed of energy tycoons. It is imagined that this paper will assist as an appreciated resource to any further work carried out in this important area of research.

14 Future Work

This paper presents a wide range of renewable energy potential and policy in India through which researchers can compare the potential and policy with other countries. Future work shall focus on the comparison of data of renewable energy potential of European countries.

References

1. National Electricity Plan Volume-I (Generation), Government of India Ministry of Power Central Electricity Authority January, 2018
2. Variable renewable energy, https://en.wikipedia.org/wiki/Variable-_Renewable-Energy
3. Executive Summary on Power Sector, Government of India Ministry of Power Central Electricity Authority, Jan. 2017, March-2018, September-18
4. Energy statistics report, Govt. of India; March, 2018
5. Ministry of New and Renewable Energy, Government of India. Available: http://www.mnes.nic.in/
6. Indian Energy Portal, [Online] https://www.re3data.org/repository/r3d100011024
7. Ministry of Power, Government of India. [Online] Available: http://powerminp.nic.in/
8. International Atomic Energy Agency (IAEA). Available: https://www.iaea.org/about
9. G. Sandeep, S.A. Raghubir, “Development of Solar Electricity Supply System in India: An Overview”, Hindawi Publishing Corporation. J. Solar Energy (2013). https://doi.org/10.1155/2013/632364

10. Wind Power Development in India: An Overview. Available: https://mnre.gov.in/file-manager/akshay-urja/october-2017/Images/20-25.pdf
11. Centre for Wind Energy Technology. [Online] Available: http://www.cwet.tn.nic.in/default.htm
12. L.G. Lee, Wind energy developments: Incentives in selected countries, Available: http://tonto.eia.doe.gov/ftproot/features/wind.pdf
13. Indian Renewable Energy Development Agency (IREDA). [Online] Available: http://www.ireda.in
14. Suzlon Energy. [Online] Available: https://www.suzlon.com/
15. Indian Wind Energy Association (IWEA). [Online] Available: http://www.inwea.org/index.htm
16. S.N. Singh, B. Singh, O. Jacob, Renewable Energy Generation in India: Present Scenario and Future Prospects. (2009). https://doi.org/10.1109/PES.2009.5275448, Published in IEEE PES General Meeting, 2009
17. Review of Performance of Hydro Power Stations 2017–18, Hydro Project Planning & Investigation Division, Central Electricity Authority
18. TERI Report to CERC on "Pricing of power from Non-Conventional Sources" [Online] Available: http://cercind.gov.in/index.html
19. Power Scenario in India 2018 – State Wise Track. Available: https://enincon.com/report/power-scenario-in-india-2018
20. Annual report 2017–18, reported by World Institute of Sustainable Energy, Pune
21. The Electricity Act 2003
22. Report on Short-term Power Market in India: 2017–18, Economics Division Central Electricity Regulatory Commission. Available: www.cercind.gov.in/2018/MMC/AR18.pdf
23. Annual report 2018-Central Electricity Authority. Available: http://www.cea.nic.in/reports/annual/annualreports/annual-report-2018.pdf
24. A consumer's guide to electricity tariff determination. Available: https://www.karnataka.gov.in/kerc/.../Electricity-Tariff-Determination.pdf
25. Power Distribution Tariffs in India 2018
26. India's Electricity Sector Transformation report of Institute of Energy Economics and Financial Analysis November, 2017, IEFE.org
27. Load Generation Balance Report 2017–18. Available: www.cea.nic.in/reports/annual/lgbr/lgbr-2018.pdf
28. Report of the Working Group on, "NEW AND RENEWABLE ENERGY" for XIth Five Year Plan (2007–12), Planning commission, Government of India, 2006, Available: http://planningcommission.nic.in/aboutus/committee/wrkgrp11/wg11_renewable.pdf
29. Renewables 2018 Global Status Report, REN21. Available: https://www.ren21.net
30. World Energy Council Annual Report 2017. Available: https://www.worldenergy.org/wp-content/uploads/2018/09/WEC-2017-Signed.pdf

Analysis of Brain Signals to Forecast Motor Intentions Using Artificial Intelligence

Nabil I. Ajali and Carlos M. Travieso

1 Introduction

In the twenty-first century, more than 120 million people around the world have some kind of motor problems due to an accident or mental illness.

As a result, amputations, atrophy, or some loss of mobility in parts of the body is common. This hinders integration into society and worsens the quality of life [1–4].

Brain Computer Systems (BCIs) use technology developed in Industry 4.0 to establish a direct communication pathway between the brain and an external device using brain electrical signals. BCIs are often focused at researching, mapping, assisting, augmenting, or repairing human cognitive or sensory-motor functions [5].

Hence, these devices are a potential solution for motor problems, when the brain is not involved directly. Signals could be analyzed and, using the machine and deep learning algorithms, a human–computer interaction is established to interpret and allocate this signal in a robotic limb or in an exoskeleton to form a complete BCI system.

In this work, a series of concepts necessary to understand the nature of the brain signal and the methods used to develop a BCI is shown. In addition, a classification of a public dataset is carried out with our system. This allows to obtain conclusions and paradigms of this application for Industry 4.0. and the inclusion in society 5.0.

N. I. Ajali (✉) · C. M. Travieso
Signals and Communications department, Universidad de Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain
e-mail: nabil.ajali101@alu.ulpgc; escarlos.travieso@ulpgc.es

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_2

1.1 State of the Art

The brain is the most important organ in the human body. As the processors in computers, it controls most of the body's activities. It processes, integrates, and coordinates the information received from the organs and the senses and sends instructions to the rest of the body. Its basic units, neurons, work by electrochemical impulses called synapses [6]. The electrical nature of the brain was discovered by Richard Caton in 1875 and the first electroencephalogram was done in 1924 by Hans Berger [7, 8].

In the years after the Second World War until 1970, advances in the field of technology promoted the society to industry 3.0. In 1970, machine–brain concept was developed by Professor Jacques Vidal. In 1988, the first non-invasive EEG control of an object was done, specifically a robot was controlled using the EEG. The authors of this work were Stevo Bozinovski, Mihail Sestakov, and Liljana Bozinovska, pioneers in the first robot control using the EEG [9, 10].

Consequently, followed by these works of BCI systems, different concepts of neurobiology such as brain neuroplasticity were discovered in this field. A range of possible applications were opened up, e.g., the implantation of electrodes directly in the cerebral cortex to handle with the brain as natural sensors or effector channels [11]. After years of experimentation with animals, the first neuroprosthetic devices implanted in humans emerged in the mid-1990s.

In the first decade of the twenty-first century, the concept of P300 or evoked potential was developed. P300 is a peak in brain response as a consequence of a sensory, cognitive, or motor stimulus [12]. This stimulus, in addition to research for medical purposes, was also focused on the video game industry and the first non-invasive brain–computer interfaces were introduced on the market.

Different works in the 2010s suggested that BCIs could stimulate the neurons and the associated behaviors through modulation of the molecular mechanisms of synaptic efficiency. Consequently, functional connection in the brain could be restored [13, 14].

BRAIN Initiative (Brain Research through Advancing Innovative Neurotechnologies) was created in 2013 by Defense Advanced Research Projects Agency of the USA to understand and map the brain activity. This initiative, has supported the work of the University of Pittsburgh Medical Center [15], Paradromics [16], Brown [17], and Synchron, among others [18].

In the decade of 2010s to 2020s, several studies have led to great advancement in human–computer interaction for pattern recognition, task classification, and diagnosis of diseases such as Parkinson's, epilepsy, or Alzheimer's. This is due to the computation and algorithms improvement. This is due to the development of Machine Learning models and Deep Learning models. These algorithms extract relevant features of brain signals from the data to classify or recognize certain patterns [19–22].

High levels of success in classifying different mental states were achieved in these years, e.g., the one obtained in 2018 by Bird, J. J et al. in prediction of mental states of relaxation, neutrality, and concentration [23]. Also, in emotional mental states such as being negative, neutral, and positive [24]. And in the theoretical framework of thalamocortical arrhythmia, allowing the understanding and detection of mental diseases such as Parkinson's or Alzheimer's [25].

On the other hand, in those years, large amounts of studies were carried out in parallel in other lines whose efforts were focused on using this AI applied to the EEG. For example, Feng, J. K et al. published, in 2019, an optimized channel selection method based on multi-frequency CSP-Rank for BCI systems using motor images [26]. This, enabled the creation of brain-machine systems capable of acting under the orders of thought. Other works propose an upper limb to assist in the rehabilitation of people with cerebrovascular accident or people with some disability or amputation, e.g., the work of Jiménez, A. R et al. [27]. Or even applications for the army, such as that obtained by the Indian exercise in the study "Brain Controlled Robot for Indian Armed Force. 2016" [28].

With the rise of intelligent applications and industry 4.0 paradigms, society is evolving towards what is known as society 5.0, where the human and the machine are increasingly interconnected and, in turn, with the Internet. In 2020, the famous philanthropist Elon Musk and his company *Neuralink* created an implantable BCI system to control systems with the mind. In 2021, a chip was successfully implanted in a pig and in a monkey. The second achieved to play video games using this device [29]. The negative note is that these BCI devices are invasive and are still in the early stages of development.

The key to the success of BCI is to define and choose the best option, depending on the nature of the problem. To recognize patterns and distinguish between different states, SVM (Support Vector Machine) algorithms, KNN (k-neighborhood) algorithms, or similar are usually used. However, the use of neural networks and combinations of various deep learning methods is becoming more and more common, since they are capable of solving a large number of problems of various kinds. The advantage of these is the automation and success of the learning process. But, on the other hand, they are too expensive in terms of computation and in the large amount of data that is required.

Furthermore, it is not only the selection of algorithms that matters. As proposed by Roc A. et al. in [30], the mental state and the type of users are relevant factors. Hence, the result of a classification differs a lot between users. This is due to the predisposition of the subjects, there are subjects who have a better response to these BCI systems. As a solution, prior training to the subjects is proposed, by Roc A. et al. to improve the classification of this type of task.

1.2 Our Proposal

A huge number of avant-garde articles has been checked to extract the most important concepts and ideas in BCI field in order to develop and complete this

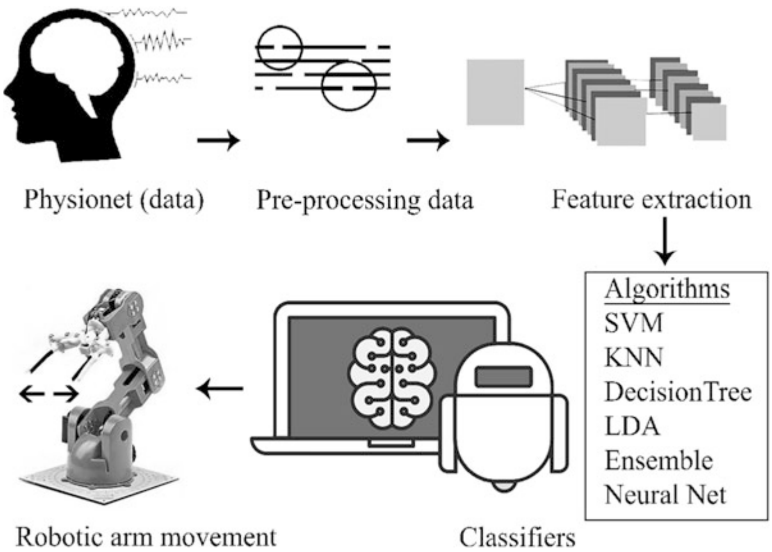


Fig. 1 The complete process is shown from the acquisition of data in the Physionet database to the movement of the robotic arm after processing and classification

work. Thus, these previous works contribute to propose a methodology in BCI field (Fig. 1).

Normally, the whole systems have work using a common scheme, as is shown in Fig. 1, which has the following steps:

1. Acquire the signal directly or in an EEG database
2. Make a Pre-process of the signal in order to increase the signal-to-noise ratio and eliminate redundant data
3. Feature extraction of the signal with the correct technique
4. Select the classification method and train the system
5. Test the system to verify its robustness and success

In this work, these steps are generally followed, but a new different feature extraction is made up. After acquiring the signal and making the pre-process, feature extraction is done taking into account certain aspects of the subjects. With this, the system is capable of making classifications by adjusting the results considering brain activity levels.

To verify the feasibility of the proposed method, a comparison with a previous work, where multiple classifiers are used, is done. Some of these studies have peaks of 75.0% in their classification rate, e.g., such as that of Joadder M. et al. [31] where the BCI algorithm is obtained using the Physionet database. Other studies also fall in their accuracy using Machine Learning, e.g., the proposal by Abbas W et al. with 61.0% success [32], Sakhavi S et al. with 74.0% success [33], and others that can be verified in the review by Craik, A et al. of the prestigious Journal of Neural Engineering [34].

2 Material and Methods

2.1 Database

In this work, different data sets were obtained from a public database on the PhysioNet website in collaboration with the developers of the BCI2000 instrumentation system [35–37]. The data collected more than 1500 EEG recordings with a 160 Hz sample and approximately 2 min per record, obtained from 109 healthy subjects. EEG signals were recorded using 64 channels and a 10/20 system.

In this dataset, these subjects have performed a series of motor and imaginary tasks. They must move the fists according to a signal on a screen for 4.1 s and then rest for another 4.1 s alternately (Fig. 2).

As the aim is to detect motor intention, to optimize time in the demonstration of the hypothesis, a simplification was carried out. Ten patients were selected, as a representative group, from among the 109 subjects.

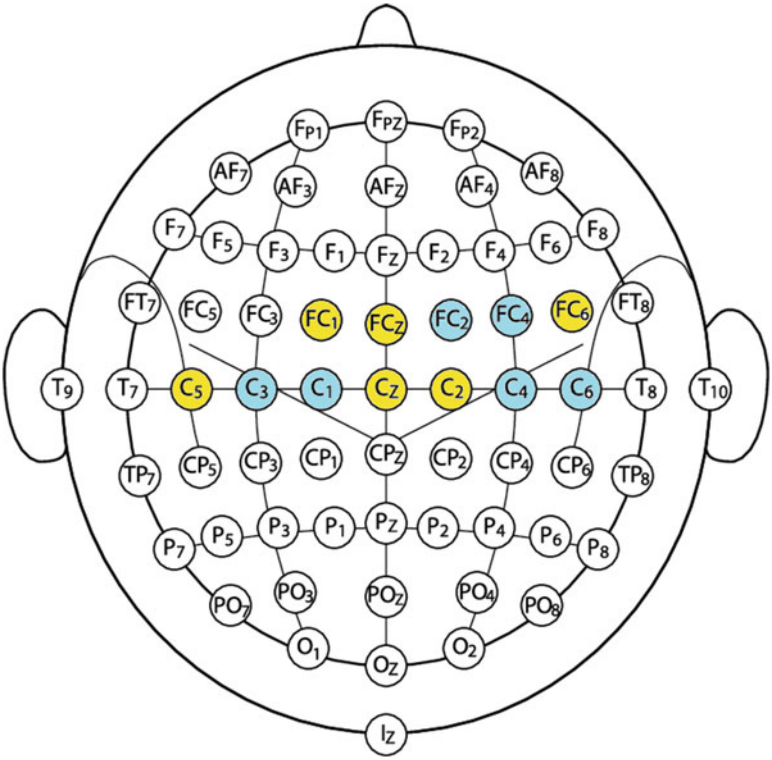


Fig. 2 EEG channels that contain hand information in real and imaginary actions. FC1, FCz, FC6, C5, Cz, C2. Six extra channels added in the frontal lobe (primary motor cortex) are C3, C1, FC2, FC4, C4, and C6

The experiments performed contain 14 tests:

- *Test 1 and 2:* Monitoring signal for 1 min with eyes open and eyes closed, respectively (baseline).
- *Test 3 to 14:* Monitoring for 2 min while the subjects perform different types of tasks related to motor movements. Each subject gets 12 tasks based on real and imaginary movements.

  Among the 12 tasks, there are four types of experiments. In this work, we focus only on two types of these tasks, as we believe that they are sufficient to develop and implement the desired system. These two types are:
- *Test 3 to 5:* A loop appears with an object to the left or right on a screen, the subject is monitored for 2 min while opening and closing the fists on the corresponding side for approximately 4 s. When the object disappears, the subject relaxes (three tests).

According to Craik, A et al. [34], the EEG channels that contain hand information in real or imaginary actions are FC1, FCz, FC6, C5, Cz, and C2. After several tests, six channels were added in the frontal lobe (primary motor cortex), up to 12, as we see that the accuracy improves. These channels are C3, C1, FC2, FC4, C4, and C6. Figure 2 shows the final arrangement chosen.

Finally, the nomenclature followed to know if the right or left hand is used goes in annotations every 4.1 s:

- *T0:* Rest (without movement information or baseline information)
- *T1:* Start of movement (real or imaginary) of the left fist
- *T2:* Start of movement (real or imaginary) of the right fist

2.2 Pre-processing

In healthy subjects under stress-free conditions, brain waves range from about 1 to 40–50 Hz (delta to beta waves). Factors such as measurement or electrical noise negatively affect the results in certain frequency ranges by masking the signal. To avoid this external noise produced by low frequencies (physical or physiological movements such as breathing or turning the head) and high frequencies (electrical noise), a band-pass filter is used as preprocessing [38–40]. By using this filter between 0.5 and 50 Hz (Butterworth type), external frequencies or noise are eliminated and a flatter signal is obtained, improving the signal-to-noise ratio.

2.3 Feature Extraction

To extract the main features of the signal, the algorithm developed used the discrete wavelet transform. The discrete wavelet transforms or DWT is a signal processing

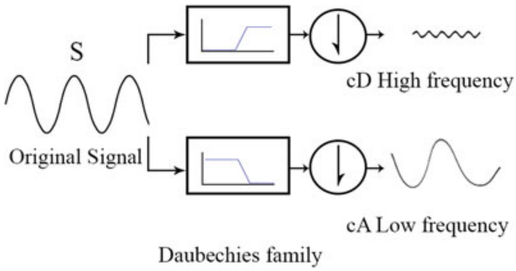


Fig. 3 After the application of the DWT, the original signal is divided into a high frequency and a low frequency. This is only 1 level of DWT

tool based on multi-resolution analysis with variable length windows. It is used to obtain a representation, decomposition, and reconstruction of certain signals. DWTs are especially useful with stochastic signals that present large changes in their time-frequency components instantly [41].

Brain signals have unpredictable frequency and intensity over time, so they are non-stationary. The DWT is capable of breaking a signal with low-pass and high-pass filters at different levels (Fig. 3). Thus, obtaining a high frequency component and a low frequency component (with different information) at various levels. These levels correspond to different types of brain waves (delta to beta) that provide different types of information.

Mathematically, the equation behind the DWT is given by [42]:

$$f(t) = \sum_{k} \sum_{j} a_{j,k}\, \varphi_{j,k}\,(t)\ . \tag{1}$$

This equation is expressed in terms of two indices, the translation time k and the scaling index j. These two indices are integer's values and the wavelet functions form an orthogonal set of functions (base) [43].

In this work, the family of the wavelet coiflet and the Daubechies are used, among the different DWT families (Haar, Symmlet, Coiflets). This is due to they have a better performance at the time of classification according to studies [35, 36].

The different coiflet and Daubechies were calculated and the level of decomposition was established in four levels of detail in order to obtain the different isolated brainwaves. After several tests, Daubechies show better performance as is shown by Alomari et al. [44], hence, after this point, only Daubechies wavelet is covered in this paper.

The equations to obtain the wavelets, formulated by Daubechies are defined below [45]:

- Orthogonal in L_2
- Tabulated coefficients
- Zero moments of father-function
- Zero moments of wavelet function

$$M_i = \int x^i \varphi(x)\mathrm{d}x = 0\,. \tag{2}$$

$$\mu_i = \int x^i \psi(x)\mathrm{d}x = 0. \tag{3}$$

2.4 Classifiers

After feature extraction, a set of classification tests is performed using different classifiers to make a comparison and find the optimal one. These classifiers are:

- Decision trees
- Linear Discriminant Analysis (LDA)
- Logistics Analysis
- Support Vector Machine
- K-Nearest Neighbors
- Ensemble methods
- Neural networks

2.4.1 Decision Trees

Decision tree algorithm is a predictive modeling approach used in machine learning, economy, statistics, and data mining. A decision tree is used as a predictive model in which a set of data is given and logical construction diagrams are manufactured. Very similar to rule-based prediction systems, which serve to represent and categorize a series of conditions that occur in succession, to solve a problem [46].

There are many types of decision trees such as; "random forest," bootstrap algorithms, decision trees, powered trees, regression trees, etc. The algorithms for the construction of decision trees usually work from the top down, choosing in each step the variable that best divides the set of elements. The mathematical theories behind these usually use entropy, Gini impurity, variance reduction, or information gain to obtain the results [47].

$$\text{Entropy} = -\sum_{i=1}^{n} p_i * \log\left(p_i\right). \tag{4}$$

$$\text{Gini index} = 1 - \sum_{i=1}^{n} {p_i}^2.$$

$$i \text{ is number of class} \tag{5}$$

2.4.2 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a generalization of Fisher's linear discriminant (technique for dimensionality reduction). This is a method used in machine learning, pattern recognition, and statistics to find a linear combination of features that allows to characterize and separate two or more classes of objects or events. The resulting combination can be used as a linear classifier or for dimensionality reduction before subsequent classification [48].

2.4.3 Logistic Analysis

Logistic model (or logit model) is used to model the probability of a certain class or event existing. Is a binary probability such as win/lose, white/black or healthy/sick. Nowadays, it can be extended to model several classes of events such as determining whether an image contains a car, a bus, or people, etc. This is used in various fields, including most medical fields, social sciences, and machine learning, where it is included in neurons of algorithms like perceptron [49].

The mathematical expression is:

$$\operatorname{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{1,i} \ldots \beta_k x_{k,i}. \tag{6}$$

β_i are parameters of the model

x_i are redictors

2.4.4 Support Vector Machine

Support vector machines are supervised learning models used in Machine Learning to address a classification or regression analysis.

When a set is provided to train the model with samples, a part of the samples is labelled to train the SVM to build a model that predicts the class of new samples. A support vector machine is a model that represents the sample points in space, separating the classes into two or more spaces as wide as possible. The model is carried out by means of a separation hyperplane defined as the vector between the two points, of the two classes, closest to what is called the support vector. SVM is present in many applications, such as in neurons that form a perceptron [50].

A hyperplane of an SVM can have different shapes, depending on its design, ranging from the simplest such as a 2D hyperplane to polynomial, irregular, or circular shapes.

2.4.5 K-Nearest Neighbors

The k nearest neighbors (K-NN) algorithm is a non-parametric classification method used for classification and regression in statistics and machine learning. K-NN finds the distances between a new sample and all the examples in the data. Selecting the examples of specified numbers (K) closest to the new sample, then choose the most frequent label (in the case of classification) or average the labels (in the case of regression) [51].

The distance functions for operating with K-NN could change depending on whether the samples are continuous variables or not. In the case of continuous variables, Euclidean or Manhattan distance functions could be used, and the Hamming distance should be used for categorical variables [52]. The equations are shown below:

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2} \,. \tag{7}$$

$$\text{Manhattan distance} = \sum_{i=1}^{k} |x_i - y_i| \,. \tag{8}$$

$$\text{Hamming distance} = D_{\text{H}} = \sum_{i=1}^{k} |x_i - y_i|$$

$$x = y \Rightarrow D = 0, x \neq y \Rightarrow D = 1 \tag{9}$$

2.4.6 Ensemble Methods

Ensemble methods are techniques that create multiple models and then combine them to produce improved results. Usually produces more accurate solutions than a single model would. In this work, Bag, Gentle-Boost, Logit-boost, Ada-boost, and RUS-boost methods are combined to select the best option. The final prediction is given by a weighted combination of all weak classifiers that have occurred previously [53]. General expression for Boosting algorithms is:

$$G(x) = \text{sign}\left(\sum_{m=1}^{M} \alpha_m G_m(x)\right). \tag{10}$$

Where $G(x)$, $m = 1,2 \ldots, M$ are the sequence of weak classifiers and $\alpha_1, \alpha_2 \ldots, \alpha_M$ are the weights of each classifier.

2.4.7 Neural Networks

Neural networks (NN) are a computational model based on biological neural networks. Similar to the biological system, NNs are made up of neurons, which are mathematical decision functions. Its structure is composed of these neurons, arranged in an input layer, one or more hidden layers, and an output layer.

Among many functions, NN allows data to be classified after supervised (or unsupervised) learning in which an algorithm "learns" to classify new observations based on labeled data examples.

In this study, the network architecture was performed using "Neural Net Fitting" included in Matlab. After the data selection process and the feature extraction that was made previously in the loop, the matrices that contain the information are now the inputs in the neural network. The number of inputs being greater than 200. The system was trained with the Levenberg–Marquardt backpropagation model, Eq. (11) [54]. It was validated in two ways, using the hold out cross validation and the k-fold 10 validation. The number of hidden layers was set to 1 and the neurons in the hidden layer were changing to find the optimal relationship between computational time and success.

$$\Delta x_k = -\left[\mathrm{J}^{\mathrm{T}}(x_K)\,\mathrm{J}(x_k) + \mu_k I\right]^{-1}\mathrm{J}^{\mathrm{T}}(x_k)\,\mathrm{v}(x_k)\,. \tag{11}$$

3 Methodology

This section explains the methodology to obtain the main characteristics used in the classification.

3.1 Feature Extraction Algorithm

In several steps, the algorithm performs a series of transformations to form a data matrix starting from EDF+ format (European Data Format). This matrix contains the motor action or motor intentions data (T0-T1-T2).

Then redundant channels are removed and rows and columns containing information of the experiment are separated in two matrices. One contains the task events T1-T2 (left/right fist) and the other matrix contains the rest intervals (T0). Channels that do not provide relevant information are discarded. The annotations are included in another attached matrix.

The second step is pre-processing with a band pass filter that is applied to optimize the noise-signal ratio.

In the third step, the data matrices go through a loop, which calculates the wavelet transform (Daubechies) in four levels. This parameter is used to create a matrix. This matrix is the input to the neural network, and it is compound by 450 columns of T1/T2 events (left/right hand), 8184 rows per column (corresponding to different levels of wavelet of the previous step) and one extra row for the labels of the events.

4 Results and Discussion

The hardware used to obtain the results is a computer with an I7-9600-K processor, 16 GB of RAM memory and a Gigabyte GeForce RTX 2060 Windforce OC 6GB GDDR6 graphic card of 7.03 TFLOPS. The computation time ranges from 12 to 6 ms, depending on the classifier used.

4.1 Results

In this section, the performance of the classifications for each user is shown. The results of tasks classifications using machine learning algorithms are shown in Table 1. The best results are painted in green tone and pointed with an arrow, reaching peaks of 93% of accuracy. In the bottom, the average accuracy and standard deviation per classifier and user is shown.

4.2 Discussion

In this section, the results shown above are discussed in order to obtain the most important keys of this work:

Accuracy in the classification is at least 80.0% for 8 of 10 users in one or more of their experiments. Coming to a 93.3% of success in some experiments. These results are higher than the results obtained by other research groups in different types of classifications with mental tasks where a 61.0–75.0% is reached [55–59]. K-NN and Ensemble are the most accurate methods. Unusually, for two users, the accuracy in their classifications were lower. At least 73.3%, as is shown in Table 1. The possible reasons, as we mentioned above in the work of Roc A. et al. [30], are the predisposition of the subjects and the mental state of these.

D Zhang et al. [60] use the same database, in which they assured to obtain 93.0% success in the classification using another methodology based not on Machine, but on Deep Learning. They take into account the spatial and temporal information and apply deep learning algorithms either cascading or parallel. A reason for this good accuracy is the use, by D Zhang et al., of the whole database and, in consequence, the more data to train the neural networks. In addition, the clustering events in pairs of T1-T0 vs. T0-T2 and vs. T1-T2 which could result in a masked result.

We have not found studies on the internet and science databases that obtain similar levels of success rates for binary classification using the same conditions

Table 1 Classification using different machine learning algorithms. There are ten users and three experiments per user

Users	Tree	LDA	Logistic	SVM	KNN	Ensemble	NN (bilayered)
S001_exp1_1	33,3	40,0	40,0	60,0	60,0	46,7	60,0
S001_exp1_2	53,3	40,0	40,0	53,3	80,0	66,7	26,7
S001_exp1_3	66,7	53,3	73,3	73,3	73,3	86,7	46,7
Average_S001	51,1	44,4	51,1	62,2	71,1	66,7	44,5
S002_exp1_1	66,7	66,7	73,3	80,0	80,0	73,3	60,0
S002_exp1_2	80,0	66,7	46,7	66,7	80,0	80,0	66,7
S002_exp1_3	53,3	33,3	60,0	53,3	86,7	60,0	53,3
Average_S002	66,7	55,6	60,0	66,7	82,2	71,1	60,0
S003_exp1_1	20,0	40,0	53,3	60,0	53,3	53,3	26,7
S003_exp1_2	26,0	60,0	40,0	73,3	73,3	66,7	73,3
S003_exp1_3	33,3	60,0	53,3	53,3	66,7	60,0	53,3
Average_S003	26,4	53,3	48,9	62,2	64,4	60,0	51,1
S004_exp1_1	26,7	33,3	93,3	53,3	53,3	33,3	53,3
S004_exp1_2	66,7	73,3	66,7	73,3	80,0	73,3	73,3
S004_exp1_3	40,0	53,3	53,3	60,0	73,3	60,0	60,0
Average_S004	44,5	53,3	71,1	62,2	68,9	55,5	62,2
S005_exp1_1	73,3	60,0	60,0	93,3	80,0	73,3	26,7
S005_exp1_2	26,7	26,7	46,7	60,0	60,0	53,3	46,7
S005_exp1_3	53,3	60,0	53,3	73,3	73,3	60,0	46,7
Average_S005	51,1	48,9	53,3	75,5	71,1	62,2	40,0
S006_exp1_1	26,7	53,3	46,7	53,3	46,7	46,7	53,3
S006_exp1_2	53,3	40,0	86,7	53,3	80,0	66,7	66,7
S006_exp1_3	60,0	46,7	20,0	53,3	66,7	66,7	33,3
Average_S006	46,7	46,7	51,1	53,3	64,5	60,0	51,1
S007_exp1_1	40,0	53,3	60,0	60,0	73,3	73,3	73,3
S007_exp1_2	73,3	66,7	60,0	80,0	86,7	73,3	60,0
S007_exp1_3	40,0	33,3	40,0	53,3	60,0	53,3	26,7
Average_S007	51,1	51,1	53,3	64,4	73,3	66,6	53,3
S008_exp1_1	80,0	60,0	80,0	73,3	93,3	80,0	46,7
S008_exp1_2	13,3	46,7	86,7	66,7	53,3	53,3	60,0
S008_exp1_3	33,3	46,7	20,0	53,3	73,3	26,7	66,7
Average_S008	42,2	51,1	62,2	64,4	73,3	53,3	57,8
S009_exp1_1	66,7	46,7	66,7	53,3	66,7	73,3	53,3
S009_exp1_2	60,0	46,7	53,3	53,3	60,0	66,7	26,7
S009_exp1_3	73,3	46,7	40,0	53,3	73,3	73,3	73,3
Average_S009	66,7	46,7	53,3	53,3	66,7	71,1	51,1
S010_exp1_1	86,7	53,3	40,0	73,3	60,0	86,7	33,3
S010_exp1_2	40,0	33,3	46,7	53,3	53,3	53,3	33,3
S010_exp1_3	40,0	26,7	40,0	53,3	53,3	46,7	33,3
Average_S010	55,6	37,8	42,2	60,0	55,5	62,2	33,3
Total_Average	51,1	48,6	53,8	62,2	68,6	63,9	49,6
STD	11,2	5,0	7,6	6,1	6,7	5,7	8,6

(Physionet dataset, Machine Learning algorithms, and number of users). This success refers to the maximum values of successes and not to the mean of the experiments.

Based on the results, we can conclude that accuracy achieved in most cases arises from repeating the experiment three times per user. We support that there exists a close, directly proportional relationship, between the training given to the user and the results obtained as suggested in A Roc et al. [30]. Hence, it would be interesting for future researches, that subjects must undergo training in the regulation of brain waves in order to optimize the accuracy in classification.

It is interesting to note the poor performance of neural networks in our case. Probably, due to the large amount of data needed to train them and the reduced sample that has been taken. It would be interesting to analyze the behavior of these neural networks with the complete set of the database as we mentioned above.

5 Conclusions

In this chapter, concepts developed in industry 4.0 have been used to create and evaluate the BCI systems' ideas. This seeks to develop innovative tools that improve the lifestyle of people with motor problems. As well as the integration of technology for a 5.0 society.

An analysis of the state of the art of BCI systems and motor control by brain signals has been made. A database containing experiments of hundreds of subjects performing motor and mental tasks (Physionet) has been obtained. The methodology to develop a BCI system has been tested. Classification of motor control tasks has been carried out with more than six types of machine learning classifiers and neural networks. The results show success rates higher than 80.0% in most cases and even cases with 93.3%, except for two users. This is possibly due to the minimal practice they have received with the tasks. Another reason is that the mental state prior to performing each task has not been taken into account, as in all studies in state of the art, and we think that is very important. We hope that with this chapter we have contributed to a better understanding of these BCI systems.

Acknowledgments This study was supported by the canary government, which promotes the investigation of the PhD students thanks to its grant program.

References

1. R. Zijdeman, F.R. da Silva, Life expectancy at birth (Total) http://hdl.handle.net/10622/LKYT53, IISH Dataverse. V1 (2015)
2. J. Bickenbach, The world report on disability. Disabil. Soc. **26**(5), 655–658 (2011)
3. National Highway Traffic Safety Administration (NHTSA). (2021). No Title. https://injuryfacts.nsc.org/motor-vehicle/overview/introduction/
4. International Labour Organization. (2021). No Title. https://www.ilo.org/moscow/areas-of-work/occupational-safety-and-health/WCMS_249278/lang--en/index.html The ILO estimates that some, of work-related illnesses annually.

5. M.O. Krucoff, S. Rahimpour, M.W. Slutzky, V.R. Edgerton, D.A. Turner, Enhancing nervous system recovery through neurobiologics, neural interface training, and neurorehabilitation. Front. Neurosci. **10**, 584 (2016)
6. CogniFit. (marzo de 2019). CogniFit.com. Obtenido de https://www.cognifit.com/es/cerebro
7. Cohen of Birkenhead, Richard Caton (1842–1926) Pioneer electrophysiologist. Proc. R. Soc. Med. **52**(8), 645 (1959)
8. L.F. Haas, Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography. J. Neurol. Neurosurg. Psychiatry **74**(1), 9–9 (2003)
9. M. Lebedev, Augmentation of sensorimotor functions with neural prostheses. Opera. Med. Physiol. **2**(3–4), 211–227 (2016)
10. M.A. Lebedev, M.A. Nicolelis, Brain-machine interfaces: From basic science to neuroprostheses and neurorehabilitation. Physiol. Rev. **97**(2), 767–837 (2017)
11. S.P. Levine, J.E. Huggins, S.L. BeMent, R.K. Kushwaha, L.A. Schuh, M.M. Rohde, et al., A direct brain interface based on event-related potentials. IEEE Trans. Rehabil. Eng. **8**(2), 180–185 (2000)
12. S.J. Luck. Introduction to the Event-Related Potential Technique (1.a ed.). Bradford Books (2005)
13. R.A. Miranda, W.D. Casebeer, A.M. Hein, J.W. Judy, E.P. Krotkov, T.L. Laabs, et al., DARPA-funded efforts in the development of novel brain–computer interface technologies. J. Neurosci. Methods **244**, 52–67 (2015)
14. M. Jacobs, A. Premji, A.J. Nelson, Plasticity-inducing TMS protocols to investigate somatosensory control of hand function. Neural Plast. **2012** (2012)
15. M. Fox, Brain chip helps paralyzed man feel his fingers. NBC News. (2020, 25 Nov.) https://www.nbcnews.com/health/health-news/brain-chip-helps-paralyzed-man-feel-his-fingers-n665881
16. DARPA, Awards $65 million to develop the perfect, tiny two-way brain-computer interface. Dent. Tech. (2017) https://techcrunch.com/2017/07/10/darpa-nesd-grants-paradromics/
17. Brown to receive up to $19M to engineer next-generation brain-computer interface. (2017, 10 jul.). Brown University. https://www.brown.edu/news/2017-07-10/neurograins
18. Minimally Invasive "Stentrode" Shows Potential as Neural Interface for Brain. (2021). DARPA. https://www.darpa.mil/
19. D. Gajic, Z. Djurovic, S. Di Gennaro, F. Gustafsson, Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition. Biomed. Eng.: Appl. Basis Commun. **26**(02), 1450021 (2014)
20. S.L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan, U.R. Acharya. A deep learning approach for Parkinson's disease diagnosis from EEG signals. Neural Comput. & Applic. **32**(15), 10927–10933 (2018). https://doi.org/10.1007/s00521-018-3689-5
21. M. Fan, A.C. Yang, J.L. Fuh, C.A. Chou, Topological pattern recognition of severe Alzheimer's disease via regularized supervised learning of EEG complexity. Front. Neurosci. **12**, 685 (2018)
22. S. Gurumoorthy, N.B. Muppalaneni, X.Z. Gao, Analysis of EEG to find Alzheimer's disease using intelligent techniques, in *Computational Intelligence Techniques in Diagnosis of Brain Diseases*, (Springer, Singapore, 2018), pp. 61–70
23. J.J. Bird, L.J. Manso, E.P. Ribeiro, A. Ekart, D.R. Faria, A study on mental state classification using eeg-based brain-machine interface. 2018 International Conference on Intelligent Systems (IS), pp. 795–800 (2018)
24. J.J. Bird, A. Ekart, C.D. Buckingham, D.R. Faria, Mental emotional sentiment classification with an EEG-based brain-machine interface, in Proceedings of The International Conference on Digital Image and Signal Processing (DISP'19) (2019)
25. S. Vanneste, J.-J. Song, D. de Ridder, Thalamocortical dysrhythmia detected by machine learning. Nat. Commun. **9**(1), 1–13 (2018)
26. J.K. Feng, J. Jin, I. Daly, J. Zhou, Y. Niu, X. Wang, A. Cichocki, An optimized channel selection method based on multifrequency CSP-rank for motor imagery-based BCI system. Comput. Intell. Neurosci. **2019** (2019)

27. A.R. Jiménez, C.A. Grisales, J.L.R. Sotelo, Diseño de un sistema cerebro-máquina de miembro superior para la asistencia a la rehabilitación de personas con accidente cerebro-vascular, in *Encuentro Internacional de Educación en Ingeniería*, (2019)
28. S. Kumar, M. Chowdhury, M. Kumar, B. Hemalatha, Brain Controlled Robot for Indian Armed Force (2016)
29. E. Pérez, Un mono jugando al Pong es la primera demostración de Neuralink, el proyecto de Elon Musk para conectar el cerebro con los ordenadores. Xataka. (2021, 9 abril). https://www.xataka.com/investigacion/mono-jugando-al-pong-primerademostracion-neuralink-proyecto-para-conectar-cerebro-ordenadores-elon-musk
30. A. Roc, L. Pillette, J. Mladenovic, C. Benaroch, B. N'Kaoua, C. Jeunet, F. Lotte, A review of user training methods in brain computer interfaces based on mental tasks. J. Neural Eng **18**(1) (2021)
31. M.A.M. Joadder, S. Siuly, E. Kabir, A new way of channel selection in the motor imagery classification for BCI applications, in *International Conference on Health Information Science*, (Springer, Cham, 2018), pp. 110–119
32. W. Abbas, N.A. Khan, DeepMI: Deep learning for multiclass motor imagery classification, in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE, 2018), pp. 219–222
33. S. Sakhavi, C. Guan, S. Yan, Learning temporal information for brain-computer interface using convolutional neural networks. IEEE Trans. Neural Netw. Learn. Syst. **29**(11), 5619–5629 (2018)
34. A. Craik, Y. He, J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: A review. J. Neural Eng. **16**(3), 031001 (2019)
35. PhysioNet. (2019). https://physionet.org. Obtained from https://www.physionet.org/physiobank/database/eegmmidb
36. A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, et al., PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation **101**(23), e215–e220 (2000)
37. G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, BCI2000: A general-purpose brain-computer interface (BCI) system. IEEE Trans. Biomed. Eng. **51**(6), 1034–1043 (2004)
38. B. Medina, J.E. Sierra, B. MEDINA, A.B. Ulloa, Técnicas de extracción de características de señales EEG en la imaginación de movimiento para sistemas BCI. Revista ESPACIOS **39**(22) (2018)
39. M.A.P. Noguera, C.E.M. Ortega, W. Castro, D.H. Peluffo Ordoñez, Análisis De Señales EEG Para Detección De Intenciones Motoras Aplicadas A Sistemas BCI
40. P.V. Tirumani, S. Das, P. Swami, T. Gandhi, A low-noise low-cost EEG amplifier for neural recording applications, in *Advanced Computational and Communication Paradigms*, (Springer, Singapore, 2018), pp. 581–589
41. Y.L. Wu, D. Agrawal, A. El Abbadi, A comparison of DFT and DWT based similarity search in time-series databases, in *Proceedings of the Ninth International Conference on Information and Knowledge Management*, (ACM, 2000), pp. 488–495
42. C.S. Burrus, R.A. Gopinath, H. Guo, Introduction to wavelets and wavelet transforms: A primer. Prentice Hall (1998)
43. D. Wei, J. Tian, R.O. Wells, C.S. Burrus, A new class of biorthogonal wavelet systems for image transform coding. IEEE Trans. Image Process. **7**(7), 1000–1013 (1998)
44. M.H. Alomari, A. AbuBaker, A. Turani, A.M. Baniyounes, A. Manasreh, EEG mouse: A machine learning-based brain computer interface (2014)
45. I. Daubechies, Orthonormal bases of compactly supported wavelets. Commun. Pure Appl. Math. **41**(7), 909–996 (1988)
46. X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, et al., Top 10 algorithms in data mining. Knowl. Inf. Syst. **14**(1), 1–37 (2008)
47. C. Strobl, J. Malley, G. Tutz, An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol. Methods **14**(4), 323 (2009)

48. G.J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, vol 544 (Wiley, New York, 2004)
49. J.S. Cramer, The origins of logistic regression (2002)
50. C. Cortes, V. Vapnik, Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
51. N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression. Am. Stat. **46**(3), 175–185 (1992)
52. S. Sayad, *An Introduction to Data Science* (2021). https://injuryfacts.nsc.org/motor-vehicle/overview/introduction/
53. A. Natekin, A. Knoll, Gradient boosting machines, a tutorial. Front. Neurorobot. **7**, 21 (2013)
54. C. Lv, Y. Xing, J. Zhang, X. Na, Y. Li, T. Liu, et al., Levenberg–Marquardt backpropagation training of multilayer neural networks for state estimation of a safety-critical cyber-physical system. IEEE Trans. Ind. Inf. **14**(8), 3436–3446 (2017)
55. T.C. Major, J.M. Conrad, The effects of pre-filtering and individualizing components for electroencephalography neural network classification, in *SoutheastCon 2017*, (IEEE, New Jersey, 2017), pp. 1–6
56. J. Feng, E. Yin, J. Jin, R. Saab, I. Daly, X. Wang, et al., Towards correlation-based time window selection method for motor imagery BCIs. Neural Netw. **102**, 87–95 (2018)
57. P. Bashivan, I. Rish, M. Yeasin, N. Codella, Learning representations from EEG with deep recurrent-convolutional neural networks (2015), *arXiv*, preprint arXiv:1511.06448
58. X. Zhang, L. Yao, D. Zhang, X. Wang, Q.Z. Sheng, T. Gu, Multi-person brain activity recognition via comprehensive EEG signal analysis, in Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, pp. 28–37, (2017, November)
59. S.L. Wu, Y.T. Liu, T.Y. Hsieh, Y.Y. Lin, C.Y. Chen, C.H. Chuang, C.T. Lin, Fuzzy integral with particle swarm optimization for a motor-imagery-based brain–computer interface. IEEE Trans. Fuzzy Syst. **25**(1), 21–28 (2016)
60. D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, Y. Liu, Making sense of spatio-temporal preserving representations for EEG-based human intention recognition. IEEE Trans. Cybern. **50**(7), 3033–3044 (2019)

Automated Detection of Covid-19 Waves with Computerized Tomography Scan Using Deep Learning

D. Suganya and R. Kalpana

1 Introduction

Corona viruses are a large group of viruses; they cause various illnesses, ranging from the normal cold to more complicated diseases like Middle-East-Respiratory-Syndrome (MERS) and Severe-Acute-Respiratory-Syndrome (SARS). In 2019, a new virus was found that had never been seen in humans before [1]. It has been proven to be a communicable airborne infectious disease transmitted not only through human droplets, but also through the air. An accurate diagnosis remains the main challenge to alleviate this pandemic situation. The Covid-19 spread conferred a major problem to the diagnostic community; however, the courageous efforts of medical professionals have created an incredible effect on improving the identification and management of infected people as well as lowering the disease spread [2].

This paper discusses the techniques and the tests available to diagnose corona virus, like RT-PCR, swab test, and antigen test, with their challenges and comparatively analyze how CT scan image provides good result in diagnosis or screening of Covid pneumonia from common pneumonia with the help of references from numerous articles. Final study shows how DMIL and Mask R-CNN techniques can screen the Covid accurately. In addition, it needs to spotlight the fast improvement and implementations of diagnostic assays are vital to save you the spread of destiny novel infectious diseases.

D. Suganya (✉) · R. Kalpana
Puducherry Technological University, Puducherry, India
e-mail: suganyajagathguru@pec.edu; rkalpana@pec.edu

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_3

The remainder of the paper is organized in this way. Section 2 discusses several methods of diagnosis with its challenges. Section 3 addresses a summary of key publications for screening of Covid-19 with CT scan images. Section 4 gives the experimental findings and analyses of various datasets, whereas Sect. 5 presents the conclusions.

2 Related Studies

Several techniques have been used for rapid screening of Covid-19 accurately. Here we discuss the preeminent testing methods along with their merits and demerits.

2.1 Molecular RT-PCR Test

Molecular test is a majorly used testing method to diagnose sequences of DNA and RNA. Various molecular detection methods were used such as PCR, RT-PCR, qPCR, and RT-qPCR. Polymerase chain reaction (PCR) is a highly sensitive method for detection of sequences of DNA. RT-PCR is a method of reverse transcription-polymerase chain reaction used to find the presence of specific genetic material in pathogens [3]. Quantitative polymerase chain reaction (qPCR) and reverse transcription-polymerase chain reaction (RT-qPCR) are to identify, categorize, and quantify the nucleic acids for clinical relevant target.

Challenges RT-PCR testing for Covid-19 detection is accurate when the sample has active viral infection. But if the test is taken later, i.e., at the abortion stage of corona virus (day 11–14), it fails to predict whether you have been infected with the virus in the past [2].

2.2 RAT Tests

Antigen test is popularly known as RAT (Rapid Antigen Test). A sample from respiratory track is taken to detect the viral proteins and identifies the antigen connected with corona virus. Antigen test is used for speedy screening of Covid-19, which is comparatively faster than all other testing methods [4].

Challenges RAT is used for direct detection of viral (SARS-CoV-2) protein, i.e., antigen from the respiratory track, where the result is accurate and rapid with active corona viruses. But it fails to provide accurate results for negative cases. The negative result needs to be validated by some other tests.

2.3 Serological Tests

The test which is used to detect or identify the antibodies against the corona viruses is diagnosed with the help of serum or blood test is known as serological test. It is also known as serum test or antibody test. This test clearly tells how your immune system reacted to the virus in the past.

Challenges The serum blood test can determine whether your body has an antibody against the virus. But it fails to reveal about the active corona virus present in your body.

3 ML Algorithm Focused on Diagnosis Using CT Scan Images

ML algorithm focuses on applications that learn from experience and improve the accuracy of decisions and prediction accuracy time [5]. As shown in Fig. 1, algorithms in ML are very well trained to identify patterns and huge amounts of various features of data is being extracted to get more effective output from medical images for accurate prediction or diagnosis. The accuracy of output in a machine learning algorithm is entirely dependent on the amount of data that is trained; as we train or process more data, the accuracy increases.

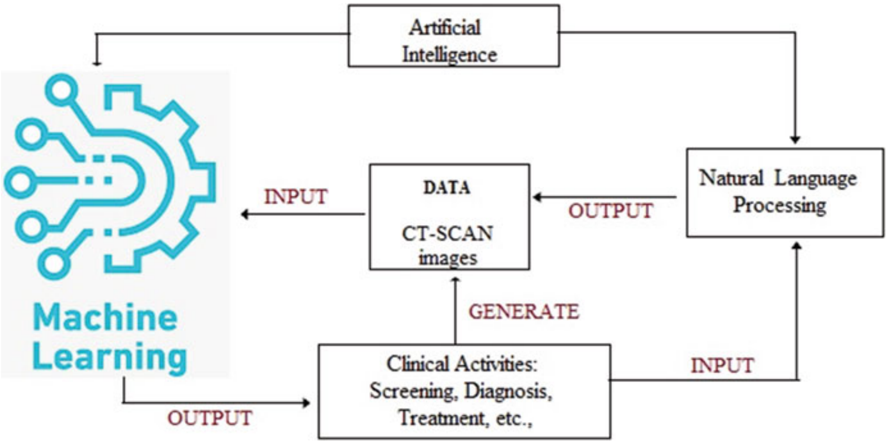


Fig. 1 Working of machine learning technique for clinical activities

3.1 Multiple Instance Learning (MIL)

MIL is a supervised ML algorithm. In general, each item is labeled individually whereas in Multiple Instance Learning, it receives a collection of multiple items in labeled bags. The bags can be labeled with positive and negative. If all the instances from the segmented image are negative the bag is marked as negative. If the sample with any one instance in a sample is positive, the bag is marked as positive. This algorithm effectively labels the individual object from a set of tagged bags.

A study has been recently published for Covid detection using chest CT images by Zhongyi Han et al. [6]. It impulsively detects the corona virus with attenuation-based multiple-instance learning algorithm. This algorithm facilitates large screening of corona virus effectively with labeled instances. It uses binary classification to classify whether the sample is labeled or tagged with positive or negative.

3.2 Why Deep Learning over Machine Learning

Machine Learning focuses on applications that learn through experiences, thereby improving prediction and making a right decision [7]. Several traditional ML algorithms are used for analysis and detection of Covid-19, but mainly focused in two ways, i.e., feature reduction techniques and model validation techniques.

Figure 2 depicts the relationship between AI, conventional ML techniques, and DL models [8]. Deep learning significantly reduces process time and human intervention [9]. The DL algorithm itself preprocesses, trains, and tests the datasets. Finally, it produces an accuracy based on the algorithm. Table 1 explains why we switch from machine learning to deep learning for detection when we have an image as input.

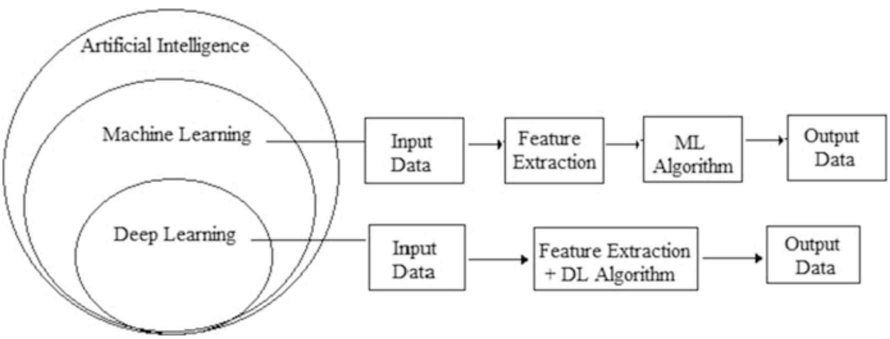


Fig. 2 Relationship among the subset of AI, ML, and DL

Table 1 shows how DL differs from traditional ML

Machine learning	Deep learning
In the problem-solving approach, the problem statement is first divided into several parts, then each part is solved, and finally, all of the parts are combined to yield a result	It will not break the problem; instead, it will attempt to solve the problem from start to the end
Multiple object detection problem	
In ML, we first use a support vector machine as an object detection algorithm to identify all objects, and then we use a histogram of oriented gradients as input for recognizing relevant objects. It takes a long time to complete the process	We use an image as input directly in DL techniques, such as Yolo net, which directly detect and provide the location as output. It solves problems directly with less amount of time

4 DL Algorithm Focused on Diagnosis Using CT Scan Images

The two most important sub-divisions of AI are ML and DL. Deep learning is necessarily a more progressive model compared to the traditional machine learning models. In ML algorithm, we take raw data as input and extract features from the preprocessed data. Then apply an algorithm to train the models for obtaining an accurate output or results. Deep Neural Learning or Deep Neural Network or Deep Learning algorithm, combine the process of feature extraction as well as training models or algorithms to achieve high accuracy of results effectively, as shown in Fig. 2. Several COVID cases can be detected accurately using chest CT scan than through RT-PCR tests. M. Chung et al. [10] recently published the features of individuals affected by novel corona virus detected using CT scan images. This method uses CT images effectively to find the how harm the virus is infected in the lungs. This paper also shows the high accuracy for prognosis of lung problems from the chest CT images.

4.1 Deep Neural Network (DNN)

DNN will create new patterns which contain multiple hidden layers. They have several types, where multiple neurons are processed like how our human brain neurons work. As shown in Fig. 3, the entire process took place in the hidden layer between an input and output layers. For example, if a DNN is trained to recognize a chest CT image, it will go over the image and calculate the likelihood that the virus is present [11, 12]. The user can examine the results and select the probabilities with which the network should display and return the tags [13, 14].

Hybrid 3D Deep Neural Network (H3DNN) paper explores the detection of novel corona virus from chest 3D CT scan images using a hybrid deep learning model

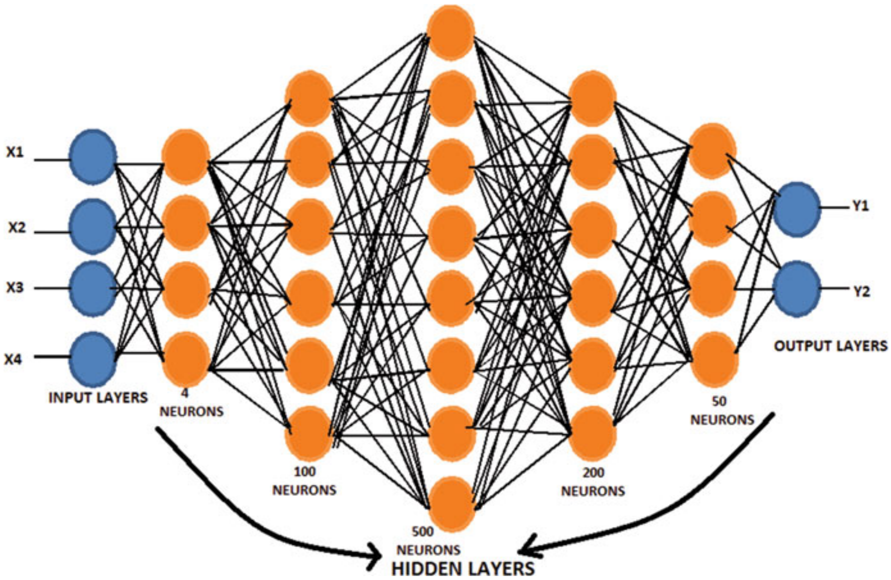


Fig. 3 Deep neural network with multiple hidden layers

which was recently published by Abdullah Aman Khan et al. [15]. It uses various datasets like CTC19 and COVIDCT, with ensemble of several transfer learning models like ResNet, DenseNet, and GoogleNet. This hybrid model works well in the diagnosis of Covid-19 with 3D-based deep learning models and acts as an effective tool in assisting the medical practitioner for large screening of novel corona virus.

4.2 Convolutional Neural Network

Convolutional Neural Network (CNN) generally analyzes, identifies, and categorizes the object in an image. ConvNet is composed of multiple layers of artificial neurons. It contains two basic operations such as convolution and pooling. CNN is a class of DNN; it contains one input layer, which combines convolution nonlinearity with pooling layers. The features are then extracted using vectors in the fully connected layers and normalization layers, as shown in Fig. 4.

O. Gozes et al. [16] recently published the corona virus detection and analysis using deep learning model with the help of chest CT. Several studies have focused on Covid-19 analysis and detection using ConvNet and medical images such as chest X-rays, computed tomographic scans, and so on. It can classify the process into distinct groups, such as Covid-19 or normal and viral pneumonia or common pneumonia or else normal/negative. With the help of various image dimensionalities and different ConvNet architecture, the output or result with minimized fully connected layers can identify Covid-19 from various medical images.

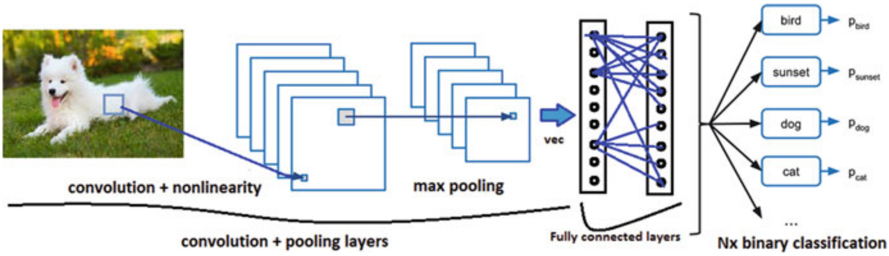


Fig. 4 Convolutional neural network working process

4.3 *Deep Convolutional Neural Network (DCNN)*

DCNN plays an essential role in image segmentation and classification. It is very effective in minimizing the number of frameworks without compromising the quality of models [17, 18]. It has the highest accuracy and is the more powerful among all the DNN algorithms such as ANN and RNN for predicting the images. Easy implementations and understand process which uses transfer learning techniques focus on gathering information for solving one problem and solve many problems related to that from the gained knowledge to get accurate output rapidly with less efforts [19–21].

Longxi Zhou et al. [22] discussed how to train the model using an accurate, rapid, and machine-agnostic segmentation and quantification method effectively. A new simulation model was created to dynamically represent the changes made for the Covid-19 infected area, and the dynamic model was used to expand additional data, which can improve the outstanding accomplishment of our segmentation model using DCNN. Through the development of a unified learning platform and heterogeneous data collection and development of preprocessing methods, the size of the data set is still limited.

Figure 5 here shows how DCNN preprocesses, extracts, and classifies the images [23, 24]. Initially, we take several chest CT images as input, then preprocess those raw input images and send for feature extraction. Here parallel deep feature extractors are used to extract the images using ResNet50V2 network and Xpection network which is much faster than the inception or GoogleNet. From the extracted features, Softmax classifier is added to classify the images into three different forms, whether it is normal or it has coon pneumonia or Covid pneumonia [25, 26].

Zhao Wang, Qi Dou, et al. [27] recently discussed about the correlative learning of redesigned Net such as COVID-Net for classification of corona viruses using CT images. It explains about the systematic study of various transfer learning models used for the classification of Covid-19. It creates a framework to work with various datasets for providing higher accuracy.

Chest CT images are used from screening and detecting Covid-19 based on Deep Learning networks and were discussed by Talha anwar et al. [28]. It clarifies the key differences between the normal CT image and viral CT image. To solve complex

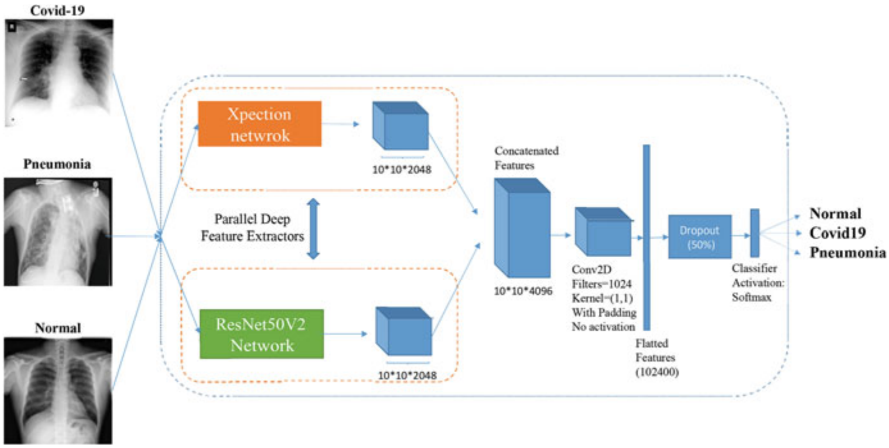


Fig. 5 Working of deep CNN to predict Covid-19

problems, various rates of learning strategies such as reducing plateau, recurring rate of learning, etc., have been used.

4.4 Transfer Learning Techniques in Deep Learning

Transfer Learning Techniques are widely used in DL for automatic detection. TL is a technique that focuses on gaining information (pre-trained models) by solving a specific problem and can be applied to a variety of related problems. The most commonly used Transfer Learning techniques for image segmentation and classification are (i) Fast R-CNN generate regions of interest with the help of selective search. (ii) Faster R-CNN is a variant of Fast R-CNN that employs the "Region Proposal Network" (iii) Mask R-CNN, a Faster R-CNN extension used for object detection and instance segmentation that is simple to train and adds little overhead to FR-CNN.

In general, CNN is used for the classifications and recognition of images, because of its high accuracy. VGG-16, ResNet50, and Inceptionv3 are a few examples of pre-trained image classification models. The process of classifying each pixel in an image as belonging to a specific category is known as image segmentation. Though there are several types of image segmentation methods, the two most commonly used are Semantic Segmentation and Instance Segmentation. Some of the Pre-Trained Models for Image Segmentation are U-Net, VB-Net, Inf-Net, or Semi Inf-Net, etc.

The related studies of Deep Learning methods used for Covid-19 diagnosis using chest CT images, as well as the datasets, transfer learning or pre-trained classifiers used by each technique, as well as its challenges and performance criteria, as shown in Table 2.

Table 2 Related studies of deep learning techniques for detection of Covid-19 using chest X-ray and chest CT scan images

References	Dataset used	DNN (TL) algorithms	Classifiers	Performance criteria (%)	Challenges
U. Ozkaya et al. [29]	Italian society of medical and interventional radiology dataset	Visual Geometric Group-19, GoogleNet, ResNet	Support vector machine	Accuracy is ~99.38 Sensitivity is ~93 Specificity is ~97.5 Precision is ~97.5	For covid-19 diagnoses, it employs a classification method. However, all deep learning models are insufficient to ensure success with the proposed method; new DL methods must be explored to ensure success
A. Ardakani et al. [30]	Clinical dataset	ResNet-101, Xception	Soft-Max classifier	Accuracy is ~99 Sensitivity is ~98.04 Specificity is ~100	Used to categorize and diagnose Covid-19 using Computer Aided Design (CAD) approach with the help of CT, whereas CAD system is not compared with radiologist
T. Ozturk, et al. [31]	Combination of different datasets	AlexNet or Image-Net	Dark Covid-Net classifier for YOLO	Accuracy is ~98 Sensitivity is ~100 Specificity is ~96	Covid-net model helps us to identify the covid-19 effectively but the dataset needs to be updated regularly in order to find an automatic detection

(continued)

References	Dataset used	DNN (TL) algorithms	Classifiers	Performance criteria (%)	Challenges
X. Wu, H. Hui et al. [32]	Clinical dataset	ResNet-50	Dense-Net-121	Accuracy is ~76 Sensitivity is ~81 Specificity is ~62	Rapid identification of Covid-19. Due to the limited data set, CT sliced pixels were used as samples for the DL model, which show fusion of inter-slice and tie-cost issues
H.S. Maghdid et al. [33]	Diagnostic imaging dataset	AlexNet, ResNet-34	Soft-max	Accuracy is ~94 Area under Curve (AUC) is ~96.35 Specificity is ~92	A large dataset of X-rays and CT image were used from various sources to provide a simple but effective detection result, but is not suitable for large data sources
L. Sun, et al. [34]	Clinical dataset	AFS-DF	Aggregation or cross-site learning approaches	Accuracy is ~91 Sensitivity is ~93 Area under Curve (AUC) is ~96 Specificity is ~89	This model is used to train a forest to reduce redundancy. It is mostly focused on feature which specifies the location whereas it is limited to validate only Covid-19 and CAP classification task

Y. Song, et al. [35]	Combination of different datasets	ResNet-101 DenseNet	Soft-max	Accuracy is ~98.5 AUC is ~99 Recall is ~93 Precision is ~96.43	Rapid and accurate identification of deep pneumonia to identify covid-19, implemented in super computer which enables parallel execution of 1000 tasks simultaneously
I. Razzak, et al. [36]	Cohens GitHub	Stacked Auto-Encoder used for Feature Reduction	Support Vector Machine	Accuracy is ~98 Sensitivity is ~69 Precision is ~70 F1-Score is ~69.5	It has a high level of accuracy for automatic diagnosis when using quantitative evaluation, but it is only useful for early screening and not for the complete screening procedure
S. Ozturk, et al. [37]	Clinical dataset	Covid-CTNet-112	Combined Decision Tree and Ada-Boost	Accuracy is ~83 Sensitivity is ~81 Specificity is ~84 Area under Curve (AUC) is ~90 F1-Score is ~77	The SMOTE method is used to eliminate unbalanced data, resulting in great accuracy in a short amount of time. It only works when we have previously trained datasets

(continued)

References	Dataset used	DNN (TL) algorithms	Classifiers	Performance criteria (%)	Challenges
S. Wang, et al. [38]	Diagnostic imaging dataset	M-Inception	Soft-max	Accuracy is ~89.5 Sensitivity is ~87 Specificity is ~88	With the pathogen images of conformed COVID cases, it employs modified Inception TL techniques. It is not appropriate for all data sources
T. Javaheri, et al. [39]	Aggregating multiple datasets	CovidCTNet open source	Soft-max	Accuracy is ~90 Sensitivity is ~80 Area under curve (AUC) is ~94	High accuracy and the model works with heterogeneous open source data, but uses very small sample sizes
X. He. et al. [40]	Clinical dataset	ResNet-V2 models, VGG-16, VGG-19	Ensemble learning approach	Accuracy is ~87.5 Sensitivity is ~87 F1-Score is ~85.0	The Self-Trans approach is used, which combines several learning techniques to produce unbiased results by reducing the over fitting problem. It has only been trained on publicly available datasets

X. Ouyang, et al. [41]	Combination of different datasets	3D ResNet-34 models	Multi-layer perceptron	Accuracy is ~87 Sensitivity is ~86.9 Specificity is ~90.1 Area under curve(AUC) is ~94.4 F1-Score ~ 82	The proposed model is being tested for tracking its consistency only when longitudinal data is ready
K. Elasnaoui and Y. Chawki [42]	Combination of different datasets	ResNet-50 models, VGG-16, VGG-19 Inception-V3 MobileNetV2	COVIDX-Net	Accuracy is ~99.5 Sensitivity is ~93	It uses automatic classification of Covid-19 with deep convolution neural networks, it uses transfer learning techniques which is not applicable for all data sets
K. Yang, et al. [43]	UCSD-AI4H datasets	CXR dataset, Inception-V1 And MobileNetV2	Softmax classifier	Accuracy is ~96 Area under curve(AUC) is ~99.4	It resizes an image during training phase, whereas the last hidden layer is learned by using t-SNE method. Quality becomes the biggest challenge while resizing the images

(continued)

References	Dataset used	DNN (TL) algorithms	Classifiers	Performance criteria (%)	Challenges
S. Rajaraman and S. Antani, [44]	Public datasets	DenseNet-121	Support vector machine	Accuracy is ~91 Recall is ~90.8 Precision is ~89.5 F1-Score is ~90	It uses datasets which represent pneumonia-related opacities; we train CNN using only DenseNet and get positive detection of Covid-19
R. Hu. et al. [19]	Combination of different datasets	CovXNet-V2	Linear layer	Accuracy is ~91 Sensitivity is ~91 Specificity is ~92 AUC is ~92	Insufficient samples also developed using augmentation of data for correct diagnosis. Misdiagnoses may occur when the necessary data is missing

Table 3 Comparison of various existing work for Covid-19 detection

Models	Accuracy	F1-Score	Specificity
Shallow CNN	95.9%	90.52%	92.9%
COVIDX-Net	93.4%	89.5%	94.3%
DenseNet-121	94.8%	92.3%	93.5%
VGG-19	95.5%	91.1%	97.5%
Mask R-CNN	96.98%	85%	**98.36%**
DMIL	**97%**	**91.3%**	96.5%

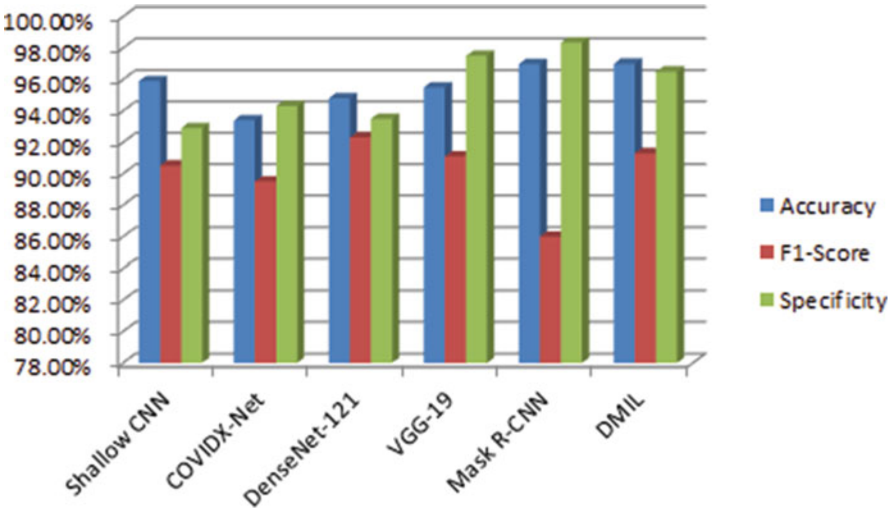


Fig. 6 Performance chart of various DL models

The performance parameters of all the six models were calculated, and the results of the parameter are given in Table 3. From this table, we came to know that among the six models, DMIL and mask R-CNN show an eminent performance for Covid-19 detection. Here, the parameters evaluated and compared are accuracy, specificity, and F1-score. By comparing these parameters with various algorithms, DMIL achieves maximum accuracy of 97%, F1-score of 91.3% and Mask R-CNN achieves the maximum specificity of 97.36%. The below chart shows the performance of various deep learning models as shown in (Fig. 6), which clearly exhibits the accuracy, specificity, and F1-score.

Table 4 shows the comparative analysis of various ensemble models with the proposed work. The performance parameters were calculated, and Mask R-CNN with DMIL technique shows an excellent performance with an accuracy of 98.96%, specificity of 97.83% and F1-score of 91.3%. These parameters are also explained clearly in the chart shown in (Fig. 7). By comparing all other existing ensemble models, the proposed DMIL with Mask R-CNN gives better accuracy for predicting Covid-19 using CT scan images.

Table 4 Comparison of various ensemble methods

Ensemble methods	Accuracy	Specificity	F1-Score
Xception65+VGG16	96.77%	97.48%	91.2%
DenseNet+ResNet-50	95.6%	97.5%	90.4%
MobileNetV2+CapsNet50	95.87%	96.13%	89%
MaskR-CNN+DMIL	**98.96%**	**97.83%**	**91.3%**

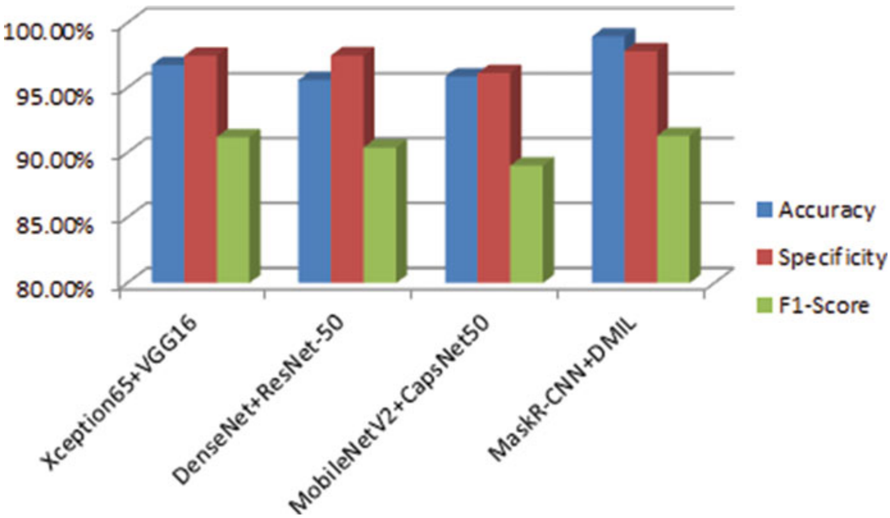


Fig. 7 Performance chart of various ensemble methods

5 Conclusion

An extensive review of research on diagnosis and prognosis using ML and DL networks was presented. A comparison of automated screening of Covid-19 from CT scanned images using various DL algorithms, including profound differential models such as MIL, DNN, Deep CNN, and Transfer Learning techniques, was discussed. Several tests were investigated for the detection of Covid-19 with its challenges, whereas the specialist can detect accurately using CT images. Various datasets were analyzed using pre-trained DNN, classifiers, performance metrics, and challenges. Transfer learning is only applicable to data that is similar in nature. The proposed ensemble models of DMIL with Mask R-CNN show a better accuracy of 98.96%. In the future, we will improve and expand the capability of transfer learning into multi-site settings as well as from other large-scale datasets to improve prognosis accuracy.

References

1. A. Nadya, J. Fouad, Editor in Chief's introduction to essays on the impact of COVID-19 on work and workers. J Vocat. Behav. **119**, 103441 (2020). https://doi.org/10.1016/j.jvb.2020.103441. Published online 2020 May 8. PMCID: PMC7205668
2. D. Suganya, R. Kalpana, Automated COVID-19 diagnosis using Deep Multiple Instance Learning with CycleGAN. 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), 2022
3. M. Eisenstadt, M. Ramachandran, N. Chowdhury, A. Third, J. Domingue, COVID-19 antibody test/vaccination certification: There's an app for that. IEEE Open J. Eng. Med. Biol. **1**, 148–155 (2020). https://doi.org/10.1109/OJEMB.2020.2999214
4. A. Crozier, S. Rajan, I. Buchan, M. McKee, Put to the test: Use of rapid testing technologies for covid-19. BMJ **372**(n208) (2021). https://doi.org/10.1136/bmj.n208
5. O. Shahid, M. Nasajpour, S. Pouriyeh, R.M. Parizi, M. Han, M. Valero, F. Li, M. Aledhari, Q.Z. Sheng, Machine learning research towards combating COVID-19: Virus detection, spread prevention, and assistance. J. Biomed. Inform. **117**, 103751 (2021). https://doi.org/10.1016/j.jbi.2021.103751
6. Z. Han, B. Wei, Y. Hong, T. Li, J. Cong, X. Zhu, H. Wei, W. Zhang, Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning. IEEE Trans. Med. Imaging **39**(8), 2584–2594 (2020)
7. M. Roberts, D. Driggs, M. Thorpe, et al., Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat. Mach. Intell. **3**, 199–217 (2021). https://doi.org/10.1038/s42256-021-00307-0
8. Y. Song et al., Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images (2020), *medRxiv*. [Online]. Available: https://www.medrxiv.org/content/early/2020/02/25/2020.02.23.20026930
9. O. Gozes et al., Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis (2020), *arXiv:2003.05037*. [Online]. Available: http://arxiv.org/abs/2003.05037
10. M. Chung et al., CT imaging features of 2019 novel coronavirus (2019-nCoV). Radiology **295**(1), 202–207 (2020)
11. J. Zhao, Y. Zhang, X. He, P. Xie, Covid-ct-dataset: A CT scan dataset about covid-19 (2020), *arXiv* preprint arXiv:2003.13865
12. J. Born, G. Brandle, M. Cossio, M. Disdier, J. Goulet, J. Roulin, N. Wiedemann, Pocovid-net: Automatic detection of covid-19 from a new lung ultrasound imaging dataset (pocus) (2020), *arXiv* preprint arXiv:2004.12084
13. F. Shi et al., Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification (2020), *arXiv:2003.09860*. [Online]. Available: http://arxiv.org/abs/2003.09860
14. M. Rahimzadeh, A. Attar, A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Inform. Med. Unlocked **19**, 100360., ISSN 2352-9148 (2020). https://doi.org/10.1016/j.imu.2020.100360
15. A.A. Khan, S. Shafiq, R. Kumar, J. Kumar, A. Ul Haq, H3dnn: 3D deep learning based detection of Covid-19 virus using lungs computed tomography. 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) | 978-0-7381-4259-3/20/$31.00 ©2020 IEEE | https://doi.org/10.1109/ICCWAMTIP51612.2020.9317357
16. O. Gozes, M. Frid-Adar, N. Sagie, H. Zhang, W. Ji, H. Greenspan, Coronavirus detection and analysis on chest CT with deep learning (2020), *arXiv:2004.02640*. [Online]. Available: http://arxiv.org/abs/2004.02640
17. X. Xu et al., Deep learning system to screen coronavirus disease 2019 pneumonia (2020), *arXiv:2002.09334*. [Online]. Available: http://arxiv.org/abs/2002.09334

18. S. Jin et al., Ai-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks (2020) medRxiv. [Online]. Available: https://www.medrxiv.org/content/early/2020/03/23/2020.03.19.20039354
19. R. Hu, G. Ruan, S. Xiang, M. Huang, Q. Liang, and J. Li, Automated diagnosis of covid-19 using deep learning and data augmentation on chest CT (2020), medRxiv.
20. S. Khobahi, C. Agarwal, M. Soltanalian, Coronet: A deep network architecture for semi-supervised task-based identification of covid-19 from chest x-ray images (2020), medRxiv
21. J.P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, M. Ghassemi, Covid-19 image data collection: Prospective predictions are the future (2020), *arXiv* 2006.11988. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset
22. L. Zhou, Y. Hong, T.L.X. Zhu, H. Wei, A rapid, accurate and machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis. IEEE Trans. Med. Imaging **39**(8), 2638–2652 (2020)
23. J. M. Banda, R. Tekumalla, G. Wang, J. Yu, T. Liu, Y. Ding, G. Chowell, A large-scale covid-19 twitter chatter dataset for open scientific research–an international collaboration (2020), *arXiv* preprint arXiv:2004.03688
24. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
25. L. Zhang, M. Wang, M. Liu, D. Zhang, A survey on deep learning for neuroimaging-based brain disorder analysis (2020), *arXiv* preprint arXiv:2005.04573
26. S. Wang et al., A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19) (2020), *medRxiv*. [Online]. Available: https://www.medrxiv.org/content/early/2020/04/24/2020.02.14.20023028
27. Z. Wang, Q. Liu, Q. Dou, Member, IEEE, Contrastive cross-site learning with redesigned net for COVID-19 CT classification. IEEE J. Biomed. Health Inform. **24**(10), 2806–2813 (2020)
28. T. Anwar, S. Zakir, Deep learning based diagnosis of COVID-19 using chest CT-scan images. 2020 IEEE 23rd International Multitopic Conference (INMIC) | 978-1-7281-9893-4/20/$31.00 ©2020 IEEE | https://doi.org/10.1109/INMIC50486.2020.9318212
29. U. Ozkaya, S. Ozturk, M. Barstugan, Coronavirus (covid-19) classification using deep features fusion and ranking technique (2020), *arXiv* preprint arXiv:2004.03698
30. A.A. Ardakani, A.R. Kanafi, U.R. Acharya, N. Khadem, A. Mohammadi, Application of deep learning technique to manage covid-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Comput. Biol. Med. **121**, 103795 (2020)
31. T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of covid-19 cases using deep neural networks with x-ray images. Comput. Biol. Med. **121**, 103792 (2020)
32. X. Wu, H. Hui, M. Niu, L. Li, L. Wang, B. He, X. Yang, L. Li, H. Li, J. Tian, et al., Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study. Eur. J. Radiol. **128**, 109041 (2020)
33. H.S. Maghdid, A.T. Asaad, K.Z. Ghafoor, A.S. Sadiq, M.K. Khan, Diagnosing covid-19 pneumonia from x-ray and CT images using deep learning and transfer learning algorithms (2020), *arXiv* preprint arXiv:2004.00038
34. L. Sun, Z. Mo, F. Yan, L. Xia, F. Shan, Z. Ding, W. Shao, F. Shi, H. Yuan, H. Jiang et al., Adaptive feature selection guided deep forest for covid-19 classification with chest CT (2020), *arXiv* preprint arXiv:2005.03264
35. Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, H. Zhao, Y. Jie, R. Wang et al., Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with CT images (2020), medRxiv
36. I. Razzak, S. Naz, A. Rehman, A. Khan, A. Zaib, Improving coronavirus (covid-19) diagnosis using deep transfer learning (2020), medRxiv
37. S. Ozturk, U. Ozkaya, M. Barstugan, Classification of coronavirus images using shrunken features (2020), medRxiv

38. S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng et al., A deep learning algorithm using CT images to screen for corona virus disease (covid-19) (2020), MedRxiv
39. T. Javaheri, M. Homayounfar, Z. Amoozgar, R. Reiazi, F. Homayounieh, E. Abbas, A. Laali, A.R. Radmard, M.H. Gharib, S.A. J. Mousavi et al., Covidctnet: An open-source deep learning approach to identify covid-19 using CT image (2020), arXiv preprint arXiv:2005.03059
40. X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, P. Xie, Sample-efficient deep learning for covid-19 diagnosis based on CT scans (2020), medRxiv
41. X. Ouyang, J. Huo, L. Xia, F. Shan, J. Liu, Z. Mo, F. Yan, Z. Ding, Q. Yang, B. Song, et al., Dual-sampling attention network for diagnosis of covid-19 from community acquired pneumonia. IEEE Trans. Med. Imaging (2020)
42. K. Elasnaoui, Y. Chawki, Using x-ray images and deep learning for automated detection of coronavirus disease. J. Biomol. Struct. Dyn.., no. just-accepted, 1–22 (2020)
43. K. Yang, X. Liu, Y. Yang, X. Liao, R. Wang, X. Zeng, Y. Wang, M. Zhang, T. Zhang, End-to-end covid-19 screening with 3d deep learning on chest computed tomography (2020)
44. S. Rajaraman, S. Antani, Training deep learning algorithms with weakly labeled pneumonia chest x-ray data for covid-19 detection (2020), medRxiv

Internet of Things (IoT) in the Agriculture Sector: Challenges and Solutions

Aakanksha Jain, Nikhil Ranjan, Sarvesh Kumar, and Sunil Kumar Vishwakarma

1 What Is Advance Agriculture?

Advance farming is a farming administration idea utilizing the current innovation to expand the amount and nature of agrarian items. Ranchers in the twenty-first century approach GPS, soil filtering, information on the board, and Internet of Things advancements. By unequivocally estimating varieties inside a field and adjusting the system as needs, ranchers can extraordinarily build the adequacy of pesticides and composts and use them all the more specifically. So also, utilizing advance farming systems, ranchers can all the more likely screen the necessities of individual animals and modify their nourishment correspondingly, in this way forestalling ailment and improving crowd well-being.

Advanced farming can be described as advanced self-adaptive; we can say the self-conscious mechanism of enhancing the quality and quantity of let's say grain, which is grown by traditional systems.

A. Jain (✉)
Department of Computer Science Engineering, Poornima University, Jaipur, India

N. Ranjan
Department of Computer Science Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

S. Kumar · S. K. Vishwakarma
Department of Computer Science Engineering, Babu Banarasi Das University, Lucknow, Uttar Pradesh, India

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_4

2 Why Is It Necessary to Adapt to the Advancement of Agriculture?

The smart development is comprehended by the astute wearables, related devices, modern machines, and modular movables. Regardless, in cultivating, the IoT has brought the best impact. We look at estimations; these estimations always reveal that the overall people will show up at 9.6 billion by 2050 (Graefe S et al. [12]). Besides, to deal with this massive population, the cultivation business is restricted to grasp the Internet of Things. Among the challenges like uncommon atmospheric conditions, climatic changes, and characteristic impact, IoT could be the obliterating factor for these troubles and in helping us to fulfill the requirement for greater sustenance.

As we know, in earlier agriculture, pedagogy of world machine is not much in use; we use animal- and human-dependent agriculture techniques, which are not only time-consuming but also have major capacity constraints and are also easily affected by environmental factors. For now, mechanical turns of events, for instance, tractors and mechanical-based tools and techniques, happened and were brought into the cultivating exercises in today's agriculture. Likewise, the agribusiness industry relies strongly upon innovative contemplations, considering the reliable creation of enthusiasm for sustenance.

The industrial IoT has been a central purpose behind extended agrarian creation at a lower cost; the usage of sharp courses of action powered by IoT will increase in the cultivation exercises. These courses and implied techniques will surely give a percentile (17–20%) growth speed in the existing business of agriculture (Sala S, Rossi F, David S (eds) [16]).

According to reports, it may be assumed that there would be a vast (in billions) increase in smart devices' usage imposed upon traditional agriculture as Fig. 1.

Because of the absence of steady and solid correspondence engineer framework, an IoT plan supplier likewise as the representatives had gone facing execution challenges in remote or less made zones. Contrarily, various structure suppliers are making it conceivable by presenting satellite openness and using cell systems.

3 Benefits of Using IoT in Agriculture

1. *The Rapid Change in Environment Condition*

   Besides, having ill-advised information about the atmosphere truly breaks down the total and nature of the yield creation. Regardless, IoT game plans empower you to comprehend the consistent environmental conditions. Sensors are put inside and outside of the agribusiness fields. The mass information regarding the quality of soil, and cultivation from their locality which is utilized to pick right, privilege procures which can make and reinforce the specific climatic conditions ([1], p. 1–12). The entire IoT implantation in the natural structure of selected landmarks, including port sensors that can perceive and

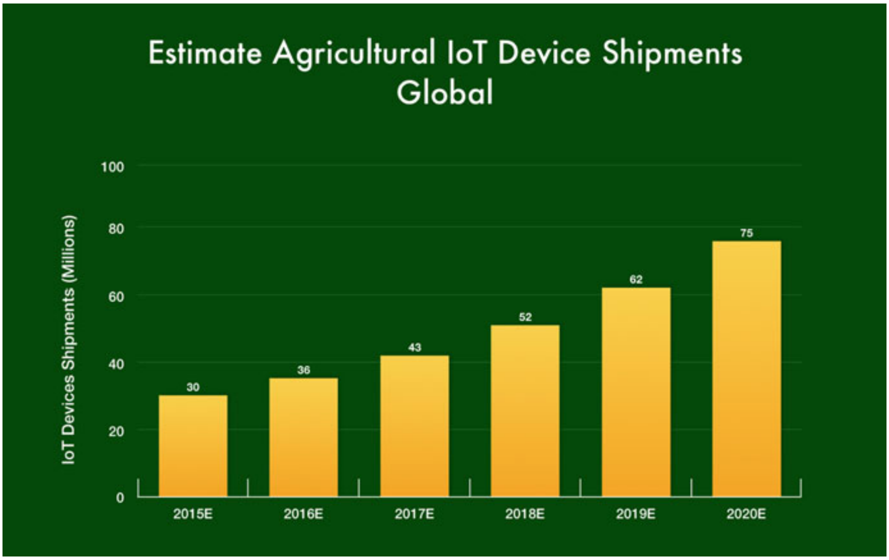


Fig. 1 Global estimation of IoT devices

determine consistent air conditions like precipitation and temperature, significantly more accurately. The distinctive number of sensors is accessible to see these parameters and organize them as prerequisites to suit your smart creating necessities. These sensors monitor the state of the harvests through various matrices and the climate fusing if any upsetting air conditions are found, by then an alarm is sent for suitable action to be taken as a solution to the perceived problem for the crop. What gets out is the need for physical closeness during upsetting climatic conditions which as time goes on broadens the profit and helps ranchers to get more agribusiness rewards.

2. *Traditional Farming Techniques*

In customary cultivating, ranchers need workers for taking care of harvests and domesticated animals. Thus, they squander assets on superfluous things. However, with an IoT-fueled framework, the ranchers can assemble information explicitly identified with crop cultivating like temperature, precipitation, soil quality, and so forth. Simultaneously, they can screen the soundness of domesticated animals and alter their nourishment in like manner. Through the broken-down information, they can recognize whether the creatures are wiped out or not as well as screen harvest's well-being and nourishment needs. On the off chance that creatures are wiped out, they are detached from others to forestall the spread of illness. On the off chance that plants are not well, at that point extra sustenance can be enhanced to them. Along these lines, ranchers can deal with their field exercises sagaciously and effectively.

Fig. 2 Monitoring using sensors

3. *Monitoring Through Sensors*
   In IoT-based agricultural techniques, we replace traditional farming tools with smart gadgets like sensor levelers and transponders. This smart equipment makes our life much easier by giving us error less monitoring and exact information about the current situation about any field let's say crop, animals, weather situation, etc.
   In animal farms, there is a provision for continuous screening of all animals, i.e., hens, goats, cow, bull, pig, etc., for various keeping track of the health of these animals by giving exact information about foreseeing infections and segregation [17]. These pieces of information make it easy for a rancher to keep its farm animal healthy and hence lead to an immense lead in quantity and quality of the product produced (Fig. 2).
   Other benefits could be listed like, by a continuous screening of its animal, a rancher can be informed about the health issue of its animal before the condition become severe, which implies a saying "healthy animal, wealthy the rancher." With smart tools of screening, a rancher doesn't need to depend on a health worker to check animal's health and well-being, temperature, movement, and change in behavior due to weather; these data will directly be stored at the server and in a form of information which is easily understandable to the cattleman [2, 7].
4. *Production of High-Quality Crops*
   When using traditional farming techniques, the crop is prone to weather conditions, as the sudden change in weather may have a higher impact on the crop

in the field. Contrary to smart farming, the weather conditions are in continuous monitoring of smart gadgets, for example, allMETEO, Smart Elements, and Pycno. These smart gadgets give exact statistical information about weather, soil, air, humidity, and several other parameters that have a higher impact on the crop [6, 18]. All this information is directly stored in the cloud by using some cloud storage service provider company; when stored in the cloud, the information is processed and analyzed with a machine learning algorithm, i.e., supervised or unsupervised or reinforcement algorithm. After this analysis process, this generated information or outcome of an action is directly accessible to the farmer on its device like mobile, laptop, desktop, etc.

Infusion of IoT in the traditional system makes it possible to avoid challenges and remove all issues that may arise during farming processes. So the quality of the product is growing and consumers get a good product of high quality. In farming, sowing seeds is a tedious and difficult task. With advanced seeding machines or robots, the task of sowing seeds becomes easier. The seeding machines or robots come with geo-mapping and sensor data [8].

5. *Use of Smart Gadgets like Drones*

   The arrival of agricultural robots is the floating disturbing influence of imaginative movements that have nearly modified the cultivating exercises. Aerial and ground machineries are utilized for yield evaluation, crop monitoring, planting, crop sprinkling, and field inspection [12]. Meandering aimlessly, growth has provided the agriculture business a high increase and makeover, thanks to the real structure and arranging susceptible to continuous data. Warm or multispectral sensors enable machines to distinguish the domains that necessitate alterations in the water framework [14, 15]. At the point when the yields start creating, sensors exhibit their prosperity and process their vegetation record. Over the long haul, splendid robots have diminished the natural impact. The results have been with the ultimate objective that there should be a colossal diminishing and much lower substance going to the groundwater.

6. *Accuracy in Data Analysis*

   The introduction of agricultural robots is the floating aggravation of inventive motions that have almost revolutionized the cultivating exercises. Robots on the ground and in the air are utilized to analyze yield prosperity, crop monitoring, planting, crop sprinkling, and field inspection. Meandering aimlessly, development has provided the farm industry a high increase and makeover, because to the genuine framework and masterminding based on consistent data. Machines with warm or multispectral sensors recognize the regions that require changes in the water framework. At the point when the yields start creating, sensors exhibit their prosperity and process their vegetation record. Over the long haul, splendid robots have diminished the natural impact. The results have been with the ultimate objective that there should be a gigantic diminishing and much lower substance going to the groundwater (Fig. 3).

7. *Less Wastage of Natural Resources like Water*

   A boundless water system strategy known as subsurface drip irrigation (SDI) is a solution to the wastage of water resources by fusing these SDI frameworks

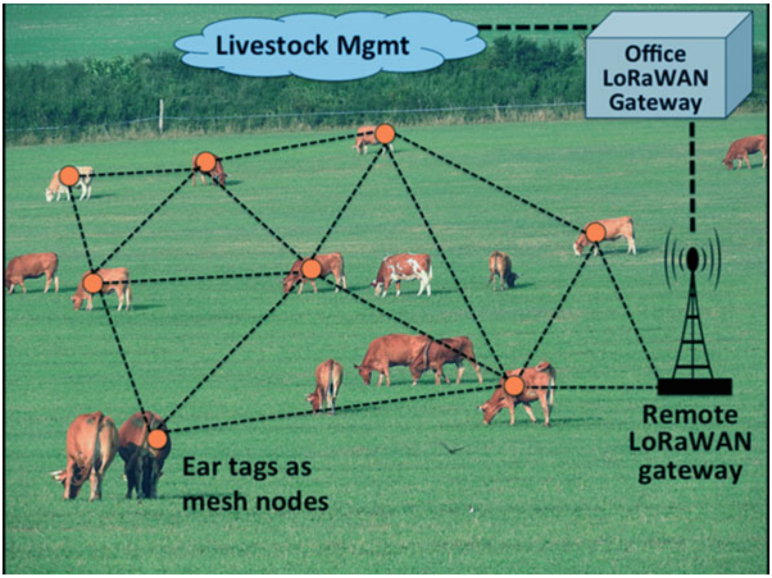


Fig. 3 Accuracy using live monitoring

ranchers can control when and how much water their yields can take this strategy [11, 12, 14–16]. With keen IoT-empowered sensors, ranchers can consistently follow dampness levels and plant well-being and intercede just when it is important.

4 Infusion of Internet of Things in Agriculture

Due to the higher population increase rate worldwide, it seemingly implies major issues in fulfilling the basic needs using existing agribusiness. Despite fighting difficulties like an unexpected natural disaster, immense change in weather conditions, increased pollution, and higher risk of the infective virus due to change in our geological cycle and cultivation's ecological effect, the demand for more nourishment will surely increase in the future [11, 12, 14–16]. For fulfillment of these expanding demands, we need a miracle in our traditional agribusiness with innovation and techniques. Newly proposed shrewd cultivating applications dependent on IoT technologies will empower the agribusiness to decrease waste and improve profitability from upgrading compost use to expanding the productivity of ranch vehicles' courses.

"The essential test for mindfulness among producers is simply the term IoT. If we somehow managed to get some information about constant sensors, climate stations or information from tractors, for instance, they are very much mindful

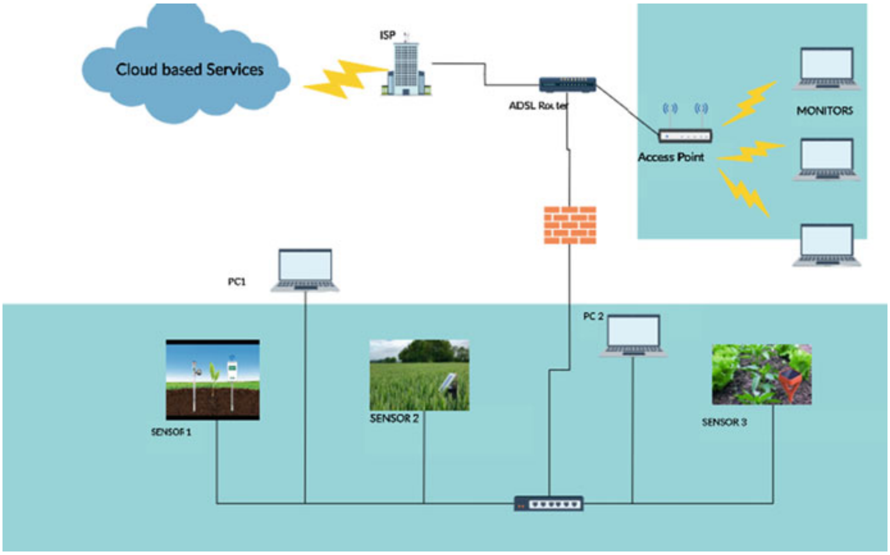


Fig. 4 Infusion of internet of things in agriculture

of the innovation and its significance, yet more should be done as an industry to manufacture familiarity with and conversation around the Internet of Things and how the investigation can be utilized to help bolster day by day on-ranch dynamic," said Harvey.

The Internet of Things (IoT) has produced a great deal of consideration throughout the years and in light of current circumstances [19]. In the present digitized age, IoT innovation is helping numerous businesses improve their activities. Indeed, even farming is beginning to grasp IoT availability. Worldwide nourishment request is developing, and customary farming basically can't deliver enough nourishment to meet the world's wholesome needs. To spare people in the future from craving and lack of healthy sustenance, ranchers should convey progressively creative methods to build their yields (Fig. 4).

5 Challenges Encountered in Infusion of IoT in Agriculture

By getting IoT in the green fragment, we get different points of interest, yet simultaneously, there are challenges glanced by IoT in-country divisions. The best challenges glanced by IoT in the agrarian portion are nonattendance of information, high determination costs, security concerns, etc. Most of the farmers don't think about the utilization of IoT in agribusiness. The major issue is that some of them are against new considerations, and they would lean toward whether or not it gives different preferences as compared to the traditional approach. The best thing about the adoption of IoT based farming approach is the rancher will use modern equipment like sensors, monitors, robotic solutions, and other hyper-tech solutions of farming for the betterment of their crops. And these sensors and monitors will take care of the current situation of the crop in a way by keeping a record of weather and crop situation through various set matrices [3, 10].

About IoT in farming, a few difficulties emerge. Initially, the hardware living at the discernment layer must be presented legitimately to unforgiving ecological marvels, similar to high sunlight-based radiation, extraordinary temperatures, downpour or high mugginess, solid breezes, vibrations, and different threats fit for obliterating the electronic circuits. The end gadgets should remain dynamic and capacity dependable for significant stretches depending on the restricted force assets of batteries. In this manner, proper programming instruments and low-power abilities are obligatory, since the continuous battery substitution or reset of the stations (if there should be an occurrence of a program disappointment), for instance, in a large-scale open field arrangement, isn't simple. Force collecting can be an answer for some degree, notwithstanding, the force utilization has still to be inside the force spending plan of little force reaping modules (e.g., sun-oriented boards, wind turbines, and so on).

1. *Absence of infrastructure:* even if the ranchers receive IoT innovation, they won't have the option to take advantage of this innovation because of poor correspondence framework. Ranches are situated in remote zones and are a long way from access to the web. A rancher needs to approach crop information dependably whenever from any area, so association issues would make a propelled observing framework pointless.

Solution

As a solution for ranchers, Internet of Things provides smart devices which are provided as infrastructure and are a part of services provided by one of cloud computing types, which is IAAS (infrastructure as a service). We can say these services are provided to smart agriculture prototype, in association with IAAS services. These services can be categorized as automatic or semiautomatic services according to purpose of service and need of service [4, 13]. With this decade the demands have been increased excessively, and according to demands, the ratio of manpower is less, or the capability and ability to work by a human are not at that distinct level as there could be a chance of human error, failure in proper analysis, etc.; all these factors give a reason in the rise of machine intervention in the area of agriculture which can be annotated as "agrobots." These agrobots have attracted the farmers in a rapid scale in this time, and according to studies, the demand will surely increase in the future. Crop production decreased by an estimated 213 crores approximate ($3.1 billion) a year due to labor shortages in the United States alone. The world is in the early stages of an agrobotic revolution with most of the products still in trial phases and R&D mode [17].

Self-loading robots with arms can detect weeds and spray pesticides on the afflicted plants, putting the plants aside like pesticide costs in general. These robots can also help with reaping and lifting. Overwhelming cultivating cars may also be explored from the comfort of one's own home using phone displays to complete tasks, and GPS can track their whereabouts at all times.

The ranches are imaged, mapped, and studied using robots equipped with sensors and cameras. They can be controlled remotely, or they can fly by encoding controlled flight designs into their implanted frameworks and collaborating with sensors and GPS [18]. Bits of knowledge on edit well-being, water system, splashing, planting, soil and field, plant tallying and yield expectation, and a lot more can be obtained from the automaton data.

Sensors, such as climate stations, are placed alongside dwellings to collect social event data, which is then communicated to explanation apparatuses for examination. They examine the yields for light, moisture, temperature, shape, and size variations. The data collected by sensors in terms of mugginess, temperature, dampness, precipitation, and dew identification aids in determining the climatic design in ranches to achieve appropriate yields. Examining the type of soil can assist in determining the supplement value and drier zones of homesteads, soil waste limit, or sharpness, allowing for changes in the amount of water required for the water system and the selection of the most beneficial type of development. PC imaging entails the use of sensor cameras placed across the ranch or automatons equipped with cameras to deliver images that have undergone advanced image processing. The images are used for quality control, disease detection, arranging, and reviewing yield and water system checks using image handling combined with AI, which compares pictures from the database to pictures of harvests to determine the size, shape, shading, and development and thus control the quality [8].

2. *Significant expense:* equipment expected to execute IoT in farming is costly. Anyway, sensors are the most economical segment, yet equipping the entirety of the ranchers' fields with them would cost more than a thousand dollars. Mechanical hardware cost more than that physically worked apparatus as they incorporate expense for ranch the executives programming and cloud access to record information. To acquire higher benefits, ranchers must put resources into these advancements; anyway it would be hard for them to make the underlying speculation to set up IoT innovation at their homesteads.

Solution

For certain farmers and those in the rustic section, a considerable amount of their work incorporates giving up control to the forces of nature: the seasons, the air structures, and the precipitation levels are in charge. Regardless, envision a situation where farmers could recover control with the help of development particularly the cloud. In this scenario we can use distributed computing as a solution, as distributed computing is used to acquire complete information and statistics from sensors, weather conditions, soil fertility ratio, and all other useful information, while the actual computing is done at a remote area; the farmer can control and see the current situation of its farm or field using a handheld device. The cloud's explanatory capacities likewise help ranchers in understanding their creation condition. "The

Fig. 5 Use of smart devices for monitoring

horticulture business is using information like never before previously," says Brian Luck, an organic frameworks building teacher at the University of Wisconsin-Madison [4, 13]. "Information is assembled on the dirt inside fields, the plants developing on fields, the climate happening during the developing season, and the apparatus utilized in fields."

As a complete scenario when we use a distributing computing solution, we facilitate with the feature like complete monitoring via real-time information acquisition and instantly computing the required calculation and generate results or notes accordingly, so the farmer/rancher can get to know about the only thing he understands and do the needful accordingly. The rancher need not know about complexity of calculation and installation procedure (all technical complexities), and all these things are taken care by remote administration only (Fig. 5).

A portion of the particular uses is as per the following:

- *Harvest-related data:* It can catch data identified with all yields developed in the past and in this manner can assist ranchers with settling on choices on what to develop straight away.
- *Climate data:* The cloud can store area-explicit climate data and the climate figure for explicit terms. Once more, these assist ranchers with settling on crop-related choices.
- *Soil information:* Yield-related dynamic relies to a great extent upon soil data as well. Aside from the soil profile, it can likewise give a pattern of previous soil, which will help in foreseeing the pattern in the future. For instance, is the dirt turning acidic/basic, or what different changes in nature and creation of soil can be seen?

- *Checking growth:* The development of different harvests can be observed in various areas and at standard interims. This empowers development examples to be contrasted and past development designs.
- *Ranchers' data:* Area shrewd rancher information can be caught, screened, and studied with the inclusion of nearby ranchers. This can help in the recognizable proof of center agrarian regions, which are useful for policymakers while surrounding their methodologies.
- *Master consultation:* Here, arrangements are accessible for typical issues those ranchers much of the time face. Close by, specialists give answers for explicit issues as well, with somewhat short reaction time (cf. [9]).

3. *Web of Things Equipment and Software Challenges in Agriculture*

   When it comes to IoT in agribusiness, a few difficulties arise. Firstly, the gear living at the observation layer has to be presented straightforwardly to cruel natural marvels, like high sun-oriented radiation, extraordinary temperatures, downpour or high humidity, solid breezes, vibrations, and different perils fit of destroying the electronic circuits. The end gadgets will have to stay dynamic and capacity dependable for significant stretches depending on the limited power assets of batteries. Thus, appropriate programming instruments and low-power capacities are mandatory, since the successive battery substitution or reset of the stations (if there should arise an occurrence of a program disappointment) [3], for instance, in an enormous scope open field sending, isn't simple. Force reaping can be an answer to some degree, be that as it may, the force consumption has to be still inside the force spending plan of little power harvesting modules (e.g., sun-powered boards, wind turbines, etc.).

Solutions

"Distributed computing requires itemized data about ranch conditions and tasks to make a trip to outside gatherings, including the cloud supplier and outsiders that help investigate the information," says Solon Barocas, a data science teacher at Cornell University. "This can raise both protection and security worries for ranchers, as they stress over who may access data that uncovers their cultivating techniques, their yield, their property, and their business activities" [7].

This is one of the issues tended to by Krintz when she created SmartFarm, an open-source, minimal effort, start to finish choice help and robotization framework for ranch forms. SmartFarm utilizes sensors to screen water and supplement development, soil conditions, and dampness levels. The information given by these sensors is handled in the cloud, together with outer information, similar to climate data and satellite pictures. Ranchers would then be able to see the outcomes through an application introduced on their cell phones.

"We have planned SmartFarm to have an on-ranch part which reflects what's done in the cloud," Krintz says [5]. "This SmartFarm box is a little registering framework that looks and runs like a DVR from your link organization. It gives the entirety of the equivalent examination, representation, and choice help usefulness that arrangements in the cloud do, yet the information never needs to leave the ranch" (cf. [9]).

By making this little however huge change, Krintz takes note of that SmartFarm empowers ranchers to look after control, possession, and security of their information. She says ranchers can even now share their information and move it to the cloud in a controlled way on the off chance that they wish to do as such.

The cloud is a promising innovation that can change how we ranch. "It's a key help that permits us to total, break down, and share information across homesteads, landmasses, and the world," says Krintz. "It's a significant bit of a lot bigger riddle—that of expanded ranch gainfulness and reasonable nourishment creation."

In reality, the cloud might just be the eventual fate of cultivating.

4. *Networking Device Interoperability Issues*

   When implementing smart agriculture in open agriculture environment, one of the most faced issues which occurs to a rancher is interoperability issue of all smart devices which he is going to install at its open agriculture environment for monitoring or screening measures. As we know IoT as the name indicates, it is "provided interconnection among things or you can say devices," and these devices are possibly connected through an internetworking framework that may be wired or a wireless one. As reported by the study, it is indicated that IoT stage is going to increase in the day-to-day life and growing in the market of IoT market for the world of the organization for the most IoT Framework to organized in world market.

Solution

Standardization Could Be a Solution

On the off chance that we take a gander at the present situation, we can say that the issue of interoperability can be improved. The market is extremely divided, particularly because of contrary qualities among brands, and a typical exertion is expected to arrive at regular measures for correspondence.

The IoT theory isn't actually to make a neighborhood shut and in restricted conditions. The way of thinking of IoT is to make a world wherein a great many gadgets can speak with one another in the most ideal and largest manner, without specialized or business restrictions, to make our lives somewhat better [2].

Nonetheless, this isn't an outlandish undertaking. It is in no way, shape, or form the first occasion when the main innovation designers and makers have consented to set commonly acknowledged guidelines. How about we think, for instance, about the Internet, and how the homogenization of correspondence conventions has prompted the development of the system?

There are a few activities to help interoperability in IoT.

One of them is, for instance, IEEE P2413 – standard for an architectural infrastructure for the Internet of Things – a normalization venture planned for distinguishing similitudes in IoT situations as different and as insightful structures, savvy transport frameworks, or human services (cf. [9]).

The other one is an EU venture, called IoT-A, made to create models that can be applied in various areas. We additionally have an open-source activity called IoTivity, with more than 300 individuals, remembering driving organizations for the division, which is planned for directing and advancing participation among organizations and engineers, or, on the other hand, the purported Industrial Internet Reference Architecture (IIRA) (cf. [9]), made in 2014 by a portion of the primary administrators in the market and concentrated on modern IoT applications.

In such an assorted and unyielding region, we accept that there have been various endeavors at unification, yet it isn't conceivable to figure out which normalization measures will, at last, be picked.

The sign of the interconnectivity challenge in IoT has been underscored by both the scholarly world and industry. The business endeavors to address IoT interoperability challenges through normalization. A few rises have been made to set up principles for giving interconnectivity between IoT gadgets, systems, administrations, information designs claimed by various suppliers.

5. *The Bandwidth of Network per Region Varies*

   Another unavoidable issue can be discussed as *network bandwidth distribution* [17] which varies region to region. This variation in distribution can have a higher impact on the network layer, which plays an important role in data transfer. And being able to transfer data and produce some analysis on it is considered as the most important thing in smart agriculture and hence could be abbreviated as the backbone of smart agriculture. Remote correspondence is the most well-known in farming arrangements, because of the absence of wiring costs. Condition is known to be one of the central points which lead to low remote connection quality, through the multiway spread impacts and its commitment to foundation clamor. Real-world arrangements have demonstrated that the presence of well-known handsets is influenced by temperature, moistness, human nearness and different hindrances inside the space in which a remote hub endeavors to convey. Along these lines, information must be moved to utilize powerful and dependable innovations, as indicated by the necessities and difficulties of the country's condition.

Solution

A remote sensor network (RSN) [17] is a combination of scattered sensors that screen physical or biological conditions, for instance, temperature, sound, and weight. Data from each sensor experiences the framework center point.

WSN centers are easeful contraptions, so they can be sent in high volume. They also work at low power so they can run on battery or even use essentialness gathering. A WSN center point is an embedded structure that ordinarily plays out a singular limit (with regard to model, assessing temperature or weight, or turning on a light or a motor). Essentialness gathering is another development that gets imperativeness from outside hotspots (e.g., sun-based power, warm essentialness, wind essentialness, electromagnetic radiation, and dynamic essentialness, and that is just a hint of something larger). The imperativeness is gotten and taken care of for use by close to nothing, low-power remote independent devices, like the centers on a WSN (cf. [9]). WSN center points are insignificant exertion devices, so they can be passed on in high volume. They moreover work at the low power with the objective that they can run on battery or even use imperativeness harvesting. A WSN center point is an embedded system that typically plays out a loan limit (with respect to model, evaluating temperature or weight, or turning on a light or a motor). Essentialness procuring is another advancement that gets imperativeness from external hotspots (e.g., sun-based power, warm imperativeness, wind essentialness, electromagnetic radiation, and dynamic imperativeness, and that is only the start). The imperativeness is gotten and taken care of for use by pretty much nothing, low-power remote independent contraptions, like the centers on a WSN. The battle about the supported frameworks organization shows it is far from being finished. There are different candidates.

Wi-Fi: The chief clear-framework organization advancement opportunity for an IoT device is Wi-Fi because it is so inescapable. Wi-Fi can be a not too bad response for certain applications. Practically every house that has an Internet affiliation has a Wi-Fi switch. In any case, Wi-Fi needs a fair proportion of power [5]. There are load contraptions that can't deal with the expense of that level of force: battery worked devices, for example, or sensors arranged in regions that are difficult to control from the lattice.

Low-power solutions: The most current framework organization progresses consider insignificant exertion, low-power game plans. These advances reinforce the creation of amazingly colossal frameworks of minimal smart contraptions.

At present, major R&D attempts include low-power and gainful radios, allowing a significant drawn-out period of time of battery life.

Imperativeness gathering as a power hotspot for IoT contraptions works by sorting out for unattended long-stretch movement without human intervention (e.g., M2M frameworks). New application shows data masterminds that enable self-administering action.

6. *Security Challenges and Power Resource Constraints*

   When we talk about IoT-enabled agriculture, i.e., smart agriculture, the main issue or challenge that comes in one's mind is being prone to some unauthorized action which could be a complete hack on implemented system or it could be being monitored by third party which is not authorized legally to do so but had a complete monitoring on one's whole system, production, expenses, and production. This fear of being controlled by some unauthorized identity is most considerable and a challenging threat in order to implement a smart system in basic environment of agriculture. Whenever a rancher or farmer trust on a smart mechanism, he is always looking forward for system's confidentiality, complete authenticated mechanism, and privacy of its communication over provided service, i.e., decision and transaction logs, etc. A person is always looking for safety measures provided to him when he chooses machine over human. In technical terms the overall mechanisms should always be secure at every layer of OSI model, i.e., application layer and network layer. Service provider should ensure the client for safety against various attacks whether it is denial of service attack, distributed denial of service attack, session reply, phishing, malware injection, message modification, etc. All the logs and transaction should be secure with encryption algorithms, and hash keys should be powerful enough to handle various kinds of attack possibilities. One more challenge that comes into consideration is resource utilization rate and power backups; the system should not be vulnerable to power failures and should not be constrained to battery consumption rate, as if rate is high the system will more depend on power resources ([1], p. 1–12).

The very basic requirement one can demand in terms of security of its system is authorized access and complete control on accessing the system; the data transaction logs should be done in confidential manner with strong firewall for the system to make it strong enough to handle attacks and demolish those attacks. In smart agriculture the rancher has to deploy all smart devices like monitoring devices and sensing devices in open environment, and it may or may not be able to check the devices and monitoring logs of devices in a continuous manner; then in that case, how will the monitoring and servicing of those device be done or by whom? In this case rancher should get a kind of assurance from service provider to give him continuous assistance term of any human or machine intervention to take care of all these things. A kind of security issue could be explained with RFID security vulnerability as it is related to disclosure of data and transaction information which include important details like transaction ID, physical location access ID, or maybe passwords. With these kinds of threats, deployment of smart system with a single security protocol can never be strong enough to ensure security to rancher or farmer.

Solution

Distributed (P2P) is a system worldview that has use cases in keen cultivating correspondence. The gadget confirmation strategies in this kind of correspondence, in any case, depend vigorously on the open key foundation. Even though the framework is reliable, it puts pointless calculation load on asset.

Ranch compelled shrewd IoT gadgets partaking in a safe P2P correspondence. In like manner, creators in cf. Chavez et al. [9] proposed a lightweight gadget confirmation arrangement in which meeting keys and open keys are joined to assist the encryption/decoding undertakings. It brings about a quick and light-weight confirmation arrangement which is a solid match for brilliant cultivating correspondence purposes. Furthermore, West (cf. [9]) presented a structure to comprehend vulnerabilities in rising innovations and the utilization of such advancements in a keen farming-specific condition. The system's objective is to measure the degree to which the utilization of shrewd cultivating new advances is powerless against digital assaults. It utilizes the regular helplessness scoring framework (CVSS) for the risk forecast model appraisal. The work shows the trade-offs between innovation development and adjustment in the brilliant cultivating condition which can prompt a framework bargain. The methodology in the paper utilizes three parameters: essential parameters, transient parameters, and natural parameters for developing a

CVSS score ([1], p. 1–12). Essential parameters show the natural and seriousness of helplessness, while transient parameters show how defenselessness may change and influence the framework after some time because of specialized changes. Ecological parameters mirror the particulars of helplessness present in a brilliant cultivating explicit condition. Even though the CVSS score has gotten a standard in the business for comprehension of the seriousness of vulnerabilities and organizing their patches, it has a few inadequacies. Shrewd cultivating is of various domain with many associated gadgets and frameworks. CVSS (cf. [9]) score bargains with singular vulnerabilities and neglects to precisely catch the effect of associations inside the whole framework.

References

1. T.S. Rosenstock, D. Rohrbach, A. Nowak, E. Girvetz, An introduction to the climate-smart Agriculture papers, in *The Climate-Smart Agriculture Papers*, pp. 1–12 (2018). https://doi.org/10.1007/978-3-319-92798-5_1
2. Action Aid, Clever name, losing game? How climate smart agriculture is sowing confusion in the food movement (2014). Available at: http://www.actionaid.org/publications/clever-name-losing-game-how-climate-smart-agriculture-sowingconfusion-food-movement
3. S. Asfaw et al., What determines farmers' adaptive capacity? Empirical evidence from Malawi. Food Secur. **8**, 643–664 (2016). Available at: http://link.springer.com/10.1007/s12571-016-0571-0
4. D.N. Barton et al., Assessing ecosystem services from multifunctional trees in pastures using Bayesian belief networks. Ecosyst. Serv. **18**, 165–174 (2016). https://doi.org/10.1016/j.ecoser.2016.03.002
5. B. Bett et al., Effects of climate change on the occurrence and distribution of livestock diseases. Prev. Vet. Med. **137**(November 2015), 119–129 (2017). https://doi.org/10.1016/j.prevetmed.2016.11.019
6. E. Bryan et al., Can agriculture support climate change adaptation, greenhouse gas mitigation and rural livelihoods? Insights from Kenya. Clim. Chang. **118**, 151–165 (2013). Available at: http://www.springerlink.com/index/10.1007/s10584-012-0640-0. Accessed 5 Mar 2013
7. CCAFS & GCF, *Sectoral Guidance for GCF Investments in Agriculture* (CCAFS, Netherlands, 2018). In press
8. A.J. Challinor et al., Current warming will reduce yields unless maize breeding and seed systems adapt immediately. Nat. Clim. Chang. **6**, 954–958 (2016). http://www.nature.com/doifinder/10.1038/nclimate3061
9. E. Chavez et al., An end-to-end assessment of extreme weather impacts on food security. Nat. Clim. Chang. **5**(11), 997–1001 (2015). Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84945295462&partnerID=40&md5=05582dc08e918c2caf26b887d45ece3a
10. K.E. Giller et al., Conservation agriculture and smallholder farming in Africa: The heretics' view. Field. Crop Res. **114**(1), 23–34 (2009)
11. R. Gledhill, C. Herweijer, D. Hamza-Goodacre, MRV and data management for CSA mitigation. Climate-smart agriculture in Sub-Saharan Africa project (2012)
12. S. Graefe et al., Climate-smart livestock systems: An assessment of carbon stocks and GHG emissions in Nicaragua. PLoS One, 1–19 (2016). https://doi.org/10.1371/journal.pone.0167949

13. B.B. Lin, Agroforestry management as an adaptive strategy against potential microclimate extremes in coffee agriculture. Agric. For. Meteorol. **144**(1–2), 85–94 (2007). Available at: http://linkinghub.elsevier.com/retrieve/pii/S0168192307000548. Accessed 25 Oct 2012
14. P.A. Minang, et al., Climate-smart landscapes: Multifunctionality in practice (2014). Available at: https://books.google.com/books?id=rii-BQAAQBAJ&pgis=1
15. K. Paustian et al., Climate-smart soils. Nature **532**(7597), 49–57 (2016). Available at: http://www.nature.com/doifinder/10.1038/nature17174
16. S. Sala, F. Rossi, S. David (eds.), *Supporting Agricultural Extension Towards Climate-Smart Agriculture: An Overview of Existing Tools* (Global Alliance for Climate Smart Agriculture (GASCA)/FAO, Italy, 2016). Available at: http://www.fao.org/3/a-bl361e.pdf
17. C. Thierfelder et al., How climate-smart is conservation agriculture (CA)? – Its potential to deliver on adaptation, mitigation and productivity on smallholder farms in southern Africa. Food Secur. **9**, 537–560 (2017). https://doi.org/10.1007/s12571-017-0665-3
18. P.K. Thornton, M. Herrero, Potential for reduced methane and carbon dioxide emissions from livestock and pasture management in the tropics. Proc. Natl. Acad. Sci. U S A **107**(46), 19667–19672 (2010). Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2993410&tool=pmcentrez&rendertype=abstract. Accessed 12 Nov 2012
19. World Business Council for Sustainable Development (WBCSD), WBCSD climate-smart agriculture action plan. Midtermreport (2015). Available at: http://docs.wbcsd.org/2017/11/WBCSD_Climate_Smart_Agriculture-Action_Plan_2020MidTermReport.pdf. Accessed 14 Nov 2017

Implementation of Women's Self-Security System Using IoT-Based Device

Sonia Das, M. R. Pooja, and K. S. Anusha

1 Introduction

In the world, the major issue for most girls is thier protection from abuse. Despite the fact that technology is evolving and many new devices are being developed, women are still facing problems [1, 2]. Crimes against women are still rising and many women have been killed and/or harassed. There is a strict law in our country, yet a number of cases continue to be reported in police stations every day. Keeping this in mind, we have built a system for women's safety [3, 4]. The basic idea is to convey emergency alert messages involving location to parents and near by police stations, so as to avoid the incidents [5].

The device has different sensors like Alarm Buzzer, sensors for measuring pulse rate, Global Positioning system (GPS) and Global system for mobile communication (GSM). All sensors are concerned with Raspberry PI, which is a controller [6–8]. Women can make use of an alert button, when they feel unsafe. When it has been activated, the buzzer will make a sound to grab the attention of people those who are near to the victim and will send out an emergency SMS and location to predefined contacts, which is already programmed in [9].

The message, which will be received by the parents, will enable them to view the location. The device works automatically, when pulse rate meets the threshold value [10, 11]. Python programming language is used to programme these sensors. The goal is to create an all-in-one security system based on IoT [12–15]. In our work, when the intruder tries to do something, the device will produce a shock and a pepper spray will also be activated; so that women can easily escape from the critical situation. The device produces two self-defence mechanisms in order to save

S. Das (✉) · M. R. Pooja (✉) · K. S. Anusha
Vidyavardhaka College of Engineering, Mysuru, Karnataka, India
e-mail: pooja.mr@vvce.ac.in; anushaks@vvce.ac.in

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_5

women, and also, the receiver shall be able to view a live video. If one fails, the other self-defence mechanism will save the women. This enables us to catch the intruder with proof and also can save the women through the self-defence mechanism.

2 Existing System

In India, the status of women has changed a lot in the last few years. They are competing equally with men, so the history of women has been eventful. Though they are equally strong, a certain percentage of women are assaulted and murdered every day. Some of the women are suffering mentally as well as physically. So, in this regard, people have come up with creative ideas. Various mobile phone applications are as follows:

[15] B.Vijaylakshmi et al. built an app. When women sense danger, they touch the app, which will record the audio for some seconds and transmit along with the location, or only an alert message.

[16] Niti shree, the author, designed a system that tracks women all the time. This mobile app tracks only the vehicle which the women use.

[8] Swapnali N.Gadhave et al. have built an app which shows the safety score based on the area the person is travelling in. That will be known as safe or dangerous area. When she goes through an unsafe area, it alerts the user with a warning message. The user can take a friend or family with her. The app also provides an alternative route to reach destination safely.

[17] Hung Nguyen et al. have created a mobile app that may be used to alert people in danger by tapping a button twice. This process initiates and sends a location or a message of a vehicle. This cycle will be repeated every two minutes. It is only for tracking the vehicle.

[18] Shubham Sharma et al. built a device. When women are in danger, they have to press and hold on the trigger button, and the system will be activated and send an SMS and location to parents with the use of a microcontroller.

[19] Ms. Deepali M et al. build a device that senses the danger upon which a buzzer may be pressed that would send a warning message to the parents as well as a location to a local police station. Some of the sensors will collect the data, if appears with irregular values then that information will also be sent.

[20] Prof. R.A. Jain et al. built a device, which takes and sends the photo of the offender to the parents and the nearest police station along with SMS and location.

[21] Orlando Arias et al. have built a smart watch. A woman when in danger can press the button, which sends the status of her to the defined contacts along with the message and address.

[22] J. H. Ziegeldorf et al. They created a smart watch and utilized it to store a few contacts. The device was activated when SOS button is pressed for 2 seconds. It then sends out an emergency message like “I am in Danger” with beep sound

to the receiver and further shares the user's location. It continuously records the oxygen level and also sends out an intruder Image.

Major Drawbacks of the Existing System

(i) Some of the systems require human intervention.
(ii) The mobile phone can be dropped at some point of time and the app may not work thereafter.
(iii) When they press the buzzer, only at that time alert message will be sent. At certain times, she may not be able to activate the button.
(iv) The system does not provide defence application for women to escape from the critical situation.
(v) The system doesn't provide complete solution to the existing problem.
(vi) The system lacks smart features like capturing and live video of culprit.
(vii) Some of the data is not useful, like every 2 minutes, the process will be repeated.

3 Proposed System

The harassment of women is still increasing, even though technologies have evolved. Because there was no self-defence mechanism to escape from that critical condition, women had to wait till parents and police reach the spot. So keeping this in mind, in the system here we produce two self-defence mechanisms, one is pepper spray and another one is shock circuit. When one fails, the other defence mechanism will work to help a woman escape from that situation. The block diagram of the proposed system is as shown in Fig. 1.

(a) **Battery:** To power the circuit, a rechargeable battery of 12v will be used.
(b) **Raspberry PI:** It is a small PC. It is integrated with the pulse rate sensors, GPS and camera modules. The emergency message will be sent automatically to pre-defined contacts. With the use of an SD card, it can store all the information concerned.
(c) **GPS module:** It is a location tracker which will track the current location if the usesr. The information is used to know the exact location like street, area or nearby junction and that will facilitate to reach the location. If GPS is not working, then it will send longitude and latitude co-ordinates through email or SMS.
(d) **GSM module:** It is used for the communication between computer and GPRS system. The Sim card is inserted within the mobile phones. To receive and send the SMS, the GSM sim card should be registered to the system. It is known as Machine to Machine communication.
(e) **Buzzer:** The buzzer is the device that produces sound in the opposite direction. This is beneficial in two ways: Firstly, it is utilized to warn those who are related to the victim, and secondly, it turns on the device by delivering an alert message and its position to the phone numbers specified.

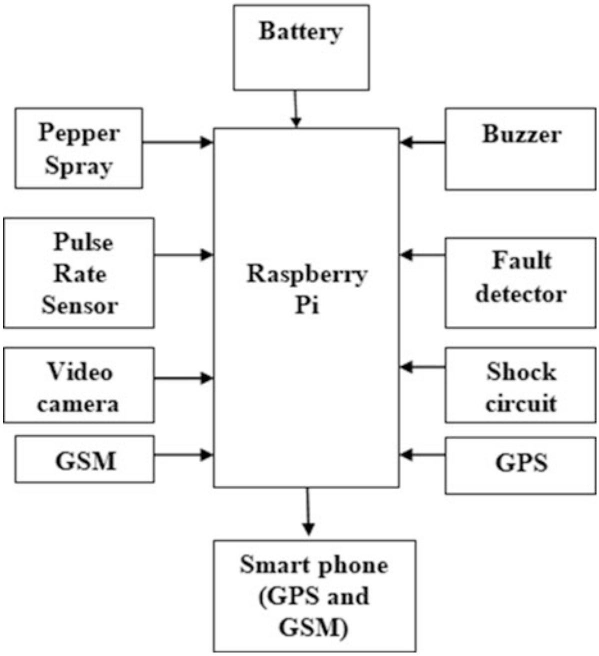


Fig. 1 Block diagram

(f) **Pepper spray:** Pepper spray helps in self-defence mechanism, where, in a critical situation, it may be sprayed on the intruder's face and thus help to escape from that situation immediately.

(g) **Shock circuit:** It is another self-defence mechanism. If one fails, the other will help to escape from that situation. The shock circuit will help by giving shock to the intruder in the critical condition.

(h) **Video camera:** While women sense danger, SMS and location will be sent, and along with it, the receiver can view the live video and can take immediate action. The live video helps us to collect the proof on the intruder, rather than having to explain to the police.

(i) **Pulse rate sensor:** This sensor will check the pulse continuously; if it reaches irregular values, then the device is automatically activated and sends out an alert SMS and location.

(j) **Fault detector:** This will help us to find the fault in the system. If the system has some fault, it will alert us through some messages.

(k) **IOT module:** It is a network of physical devices, things and alternate things like sensors and actuators. It allows us to gather the information and also to exchange the information. It also allows objects to be sensed and remotely controls the things without human intervention.

Applications of Proposed System

(i) Used for the safety of women.
(ii) Used as evidence to punish with an accurate documentation.
(iii) Used for the purpose of kid surveillance.
(iv) Used for the system of vehicle tracking.
(v) Used to keep handicapped individuals safe.
(vi) Used for aged persons.

4 Methodology

The system is functioning in two scenarios: when a woman feels unsafe, she will press the buzzer, and when she does not feel unsafe, she will not touch the buzzer, or when she is unable to press the buzzer for some reason, the other procedure is used. It detects the pulse rate and automatically begins the system – it operates the same way when the buzzer is pressed. The working of both the scenarios is as follows:

First Scenario The device gets activated by tapping an emergency button. The IOT device, which will retrieve the location by GPS, will be updated in Raspberry PI and that will initiate and send message to the pre-defined contacts through GSM [23, 24]. With the alert message an option will be provided for to view the live video.

Second Scenario If pulse rate sensor reaches an irregular value, then the device automatically gets activated. The irregular values will be sent to raspberry PI, the raspberry PI will then send the alert message to pre-defined contacts. The system model and the flow diagram are as shown in Figs. 2 and 3, respectively.

Women's Security System's Workflow

Step 1: Start.
Step 2: Emergency button is pressed [Buzzer].
Step 3: When GPS detects a signal, it begins calculating the user's current values and sends the location via SMS to pre-defined contacts using GSM. Live video can also be viewed.
Step 4: If pulse rate reaches irregular values, then it automatically sends the SMS and location to the contacts. Live video can also be viewed by the receiver.
Step 5: Pepper Spray is turned ON, to spray the pepper on the intruder.
Step 6: Shock circuit is turned ON, to apply shock to the attacker.
Step 7: The buzzer has been activated to alert the individuals in the area.
Step 8: Stop.

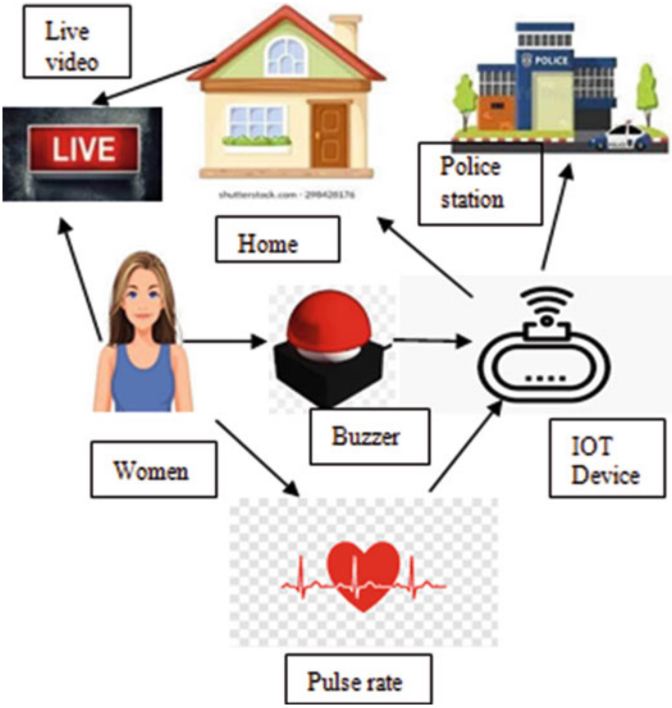


Fig. 2 System model

Advantages of the Proposed System

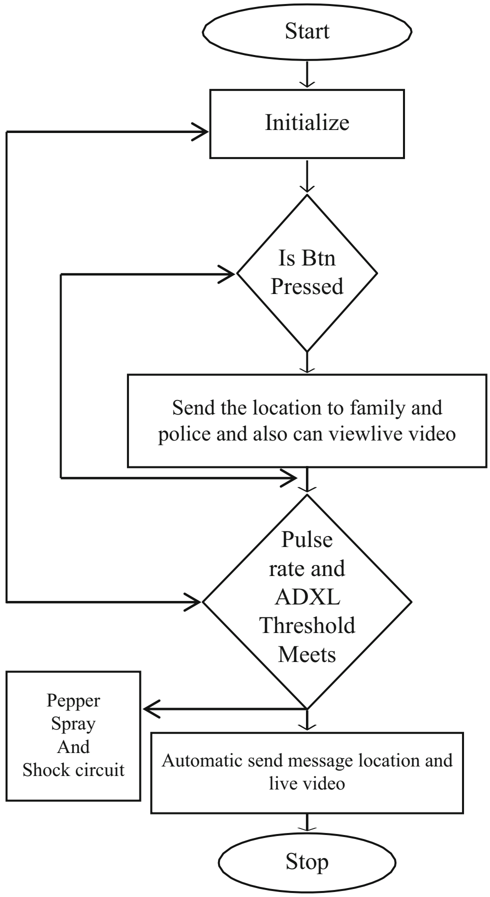
(i) Women can go anywhere with security device.
(ii) Intruders will be found easily with proof.
(iii) An automated system without human intervention.
(iv) Easy and convenient to operate.
(v) A low-cost device with greater performance.
(vi) Data with greater accuracy and quick response in emergency.
(vii) It is environment-friendly.
(viii) Detects fault in the system.

5 Result and Discussion

The purpose is to provide safety for women in critical situations. The system has switches 1 and 2; when the device is on, it displays both the switches.

When the first switch is activated, it sends the alert message, pulse rate, temperature and humidity to the assigned number along with the location. When she feels unsafe, the buzzer will be pressed. Once it gets activated the Raspberry PI

Fig. 3 Flow diagram



gets the commands and it calculates the current values of the user. The calculated values will be sent through SMS, in which raspberry PI stores the values and then sends it through message as in Fig. 4. The message is directed to the pre-defined contacts. The GPS will track the location as in Fig. 5 and update it. Through buzzer sound, the nearby people can save her.

When switch 2 is activated, it makes a live video as shown in Fig. 6 and sprays the pepper spray on the attacker, as in Fig. 7. The woman can thus be saved through the IOT device. If intruder attacks woman or tries to damage the device, it senses the motion and produces shock for the aggressor; for analysis, a bulb is used which lights up on sensing any motion instead of shock producer as in Fig. 8. The proposed model is as shown in Fig. 9.

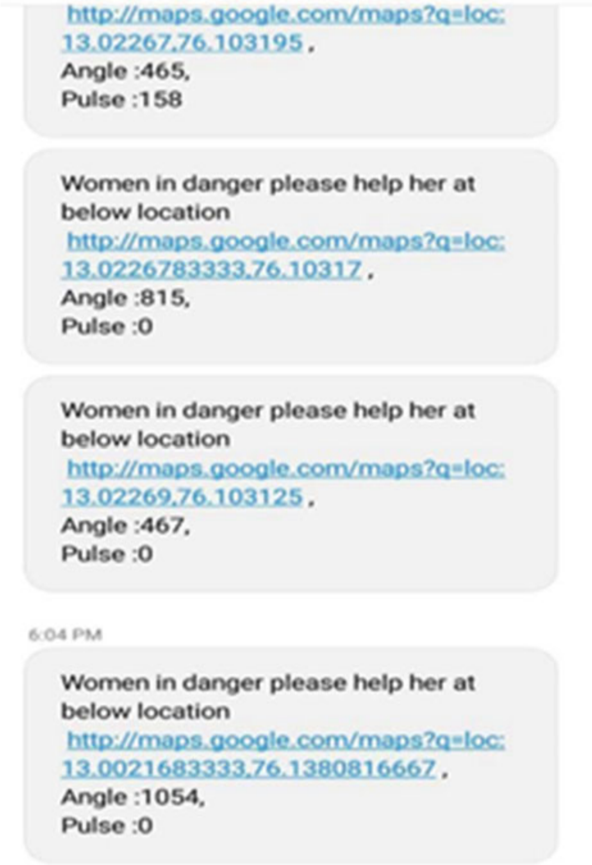
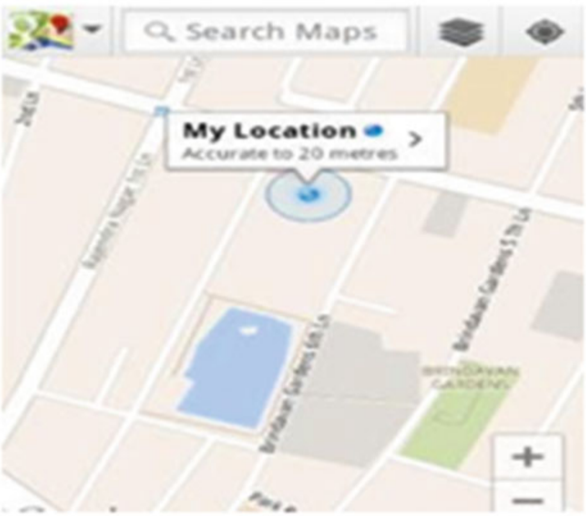


Fig. 4 Tracking the woman through message

Fig. 5 Woman's location



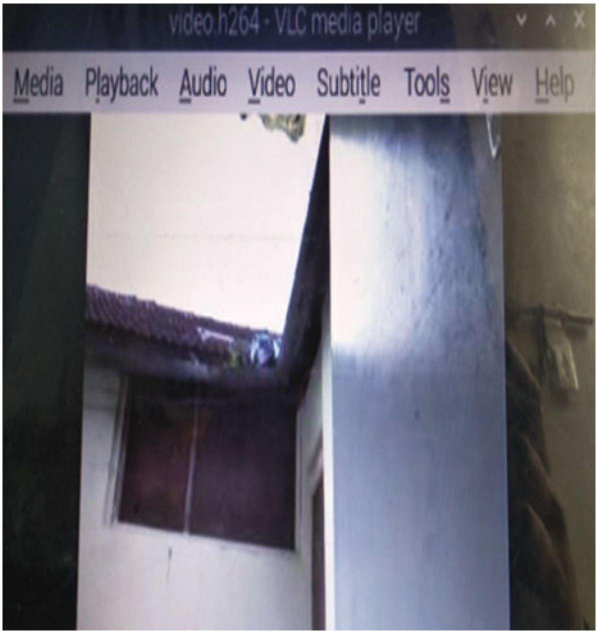


Fig. 6 View of live video

Fig. 7 View of pepper spray

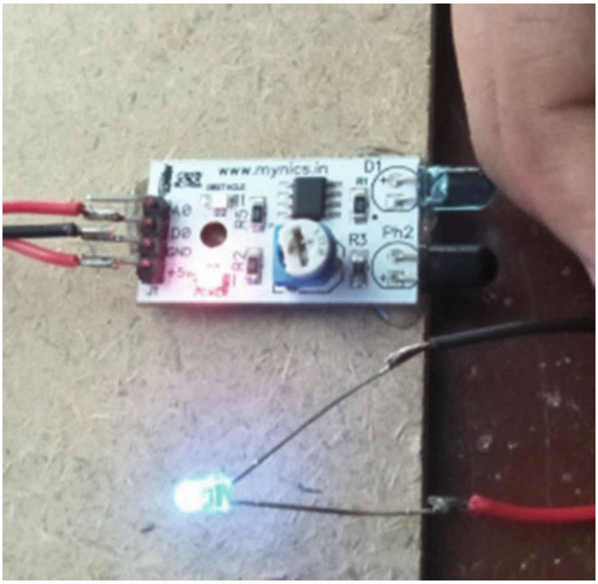


Fig. 8 View of shock LED bulb

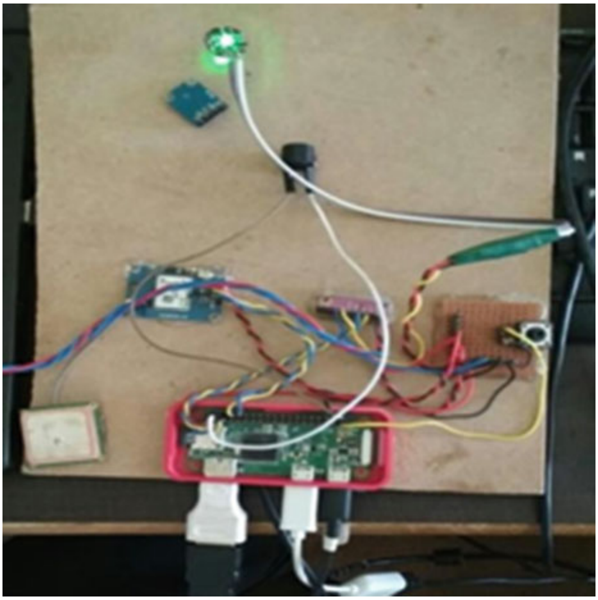


Fig. 9 Proposed model

6 Conclusion

This proposed system ensures protection and security for women. However, with the assistance of wireless technologies, with the secure medium, it would be possible to connect and send alarm messages to pre-specified numbers. The system will aid in the speeding up of the process of ensuring women's safety via the Geolocation tracking mechanism. The system provides security for women by detecting problems automatically, easing intimidations and sends help messages to the kin and to the police station that is nearby by means of IoT. Therefore, crimes of such nature can now be brought to an end with the help of the proposed system.

References

1. V. Krishnamurthy, S. Saranya, S. Srikanth, S. Modi, M-WPS: Mobile based women protection system, in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, (IEEE, 2017), pp. 1701–1706
2. H. Nagamma, IoT based smart security gadget for women's safety, in *2019 1st International Conference on Advances in Information Technology (ICAIT)*, (IEEE, 2019), pp. 348–352
3. Z. Ali, M.A. Khan, O.B. Samin, M. Mansoor, M. Omar, IoT based smart gloves for women safety, in *2021 International Conference on Innovative Computing (ICIC)*, (IEEE, 2021), pp. 1–6
4. D. Seth, A. Chowdhury, S. Ghosh, A hidden Markov model and internet of things hybrid based smart women safety device, in *2018 2nd International Conference on Power, Energy and Environment: Towards Smart Technology (ICEPE)*, (IEEE, 2018), pp. 1–9
5. H. Nguyen, K. Lebel, S. Bogard, E. Goubault, P. Boissy, C. Duval, Using inertial sensors to automatically detect and segment activities of daily living in people with Parkinson's disease. IEEE Trans. Neural Syst. Rehabil. Eng. **26**(1), 197–204 (2017)
6. S. Sharma, F. Ayaz, R. Sharma, D. Jain, B.E. Student, IoT based women safety device using ARM7. IJESC **7**(5), 11465–11466 (2017)
7. K. Venkatesh, S. Parthiban, P.S. Kumar, C.V. Kumar, IoT based unified approach for women safety alert using GSM, in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, (IEEE, 2021), pp. 388–392
8. S.N. Gadhave, S.D. Kale, S.N. Shinde, A.C. Bhosale, Electronic jacket for women safety. IRJET (2017)
9. V. Hyndavi, N.S. Nikhita, S. Rakesh, Smart wearable device for women safety using IoT, in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, (IEEE, 2020), pp. 459–463
10. O. Arias, J. Wurm, K. Hoang, Y. Jin, Privacy and security in internet of things and wearable devices. IEEE Trans. Multi-Scale Comput. Syst. **1**(2), 99–109 (2015)
11. R. Khan, N. Mahfuz, N. Nowshin, A novel approach of women safety assistant device with biometric verification in real scenario, in *2020 IEEE International Women in Engineering (WIE)*, (2020)
12. P. Podder, M. Mondal, S. Bharati, P.K. Paul, Review on the security threats of internet of things (2021). *arXiv preprint arXiv:2101.05614*
13. A. Mosenia, N.K. Jha, A comprehensive study of security of internet-of-things. IEEE Trans. Emerg. Top. Comput. **5**(4), 586–602 (2016)
14. W. Akram, M. Jain, C.S. Hemalatha, Design of a smart safety device for women using IoT. Procedia Comput. Sci. **165**, 656–662 (2019)

15. B. Vijaylashmi, S. Renuka, P. Chennur, S. Patil, Self defence system for women safety with location tracking and SMS alerting through Gsm network. IJRET: Int. J. Res. Eng. Technol. ISSN: 2319-1163 ISSN: 2321-7308
16. Niti shree, A review on IOT based smart GPS device for child and women safety, IJERGS, May–June 2018
17. H. Nguyen, K. Lebel, S. Bogard, E. Goubault, P. Boissy, C. Duval, Using inertial sensors to automatically detect and segment activities of daily living in people with Parkinson's disease. Trans. Neural Syst. Rehabil. Eng. **26**(1), 197 (2018)
18. S. Sharma, F. Ayaz, R. Sharma, D. Jain, IoT based women safety device using ARM7, IJESC, May 2017
19. D.M. Bhavale, P.S. Bhawale, T. Sasane, A.S. Bhawale, IOT based unified approach for women and children security using wireless and GPS. IJARCET (2018)
20. R.A. Jain, A. Patil, P. Nikam, S. More, S. Totewar, Women's safety using IOT. IRJET (2017)
21. O. Arias, J. Wurm, Y. Jin, Privacy and security in internet of things and wearable devices. IEEE Trans. Multi-scale Comput. Syst. **1**(2) (2018)
22. J.H. Ziegeldorf, O.G. Morchon, internet of things: Threats and challenges, Security Commun. Netw., vol. 7, no. 12, pp. 2728–27422, *Conference on Electrical and Computer Engineering (WIECON-ECE)* (IEEE, 2017), pp. 426–431
23. T. Sen, A. Dutta, S. Singh, V.N. Kumar, ProTecht–implementation of an IoT based 3–way women safety device, in *2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, (IEEE, 2019), pp. 1377–1384
24. G.C. Harikiran, K. Menasinkai, S. Shirol, Smart security solution for women based on Internet Of Things (IOT), in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, (IEEE, 2016), pp. 3551–3554

Evolution of Hadoop and Big Data Trends in Smart World

Neeta Awasthy and Nikhila Valivarthi

1 Introduction to Big Data

Big Data is data that is too massive, complicated, and dynamic to be captured, stored, managed, and analysed using traditional data methods [1]. We need this data to make informed decisions with real-time dynamic updates in a business so that the leaking losses are minimized and the business runs to its optimum. The ultimate vision is to move towards machines that discover and an Internet that learns.

2 The Three V's

There is no size to categorize 'Big Data'. To start with, it is imperative to understand thoroughly the three basic V's of Big Data:

- *Volume* – Today, there are organizations and societies producing terabytes of data in a day. In traditional analysis, we take a sample of data, whereas Big Data analytics is used when all data has to be processed. If we keep using the traditional tools, we will have to leave some data without analysing it because of such a large volume of data. This volume of unanalysed data would keep increasing with the increase in data size, thereby influencing the decisions and policies [2]. Many organizations are stuck to this. There is so much data on the databases, but, there is no comprehension. This is because the data is not being

N. Awasthy (✉)
AKTU, Lucknow, India

N. Valivarthi
Waterloo, Waterloo, ON, Canada

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_6

analysed. This is pointing to a requirement, i.e. the systems are not able to handle that much data. Hence, some up-gradation is required in the organization.

- *Velocity* – If we are unable to analyse data at a speed faster than its creation and inflow, then the whole idea of Big Data Analytics is forfeited. The speed with which a company can analyse data and the pace at which data gets generated has to be competitive. Cloud Computing has increased five times in the last 5 years. Due to inexpensive storage and powerful computing, Moore's Law has become a thing of the past where it's concerned. The data explosion has taken an exponential vertical curve. It is astonishing to note that 90% of the data has been created in the last 2 years. To have a snapshot, let us bullet a few points:

  – The world is connected with one billion Facebook accounts, 200 million tweeters, 160 million Linkedin connections, 10.4 million Pinterest accounts, and many more active users in various social media platforms.
  – In one second, there are 8963 tweets, 989 Instagram photos, 1696 Tumblr posts, 4630 Skype calls, 92,204 GB of Internet traffic, 82,659 Google Searches, 83,943 YouTube videos viewed, and 2,908,827 emails sent as one date. Well, we cannot forget 54,977 Facebook posts and 890,740 WhatsApp messages in one second.
  – All this consumes 375 Megabytes of Data every day. Add Netflix, Prime, all entertainment, and gaming subscriptions to it. It makes an unending list. Now add to it the connected devices.
  – There are Wearable Devices, Smart homes, Industrial Internet, Connected Cars, Agriculture, Smart Retails, Energy Engagement, Smart Cities, Healthcare . . . to name a few.
  – In the year 1990, the number of connected devices was 0.3 million. In 1999, it grew to 90 million. In 2010, it rose to 5 billion, escalating to 9 billion in 2013 and expected to exponentially increase to 1 trillion by 2025.

- *Variety* – Though the volume and velocity of generation of data cannot be denied to be the two most important aspects, actually it is Variety that makes the traditional tools obsolete. Traditional tools are designed to work best with structured data. Obviously, the data has to be fixed in a given format before any possible analysis. Big data can be classified into basic three categories so far as the variety is concerned, i.e. structured data, semi-structured data, and unstructured data.

  – Structured data resides in a relational database or sometimes in a spreadsheet. Its application can be railway and airline reservation systems, inventory control, and other ERP. The likes, shares, downloads from a social media site are structured data.
  – Semi-structured data has no fixed format. It can have some loose format, but not all data might fit in there. This is because different data objects may have different attributes that are not known in advance. For that, we generally leave a 'Remark' column in the database while dealing with structured data. Its structure is implicit, irregular, nested, and heterogeneous. For example, web

pages, information integration, and XML files. Extensible Markup language files, or XML files, describe the transportation, structure, and storage of files. Really Simple Syndication or RSS feeds are an example of XML files. RSS feeds may be an update to a favourite blog. It can be a software upgrade or a feed from a favourite news channel. All these form the semi-structured dataset.

– Unstructured data can be text, documentation, audios, videos, and surveillance imagery. The flood of data coming from emails, customer comments, social media forums, customer journeys on websites, and call-centres are unstructured in nature. It resides in applications, such as NoSQL database, data warehouse, and data lakes.

3 Challenges with Big Data

Following are the challenges while dealing with big data:

1. It takes an enormous amount of time to process this huge and exponentially growing dataset if processed on a single unit.
2. There is difficulty in query building when the data is having a complex structure, i.e. it is semi-structured and/or unstructured. Also, in the case of a long query, the last step error wastes and expects to reiterate the whole process.
3. Finally, a large amount of data comes to the computational unit with such high processing capacity. Thereby making a huge capital investment.

The problem is challenging because the data is semi-structured or unstructured coming from all types of sources. It becomes even more complicated when a sample of data is not enough, meaning thereby, all data needs to be processed. It becomes the most challenging due to the last mile errors, wherein all the data needs to be reprocessed due to some error at the last operation. This reiteration creates a huge pressure on all the resources.

To solve this problem, the obvious solution seems to be dividing the data into small chunks in racks and then processing them in parallel. Also, divide the jobs in the form of Mapping them and then pointing them by reduction process to find the winner.

The technology required for such storage, access to clean data, its parallel processing, and indication for decisions is massive [3]. The design of the solution needs to have failure tolerance, which is possible only with redundancy. This is why multiple copies of each data block are created in different racks. While the traditional systems are designed for some scales, the big data solutions are designed to add or remove the data racks without any problem of scalability or flexibility. The system is cost-effective as one needs to pay-per-use as commodity hardware, and not bear the cost of expensive servers and failure-resistant disks [4]. Nowadays, not only data but the task handler to query and analyse is also available over the cloud on a ‘pay-per-use’ basis. This is what exactly Hadoop does. It makes three copies of data in two separate tracks on a pay-per-use basis stored in cloud servers maintained by third parties. In such a scenario, the business has to concentrate only on "business" and not on IT infrastructure and its maintenance.

Hadoop has become almost synonymous with Big Data, leading to social analytics and Algorithmic Approach to Business. From here, the need starts for a good storyboarding and good storytelling. A good storyboarding is asking intelligent questions which create the required business insights. A good storytelling will land up in making informed business choices or Algorithmic Business. The six-prong approach to this is based on human, machine, product, platform, organization, and crowd.

In order to Learn, Reason, Respond, and Interact with such data in hand and expanding in an explosive manner to address new areas like Health, Contactless personal assistants, and robots; we need to have a file system like Hadoop.

4 Hadoop Architecture: HDFS and Map-Reduce

Apache Hadoop is a collection of open-source software utilities that provides a software framework that is designed with the objective of having less expensive Data Storage that would allow the firms in carefully storing, efficient handling, and analysing the data [5]. Hadoop is designed with *Master–slave Architecture* for data storage and processing. This architecture provides a well-grounded, flexible, and distributed computing framework.

Let us understand some of the terminologies present in the Hadoop architecture (Fig. 1).

- *Data Node*: Data node is used for storing user's data from the file system into the HDFS by providing services like write and read requests from the clients. The file gets split into one or more blocks and these blocks are stored inside the data node.
- *Name Node*: Name Node is considered as the centrepiece in the Hadoop file system which is also known as Master. It manages the metadata of HDFS like namespace information, block information, etc., hence it is configured with a lot of memory because metadata is held in RAM.
- *Secondary Name Node*: The main responsibility of the secondary Name Node is to take checkpoints of the metadata present in the file system on the Name Node. It is also responsible for merging edit logs with a file system image (fsimage) from the Name Node for better functioning.
- *Job Tracker*: Job Tracker is responsible for resource management, tracking resource availability, and tracking the progress of fault tolerance.

  The Job Tracker interacts with the Name Node to determine the location of data. It finds the task tracker nodes to perform the tasks on the given nodes. It also monitors the execution of MapReduce from local to the Slave node.

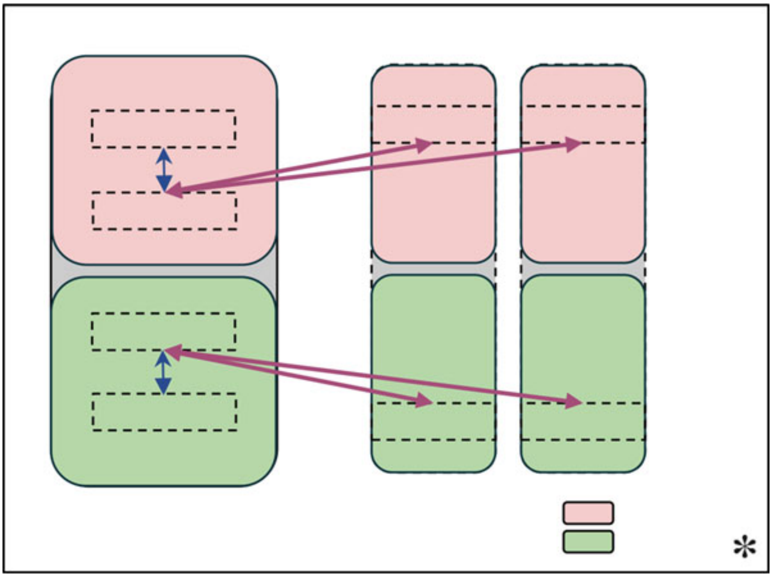


Fig. 1 Basic Hadoop architecture

- *Task Tracker*: Task Tracker runs the tasks and reports the status of tasks to the JobTracker. TaskTracker runs on Data Nodes. It should follow the orders of the Job Tracker and update the Job Tracker with its progress status periodically.

  The illustration of Hadoop architecture can be shown as the following:

5 HDFS

Hadoop Distributed File System (HDFS) is the main data storage system used by Hadoop applications. It utilizes a Data Node and Name Node (master–slave) architecture to execute a distributed file system that allows high-performance access to data across highly reliable and scalable Hadoop clusters.

Hadoop architecture basically consists of two main layers:

1. HDFS Layer (for storing data)
2. Map Reduce Layer (for processing data)

5.1 Features of HDFS

The main features of HDFS are as follows:

- Replication – Data loss is reduced by replicating each data block in two or more different machines.
- High Availability – We can access the data even if there is a data loss in one machine through replication.
- Fault Tolerance – It is the ability of a system to work without any interruption irrespective of the failures.
- Reliability – It can store data ranging up to 100s of petabytes with high performance in a cluster.
- Scalability – It can successfully handle growing amounts of data by adding nodes.

These features ensure zero leakage or loss of data, anywhere access of data, easy scaling up and scaling down business enterprise, and reliability of the system.

5.2 Storage in HDFS

The Hadoop Distributed File System employs a master–slave design, with the master serving as the Name Node and the slaves serving as data nodes. A single Name Node manages the file system metadata and controls user application access to the files in an HDFS cluster, and multiple data nodes store the data (in hundreds or thousands). Each data node is in charge of managing file storage and storage devices. When a file is loaded, HDFS splits it into one or more blocks, each of which is 64MB in size by default, but this can be changed at the cluster level. Data Nodes are a collection of slaves that store these blocks. To increase fault tolerance, various copies of each data node are stored according to the replication factor, which can be configured at the cluster level or at file creation [6].

Users connect with the Name Node to get metadata; data is sent between the client and the data nodes. In the actual data transport, the Name Node is not engaged. If a programme wishes to write a file to HDFS, it must first present the file's information to the HDFS Name Node. Based on the size of the block, file, and replication factor, the Name Node responds with data. The number of blocks in the file, the replication factor, and the data nodes where each block will be stored are all contained in the information provided by the Name Node. Users or apps divide the files into several blocks and begin organizing the data nodes based on the information gathered from the Name Node. Because of the write affinity, the initial replica is usually written to the data node that creates the file in order to increase write speed. Pipelining, HDFS write, and acknowledgement are the three steps of the procedure (Fig. 2).

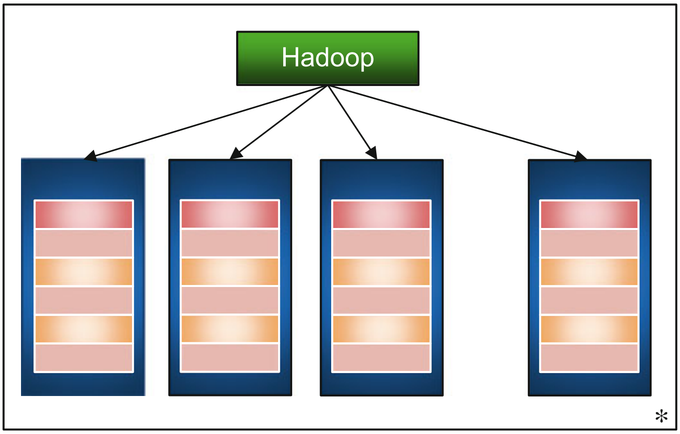


Fig. 2 Illustration of storage of data blocks in HDFS

6 MapReduce

With the turn of the century, Google implemented a programming model called MapReduce. Typically, MapReduce can process many terabytes of data on thousands of machines simultaneously at different locations. MapReduce has the ability to process more than 20,000 Petabytes of data each day. Google processed these Map-Reduce operations in Google File System. Unfortunately, this special file system was not open source.

Google ran these MapReduce operations on a special file system called Google File System (GFS) [7]. Now, it was the time for open source software with creative commons license instead of copyright solutions. This task was taken up by Doug Cutting along with Yahoo. They reverse-engineered the GFS model and made a Hadoop Distributed File System (HDFS). Hence, the software or framework that supports HDFS and MapReduce is known as Hadoop. Hadoop is distributed by Apache.

We need to have hundreds and thousands of CPUs to process petabytes of data produced by different Internet of Things devices. To solve this, a MapReduce programming model was introduced. MapReduce programme provides automatic parallelization and distribution. Each set of MapReduce has an input and output function; which are the key-values pair. The programmer specifies two functions which, i.e. Map, that produces and processes input key-value pairs, and it produces a set of intermediate pairs. The reduce function combines all intermediate values for a particular key and it produces a set of input values which is usually one. To define MapReduce, here is an example of a reality show in which we have to identify the best singer from the entire country. We do not call all the singers of the entire country in one place. But we divide them into different zones and find the winner of different zones (This is the Map process). Then we call all the winners of different zones into the finale and we find the best singer (that's finding the zonal winner out of key-value pairs and reducing the same to output). This is exactly what MapReduce does to us.

Extend this process to the Parallel Execution and we have the output for Task 1 to Task n ready at the same time. This process divides the tasks to granular level, i.e. very small tasks, and fulfils the following:

(a) Minimized time for fault recovery
(b) Graceful Degradation
(c) Better dynamic Load Balancing and
(d) Pipeline shuffling with map execution

Typically, there is a ratio of 40:1 for maps: reduce machines. On the fail of Map and Reduce, the failure of the worker machine is checked by Detecting failure by periodic checks, Map Task in progress re-executed, Reduce Tasks in progress re-executed, and The Task Completion is committed through the master.

It did not have the ability to handle the Master failure but it was said that the Master Machine of Google was not likely to fail. The process was so robust that during testing, 1600 out of 1800 machines (each with 4 GB memory, 160 GB IDE disk, and 2GHz Xeons processors) were turned down in one test, but still, the process finished 'Fine'. This was the performance testing. This infrastructure could scan and sort 10^{10} records of 100 bytes each. This way, a maximum of 24 MapReduce operations were tested.

When we say 'workers', we mean 'machines'. The slow workers typically lengthen the process completion time. The workers were slow due to other jobs running parallel, there was low data transfer due to soft errors on faulty disks, Process cache disables, or RAM not efficient, etc.

The solution was to distribute the job to different sets, and whichever finishes 'first' is called a winner. The backup copies were created at the near end of the phase. This correction on Redundancy execution shortened the time dramatically. To improve upon the Locality Optimization, the GFS asks for locations of replicas in input file blocks. All the tasks are divided into 64 MB each, which is actually the GFS block size. Map and reduce functions sometimes fail for typical inputs; though in such cases, the best solution is to fix the bug, it may not be possible for all cases. In such cases, the bugs are handled in third party libraries. Other characteristics of Map Reduce system are:

(a) Sorting is guaranteed within each reduced partition
(b) Use of combiner for saving network bandwidth
(c) Counters defined by users
(d) Compression of Intermediate data
(e) Local testing and debugging

The MapReduce comprises two daemons:

1. Job Tracker (Master process)
2. Task Tracker (Slave process)

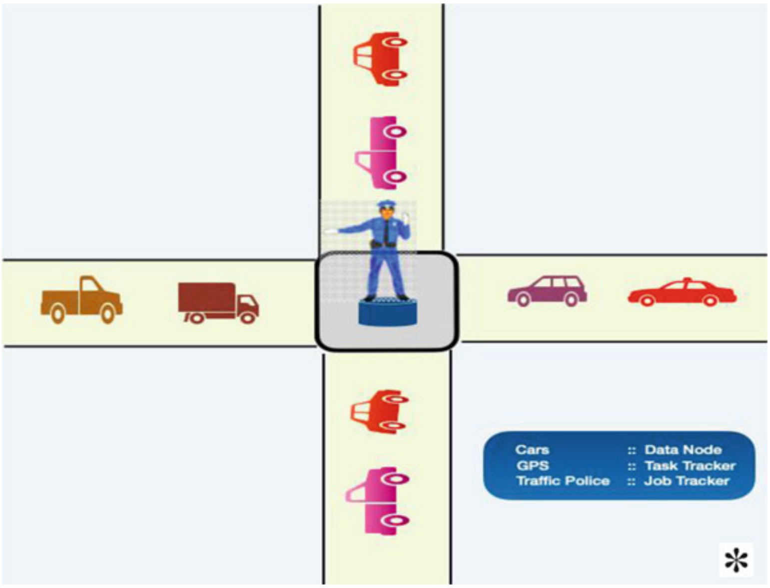


Fig. 3 Example of understanding MapReduce daemons

A single Job Tracker can be found in each cluster; however, many-task trackers can be found as well. In a small cluster, the Name Node and Job Tracker are usually on the same machine, whereas in a large cluster, the Name Node and Job Tracker are usually on separate machines. The job is divided or broken into tasks when a customer submits the job to the Job Tracker. The Job Tracker now determines which worker node should perform which job. Worker nodes are the slave nodes' counterparts. Task scheduling is the procedure for allocating jobs to worker nodes. The Job Tracker is also in charge of keeping track of all the tasks assigned to the worker nodes in the cluster and monitoring their health [8]. Clients and task trackers communicate using Job Tracker, which is used to make Remote Procedure Calls (RPC). RPC is a language that processors use to interact with one another. Because the Job Tracker keeps track of all jobs and their associated tasks in the main memory, the memory requirements for a Job Tracker are quite high, as the number of tasks varies from one job to the next (Fig. 3).

A task tracker is a slave process. The task tracker is in charge of completing all of the duties that the Job Tracker has allocated to them. Only one task tracker process is allowed to operate on each worker node. There are several maps and decrease spaces available in each Task Tracker. Task Slot is the name given to these maps and reduce slots, when combined. The number of map and red slots determines the number of map and reduce jobs that may be run at the same time. Through a mechanism known as the heartbeat, the task tracker occasionally reports the current health state and task progress. The Job Tracker considers the task tracker is dead if it does not get the task tracker's heartbeat for a certain amount of times.

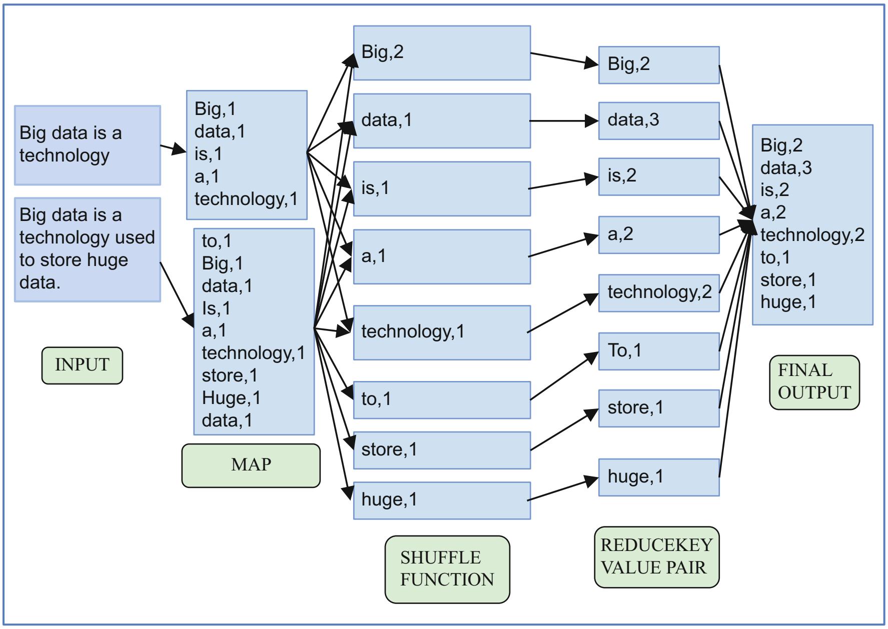


Fig. 4 MapReduce flow using wordcount example

MapReduce models are often run in three stages:

- Stage 1: map; stage 2: shuffle; stage 3: reduce

In Map Phase, the data from the directory or the file is taken and eventually stored in the HDFS. The input is pulled line by line by the mapper function. This data is processed and gets separated as small chunks of data, which are in <key, value> format.

The shuffle phase inputs the results from the mapper function and its tasks are to consolidate the relevant records. This function usually sorts and shuffles together all the grouping elements in the output obtained from the map phase. Reduce phase is the final task wherein we have to process the result from the shuffle phase, by combining values based on a function given by the user in order to produce a single output value (Fig. 4).

Let us consider an example of wordcount through Map Reduce. The functioning of each phase is performed as follows:

7 Hadoop Ecosystem and Components

Hadoop is an open-source framework. It is an ecosystem of open source projects that gives us the necessary framework to deal with big data. It stores data and uses commodity hardware to run applications. Thus Hadoop has the enormous processing power and gigantic storage for structured, semi-structured, and unstructured data [9]. Simultaneously handling an infinite number of jobs is the wonderful ability of Hadoop.

Following is the diagram exhibiting dependencies of Hadoop, deployed by various organizations. Each component of the Ecosystem has been developed to deliver an explicit function. Also included is their year of development (Fig. 5).

Hadoop Ecosystem is a group of tools that are used together by various companies in various domains for various tasks. Hadoop can only fulfil the requirements of storing and processing the data, but there are many other requirements such as clustering algorithms, recommendation systems, etc., where Hadoop fails to provide services [10]. The tools present in the Hadoop ecosystem help to resolve this issue. Let us discuss some of the main components in the Hadoop ecosystem.

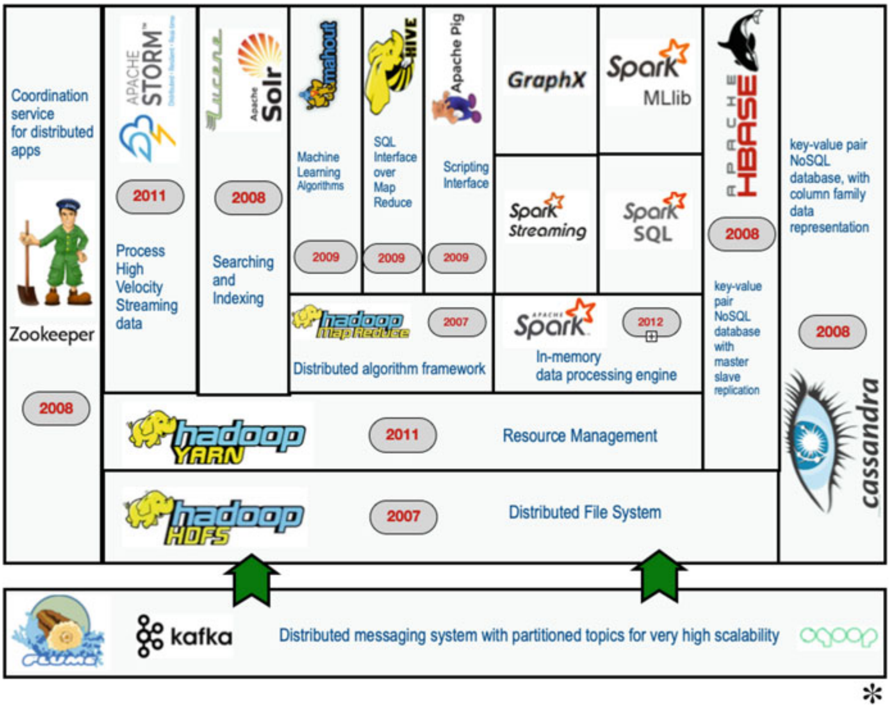


Fig. 5 Hadoop ecosystem and dependencies

7.1 YARN (Yet Another Resource Negotiator)

YARN is responsible for allocating resources to run particular tasks over the Hadoop cluster [11]. It performs all the processing activities for resource management. YARN has two components.

Resource Manager It is responsible for managing the resources and schedule applications on top of the YARN. It acts as a master daemon.

Node Manager It is responsible for managing containers and looking after resource utilization in each data node. It acts as a slave daemon.

7.2 Pig

Apache pig is a data processing tool that runs over Hadoop. It was developed by Yahoo in 2009. Apache Pig has its own instructional language which is called Pig Latin. Using Pig Latin, the life of the developers became very easy. They need to write the entire map-reduce program, but now the compiler converts the Pig Latin to MapReduce. It is said that approximately one line of Pig Latin is equal to 100 lines of map-reduce job. Pig provides a platform in building data flow for ETL operations (Extract, Transform, and Load).

Pig loads the data from the file system and then performs many functions on it like joining, sorting, filtering, grouping, etc. In the end, it also dumps the final data on the screen or stores in HDFS. Some of the PIG commands are:

- Data = LOAD 'desktop/pig-examples/sample.txt' AS (name:char array, score:int,class:int)
- filter_data = FILTER data BY score <= 75
- STORE filter_data INTO 'desktop/results'
- DUMP filter_data

7.3 Hive

In a distributed context, Hive employs an SQL-like interface to analyse data. Facebook has produced one of the most significant tools. Hive, like the pig, is used to write queries and run jobs on the Hadoop cluster, allowing Big Data to be leveraged.

Facebook realized that they are inundated with large amounts of data on a daily basis and that they require a specific mechanism to store, mine, and aid in data analysis. Hive was born from the concept of mining and analysing massive volumes of data. It's a Hive that allows Facebook to handle tens of terabytes of data on a regular basis. The following are some of the distinctions between pigs and bees (Table 1).

Table 1 Differences between Pig and Hive

Pig	Hive
It is a procedural language used for programming.	It is a declarative SQL, mostly used for reporting.
Pig doesn't have dedicated metadata of a database.	*Hive* defines tables beforehand and also stores schema information in a database.
It cannot perform partitioning.	It is very effective at performing partitions on data.
Pig is useful for ETL kind of workloads.	It is built on top of HDFS mainly for warehouse processing.

7.4 HBase

Apache HBase is a Hadoop-based open-source NoSQL database that handles all forms of data (structured, unstructured, and semi-structured) [12]. It's inspired by Google's BigTable. HBase may be used as a backend for a website to query in real-time, which is something that technologies like Pig, Hive, HDFS, and others can't accomplish. It provides a fault-tolerant method of sparse data storage. When used with Hive, it works well. Hbase is a column-oriented database, similar to a regular database, that consists of a set of tables with rows and columns. Any table should have a main key property, and every attempt to access the tables in HBase must utilize this primary key. For grouping together as column families, it accepts a variety of properties.

In this design, the master node manages the cluster, while the servers store table fragments before performing operations on the data. Let's say we have a table with billions of rows. If we need to find a certain entry in a large database, HBase is the best option, since query fetch time is reduced. HBase is most commonly used for analytical purposes.

7.5 Cassandra

Apache Cassandra is one of the most popular NoSQL databases which provides high availability with no single point of failure. The integration of Cassandra with Kafka (The message broker) and Spark makes it very reliable for real-time streaming, storing, and querying systems.

Some of the features in Cassandra are:

1. *High availability with fault-tolerance*: Each node is replicated as per the replication factor. Suppose if data is not found at a particular node, it could be found on the other node. There is no single point of failure.

2. *Reducing admin task*: Being masterless, each task can run independently without any requirement of engineers.
3. *Columnar databases*: It has a very high-level data model. Other databases where only metadata could be saved in columns, here actual data can be saved in columns. It is schemaless. So, there can exist multiple columns. Cassandra stores columns based on column names.
4. *Elastic scalability*: New nodes can be added or deleted in the cluster without any zero downtime. The server can be scaled up and scaled down without any disturbance. There is High throughput due to the availability of a larger number of nodes.

7.6 Sqoop and Flume

In order to capture data or load them into Hadoop, There are two tools called SQOOP and FLUME. *Sqoop (SQL-to-Hadoop):* As the name reflects, this tool is used to import data from RDBMS to HDFS. *Flume:* This tool is utilized to primarily insert semi-structured or unstructured data from streaming data sources to HDFS.

The main difference is how they bring the data into the Hadoop file system, the flume is a single process tool designed to get a change in data, whereas Sqoop is designed specifically to import data in bulk from RDBMS by distributing the task across the Hadoop cluster.

7.7 Apache Spark

Apache Spark is an open-source data processing tool used mainly for the purpose of storage and processing real-time data across various clusters of computers using simple programming constructs [13].

Spark supports different programming languages like Python, Java, R, and Scala.

As it is done in-memory, Spark generally processes data 100 times faster than MapReduce. It has optimized code and stands up for authentication via a shared secret (Fig. 6).

Some of the vital features of Apache Spark are as follows:

- Fast processing
- In-memory computing
- Flexible in nature
- Fault-tolerant

Apache Spark has five components:

- *Spark Core* is the brains of Apache Spark. Memory management, job scheduling, fault tolerance, storage system interfacing, and other critical functions are all

Fig. 6 Components of Spark

handled by it. The Scala, Java, and Python APIs make it simple to use the SparkCore capabilities.

- *Spark SQL* is responsible for providing SQL support to Spark and for streamlining the querying process on data stored in Spark's distributed datasets (RDD) as well as data from other external sources.
- *Spark Streaming* is a lightweight API that makes batch processing and real-time data streaming simple for developers.
- *Spark MLlib*: It's a scalable machine learning library that's easy to use, low-level, and works with Java, Scala, Python, and R.
- *GraphX*: It's a component for graph-parallel computations that's based on a branch of mathematics known as graph theory. On top of the Spark core, GraphX is a distributed graph-processing platform.

8 Hadoop 3.0 Upgradation

The upgradation of Hadoop 3.0 is a huge milestone in the line of Hadoop releases. Let us now discuss feature enhancements provided in Hadoop 3.x over Hadoop 2.x. Let us see the illustration first (Fig. 7).

- In order to access Hadoop 3.0, all the existing users must upgrade from Java 7 to Java 8. In Hadoop 3, the jar files are compiled targeting a version of Java 8 in runtime.
- Hadoop 3 has introduced the concept of execution type. At any particular moment, if there are no resources present, then these containers wait at the Node Manager. These Opportunistic containers have lower priority than Guaranteed containers. If supposed Guaranteed containers arrive in the middle of the execution of opportunistic containers then later gets pre-empted. This happens to make room for Guaranteed containers.

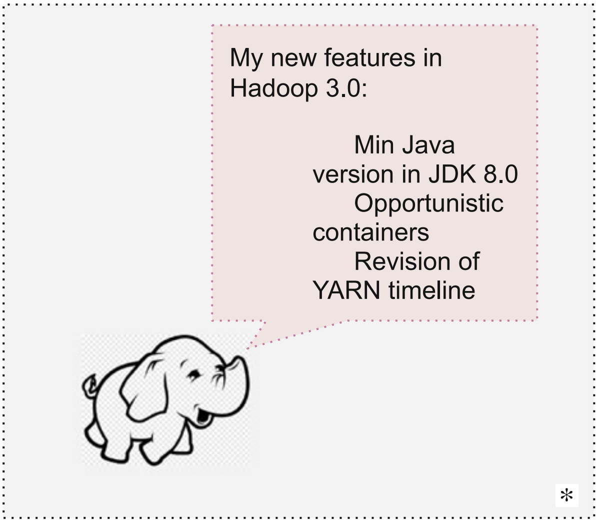


Fig. 7 Hadoop 3.0 upgradations

- Introduction to Erasure coding is a vital update, where HDFS can reduce its storage overhead nearly up to 50% when compared with replication of data blocks in different storages. This results in saving cost-storage bills on hardware expenditure. Erasure coding uses RAID (Redundant Array of Inexpensive Disks) which implements stripping, in which sequential data is divided and stored in consecutive units on different disks. For each strip, a certain number of parity cells are calculated and stored, which is called encoding. It provides the same level of fault-tolerance with less storage overhead.
- YARN timeline services are those which make use of a timeline collector to put data into its backend storage by running in a distributed manner across the cluster [14]. This feature gives RESTful API service to allow the customers to perform their queries.
- MapReduce Task-level optimization is done by adding a native implementation of map output collector. It provides above 30% speedups in the performance of shuffle-intensive jobs. Its basic motive is to add a Native Map Output Collector to handle key-value pairs that are resulted from a mapper, sort them, and serialize in native code [15].

Big Data is used in the industry massively. From large MNCs to small start-ups, whoever needs to work with big data is now moving to Hadoop. Some of the Top companies using Big Data Technologies are:

- *At Google:*
  - Index building for Google Search has come to doorsteps by having extremely large datasets and performing statistical analysis in indexing through large scale batch processing.
  - Google Cloud Dataproc is a Hadoop ecosystem in the version of Google. It contains the HDFS and Map/Reduce-processing framework which make the tasks easier [16].
  - Statistical machine translation.
- *At Facebook:*
  - The most popular social networking service provider is Facebook. There are two primary clusters in it. One of them is a 1100-node cluster with 8800 cores and 12 terabytes of storage.
  - A 300-node cluster with 2400 cores and 3 petabytes of storage is another option.
  - Facebook's Hive/Hadoop cluster contains over 2 PB of uncompressed data and loads 15 TB of data on a daily basis.
- *At eBay:*
  - At the moment, eBay is experimenting with Apache Spark, Storm, Kafka, and Hortonworks HDF, among other Big Data tools.
  - With 180 billion active users, the company's Hadoop clusters and Teradata installations generally handle 10 petabytes of raw data.
- *At Twitter:*
  - Because Twitter has such a large amount of data, storing and processing it is a difficult task. Hadoop is a technology for studying data that stores and processes large amounts of data.
  - Twitter's ranking algorithm has ingested a large amount of data, processed it using deep neural networks, and learned over time what content is relevant to each individual user.
- *At Spotify:*
  - Streaming music platforms like Spotify are using consumer interaction data to fine-tune their algorithms, improve user experiences, target audiences with ads, and make better-informed business decisions.
  - The data is gathered and stored in large databases, then processed for streaming according to the needs of the consumers.

9 Applications and Future

(a) Predictive Policing
(b) UID Project: Cradle to Grave
(c) Smart Meters

(d) Intelligent Transport system
(e) Defence Mechanism
(f) Intelligent Hiring

9.1 Predictive Policing

The fundamental underlying assumption of predictive policing is that crime is not distributed randomly, not in a Gaussian or Poissons distribution. Instead, it's a set of functional environmental factors that, at any given time, create vulnerabilities for victims and opportunities and spaces for criminals. Despite the fact that it is based on the "Near Repeat Theory." Researchers have discovered that if a burglary occurs in one house in a given neighbourhood, the other nearby houses become more vulnerable for the next 3–4 weeks.

Clearly, this is not an issue if data samples can be analysed and outcomes determined. Instead, all available data will be analysed to establish a pattern and algorithm to determine the crime's future location and time. The goal of Big Data Analytics in this case is to recreate the circumstances of previous crimes and patterns, as well as to predict which groups of people are more likely to be victims or perpetrators of crime [17].

CCTNS (Crime and Criminal Tracking Network System) is an e-governance initiative that connects 14,000 police stations as part of the Digital India mission. It operates by anticipating the locations and times that provide opportunities for criminals and are associated with a higher risk of crime. It also serves to identify possible criminals (Fig. 8).

In the first step, Data is gathered from a variety of sources, including police station databases, environmental data, such as seasonal patterns of crime, neighbour-

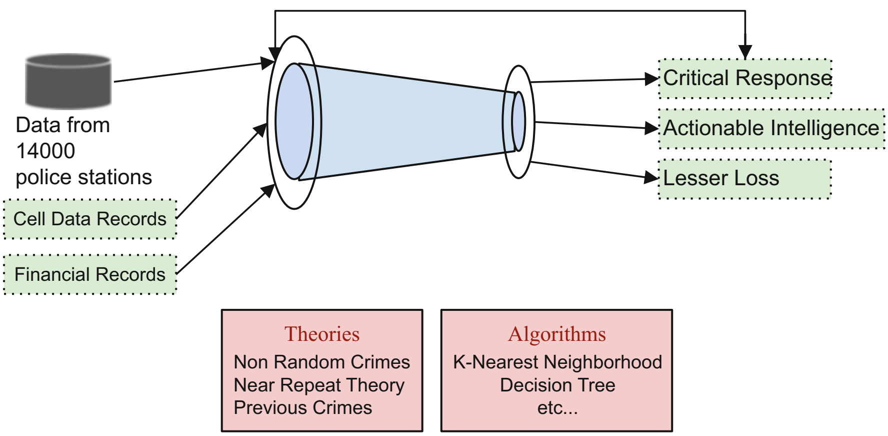


Fig. 8 Predictive policing

hood formation, call data records, financial information, and other mobile phone data. The data obtained is then examined in parallel using predictive algorithms in the next step, Analysis. Near-repeat theory, social network analysis, educational financial stratification, and risk factor regression models are among such examples. Following this study, the actionable item might be anything as simple as increasing the number of forces in a certain area. It's worth noting that the act of predictive policing itself creates fresh data [18]. Finally, the question is whether these efforts have resulted in any reaction, such as a shift in crime patterns or relocation of criminals.

Better resource allocation, preventive policing, and more comprehensive analysis are all advantages of these systems. The drawbacks with this methodology include its inherent biases and the prediction model's opacity. The slums of Mumbai, for example, are well-known for their criminal activities. If a member of this community has a police record, he will be referred to as a "history sheeter," and it is possible that the factors or data that led to his suspicion are biased data from the police database. Systemic bias may lead to patterns and conclusions when combined with irrelevant data and suppressing [19]. Implicit biases in police data about this neighbourhood can lead to increased prejudices in everyone who lives there.

Because he comes from a community with a high number of history sheeters, the police may find many connections to other criminals/suspects/history sheeters through social network theory analysis. As a result, any use of predictive policing software must be accompanied by training for law enforcement officers on how to critically question the software's findings before implementing them. However, the over-reliance on statistics and the omission of other aspects remain suspect. What assumptions the algorithms make and what kind of contextual questions the algorithm asks during the analysis with predictive models and algorithms are completely opaque. As a result, greater openness and information accessibility are required.

Each action carried out as part of predictive policing adds to the data and biases that influence judgments. Actuarial science is a branch of mathematics and statistics that is used to estimate risk in insurance, finance, and other fields. Predictive policing is a multidisciplinary process that combines insights from various fields such as actuarial science statistics, criminology, and a contextual understanding of the local environment [20]. As a result, it is envisaged that more thorough and holistic studies of crime trends, as well as meticulous and critical analysis, would result.

9.2 Unique Identity Project: From Cradle to Grave

With 121 crore registrations, the Unique Identity Project, Aadhaar, is a paperless identity system that can be used to access a variety of digital services. The National Security Depository Limited (NSDL), with the help of its 276 partners, collects personal information like location, contact information, race, gender, etc., along with biometric details like retinal scanning and fingerprints, an individual is mapped in a central database. It's a biometric identity system in which a person is recorded in the Central Identities Data Repository, a government database (CIDR). It has all of the Aadhaar numbers, as well as biometric and demographic information. One of the most profitable use of Aadhaar is that it allows the user to get all government benefits. When you connect it to your bank, all payments are sent immediately to the beneficiary's bank account. This system generates a significant quantity of data that is valuable to both the public and business sectors. This is a single-window system for citizen verification and service provision. As a result, service providers can concentrate on their products and services rather than worrying about verification and other systemic issues [21]. In India, the database and technology are now in place to handle such massive data. As the platform matures, hundreds more services have migrated online. India Inc., now that it is digitally equipped, has generated a massive amount of data.

A Digital Identity system has the following components:

1. *Identification* – Through a biometric system,
2. *Authentication* – In India, as a digital online evidence of identity and address. Because the biometric information is available online, authorized entities can cross-verify and confirm the identity before proceeding with the process of providing services and subsidies. The CIDR is used to connect biometric data to the Aadhaar number's biometric data. The false-positive ratio (the probability that the identifiers of two persons will match) of this procedure has been estimated to be extremely high, and in a population size of 1.37 billion people, and a 12 digit random number is generated, the expected proportion of resulting duplicates could be as high as 1/121 2016 (S.M. Hans Mathews) [22].
3. *Authorization* – The process of ascertaining whether a particular service or subsidy can be given to this person.

Seeding: The service delivery database is updated with all unique Aadhaar numbers. Seeding is the term for this procedure. This is done to verify that no data is duplicated and that authentication is flawless. To facilitate large data processing, unique identities are added to various databases and tagged together for relationality. The seeding procedure can be done manually or using organic methods. It may also be created by specified databases, or inorganic processes as they are commonly referred to.

The organic process involves building a database through door-to-door collection, NSDL agents, banks, Interactive Voice Recognition System (IVRS), SMS, File Hosting Systems, or DropBoxes. Point of Sale, Know Your Residents, Enrolment ID, Unique ID, a database of beneficiaries, smart devices, and other inorganic processes are used. In India, where people rely on technology for essential services such as rations, benefits, and subsidies, as well as credit opportunities, such technological failures come at a high cost (Fig. 9).

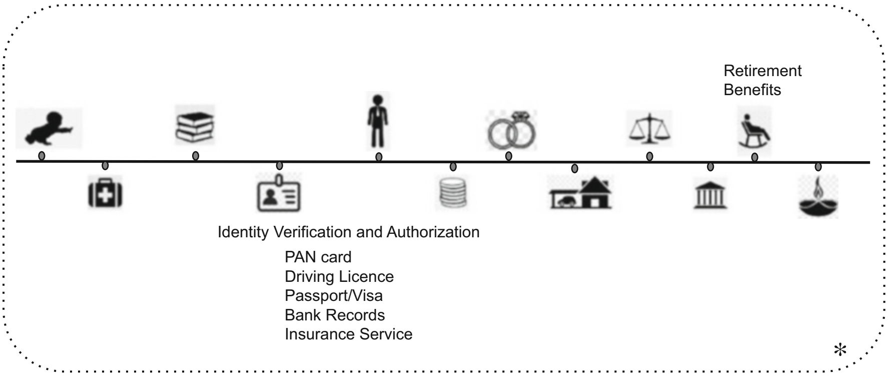


Fig. 9 Janam Mrityu: from cradle to grave

This graphic depicts how an individual's whole lifespan is digitally documented. The e-District creates a birth certificate when a kid is born. At this point, he can create his first UID. It's his first foray into the digital realm. His immunization and other health data are also being digitized at the same time. He matures to the point when he can attend school, and his educational records are established. He takes advantage of educational programmes, loans, and scholarships. At the age of 18, he obtains an adult Aadhaar card, a bank account, a passport/visa, and a driver's licence through this process [23]. He registers with an employment exchange after graduation, tries on his own, and obtains a job. When he gets a job, he becomes insured and pays income tax as well as other indirect taxes. In the meanwhile, he registers a car in his name and purchases a home or property. Meanwhile, he has experienced another major life event: marriage, for which he holds a marriage certificate. If he is hospitalized or has a family issue, he will have medical documents, court hearings, and so forth.

The original press release announcing the Cabinet of Ministers' approval of the ambitious Digital India project talks about "Cradle to Grave," which we might rename "Janm se Mrityu tak." As one of its vision areas, digital identity was mentioned. "Unique, Lifelong, Online, and Authenticable" are some of the features of this identity (Fig. 10).

To map the e-governance programmes under Digital India with those who desire services for individuals to establish a welfare state, a survey was performed. These systems collect and retain information about a person's whole life in order to provide successful services. As a consequence, a database of persons is created. We've set up the seeding procedure on the Ginger platform and have the necessary technologies in place [24]. As a result, it may be used as a 360-degree digital identity platform. All of Digital India's initiatives are based on a single digital ecosystem. The major key function is Aadhaar. A Pan card, passport, and driver's licence are among the other essentials (Fig. 11).

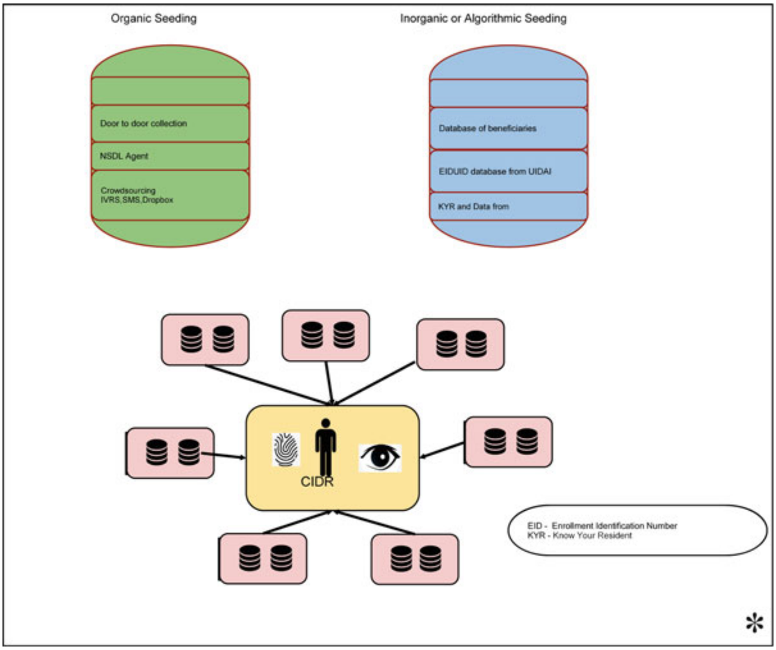


Fig. 10 Aadhaar seeding system

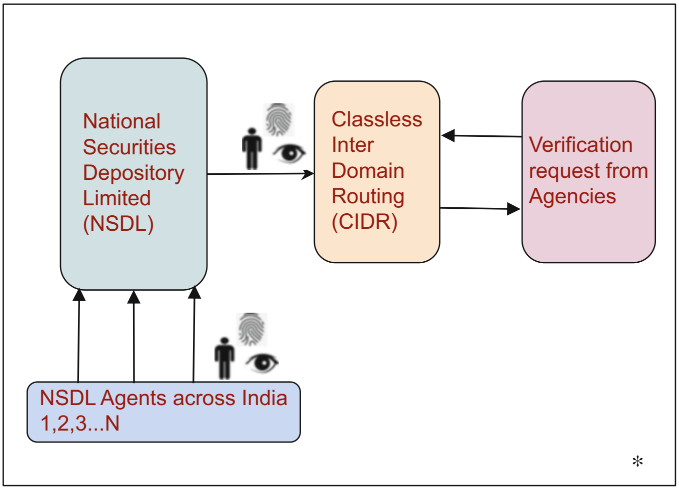


Fig. 11 The UID system

Pros and Cons

Pros:

1. The envelope containing all of the data points policymakers in the social welfare department in the direction of developing new policies, changing existing ones, and eliminating bad ones in favour of countrymen [25]. It also prevents the same person from receiving benefits from two districts.
2. All data inputs are fed into the system just once, resulting in a Single Window System or Data Convergence. Citizens are not repeatedly asked for the same information. As a result, adjustments such as address/phone number changes do not need to be submitted in numerous databases and agencies. Money laundering and tax fraud are prevented by bringing all departments' databases together on a single platform [26].
3. Also, because all data is centralized, any information about a new government scheme, as well as the procedures for completing paperwork and receiving services, can reach the targeted group of people directly.

Cons:

1. Data graining with divergent data gathering by multiple agencies is used to create a 360-degree profile of an individual. This results in a demographic that is suitable for policymakers and allows for more surveillance. At the same time, privacy is compromised in this database. Citizens' rights to free speech and expression may be eroded as a result of this.
2. This information is shared throughout all government departments and organizations. Citizens have no idea where data is stored, who owns it, or how it is used. This opacity might lead to mistrust between governments and populations [27].
3. If policy judgments are based on unconnected and faulty facts, they may be incorrect. If security is not properly handled, the volume of data collected over the last decade, combined with sophisticated analysis and mining techniques, can lead to dangerous situations.

9.3 Intelligent Transportation System

Intelligent Transport Systems (ITS) is an integrated application of communications, information and control technologies to monitor and manage a transportation network [28]. This system enables the collection of data and intelligence. This database is then analysed and thus the resultant insights are shared back to traffic managers and road-users (Fig. 12).

The main objective of ITS is to improve commuter service, better traffic management, improved safety of drivers and commuters, optimum use of fuel, use of the shortest route, informed decision-making, and improved profitability of the service provider. These systems are integrated and use standalone applications also.

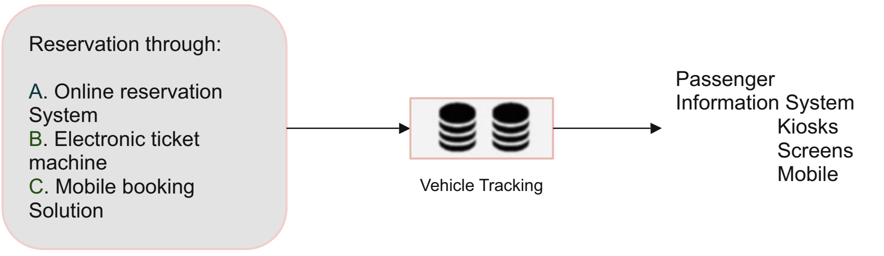


Fig. 12 Intelligent transport system

The components of Intelligent Transport Systems are sensors network, connected vehicles, dynamic smart traffic lights, passenger information kiosks, automatic navigation facilities, GPS-enabled public transport, CCTV systems, and mobile application.

The mobile application collects and reflects the information in terms of Camera, Wi-Fi information, device id, call information, and sends related SMS to the passenger.

In India, if someone wants to use the ITS mobile app, one has to give permission to all access to information on the phone. An individual has to register mandatorily before he/she makes an online booking [29]. The personal information so asked is really exhaustive. If a person is a regular user of the app of Bengaluru Metropolitan Transport Corporation, the BMTC servers have his details of the journey, frequency, location, and even personal data on her mobile phone, making her vulnerable.

Since the data is being used by the city planners and policymakers, individuals with limited digital literacy may be unrepresented in planning and decision-making. In India, such individuals are in crores.

Like any other system, this also has its pros and cons. The benefits of ITS for passengers and organizations are the following:

1. ITS helps to identify the loopholes in the system and identifies the pilferage, It improves revenue management and reduces fuel consumption.
2. The Tracking location of vehicles ensures preventive maintenance of vehicles and avoids potential accidental threats. The Voice kit enables instant communication to inform about mishaps, thereby arresting the loss of vehicles and precious mankind.
3. The Increased efficiency of the vehicles results in better output of the ITS system and improved forecasting promotes reliability on transports [30]. The reduced waiting time for passengers straightway relates to helping prevent wasting of the man-hours and the trust increases the passenger base.
4. Traffic congestion on the roads is a major task for Indian cities. Route planning and route rationalization with the help of VTUs can help here.

5. Components like Efficient Transport System help reorganize the fleet of vehicles as per the demand prediction.
6. The traffic patterns are examined via real-time data. It can help transport authorities to understand commuters' behaviour. The targeted information is provided and policy interventions are identified and executed. Also, thereby proposing new roads for traffic decongestion and lower accident possibility.

The disadvantages are the same as any digital platform:

Like most of the apps, in ITS also, the passenger is supposed to permit all kinds of information sharing without a clear purpose. Whenever a person downloads an app, it is clearly a requirement. The psychology is that at that time, he/she is desperate to have that app at all costs and conditions and allows all kinds of personal information sharing with the vendor [31]. At first, the vendor is also clueless about what to do with this data, later, other organizations and vendors start demanding for that information. The person is traded for his/her desperation for information on using that app. This over-collection of personal information helps the vendor get information about the location and route of a vehicle. This clearly violates the principle of data minimization, poses a risk of identification and other risks beyond the imagination of any healthy mind. The privacy policies in India along with inadequate guidelines to regulate pose a risk to individuals for identity theft. The unclear use of collected data across the platforms and cities poses a real risk for individuals.

The data is collected through different systems, and organizations having different sets of analysts. It is stored in different servers and databases. This can result in poor quality of data collection, retrieval, and processing. This may lead to unreliable analysis, results, thereby making wrong decisions [32]. The unrepresented lot further increases the exclusion and bias. Thus, the potential of smartphone apps to address social problems of transport management should not be overstated. It has to be a balanced one, as these apps improve the system efficiency but may not be able to bridge the social divide.

9.4 Smart Meters

The traditional energy meters record the energy consumption, the meter is read manually, and accordingly, the energy services company (ESCO) creates and sends the bill by email or snail mail. The consumer deposits the payment at the cash counter, and now electronically, through banking apps and other payment apps.

On the contrary, the Smart Meters measure real-time energy usage, analyse it, and send this information to the energy service company remotely. These meters collect the data, monitor supply, transmit data, and also communicate with other smart devices at home [33]. This analysis is shared with the consumer to understand their energy usage pattern and smartly manage their energy bills. For example, an Energy Service Company has four slabs:

- 0–500 units @Rs. 4.50
- 501–1000 units @Rs. 7.50
- 1001–2000 units @ Rs. 20.00
- 2001–beyond @ Rs.100.00

The consumption always goes a little beyond 2001, and the customer repents paying @Rs.100.00 for the entire bill. On close monitoring, it is found that the street lights of the institution keep "On" even during the day, and switching them off can easily bring the bill well below 2000. Making a huge saving on the entire bill.

These kinds of insights help consumers save on money and ESCOs to help distribute that energy to needy industries and optimize the processes. They can redistribute the services and innovate further [34–35]. Thus the consumer is able to understand the energy consumption, make the corrections, and manage their energy usage more efficiently. Moreover, since all this is done in real-time and remotely, the consumer can manage their energy usage in real-time more efficiently by collecting granular data. Since the smart meter can well communicate with other appliances through IoT, the potential of data collection and mining increases exponentially.

Let us take another example. Let there be a sample citizen, maybe a man, whose activities are as follows as per the timeline of round the day (Fig. 13).

The following insights (and more) are created by doing this.

1. What time does he get up?
2. How long does he take to shower?
3. How much time does he take between the shower and leaving for office?
4. What is the usual temperature of air conditioning required for him?
5. How long does he watch TV, play video games, or watch any DVD (DVD is original or pirated)? Which TV does he use, LCD or LED, and what size?
6. What channels does he watch (remember the digits on black stripes appearing randomly)

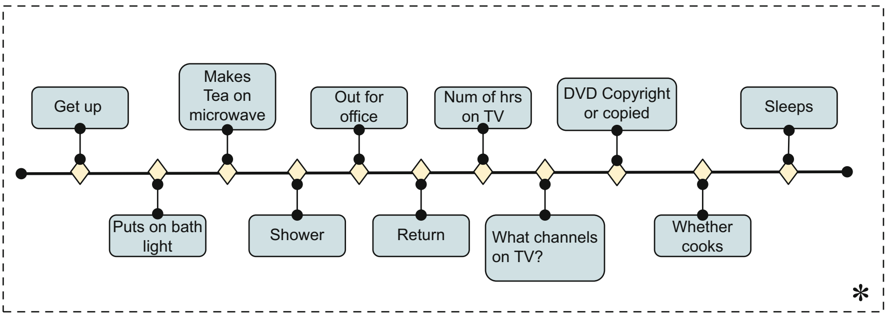


Fig. 13 Smart meters

7. How often does he order food from outside?
8. How often does he cook and how often he has friends? If he is a loner.

These revelations may have an impact on his insurance cost. It's worth noting that these appliance load signatures can be hidden using power management models, power mixing algorithms, and rechargeable batteries, among other things.

Despite the fact that these metres appear to be there to increase energy savings, reduce energy spending (and thus distribution losses), and provide insight into energy consumption, they are not. SCADA technology, which collects data for control and supervisory management, is used in tandem to provide these insights [36].

Smart metres provide information on how much excess energy will be available, whether the grid will be balanced, whether the grid will be able to transmit it, and whether it will be sold [37]. Smart metre data may also give you a heads-up on when and where equipment breakdowns and power outages are most likely to happen. As a result, improved load management, prediction models, and resource allocation may be achieved. They improve efficiency at all stages of the energy supply chain by using big data analytics. This means a more consistent and secure electrical supply. However, the data generated as a result of this procedure is fraught with security dangers and problems. This information compiles a detailed map of daily routines, habits, and other socioeconomic elements. Smart metre owners, for example, are subject to thieves, marketers, police enforcement, creditors, insurance agencies, and so on. There must be a balance between protecting individual privacy via data aggregation, over-regulation, and maximizing Big Data's potential. The best alternative is to tell consumers about the potential uses of big data and the numerous entities with whom their information will be shared. Maintaining data dissemination openness should be the best policy. Again, the population that does not have a working electricity connection is severely marginalized, or does not use smart metres may be underrepresented in the secondary market.

Before Corona, it was estimated that 780 million smart metres will be installed by the end of 2020. The Asia-Pacific area was anticipated to account for 65% of this, or about 500 million people. India is expected to be "Fully Smart" by 2021–2022, thanks to the government's promise to provide smart metres to every home.

References

1. R. Paul, Big data analysis of Indian premier league using Hadoop and MapReduce, in *Computational Intelligence in Data Science(ICCIDS) 2017 International Conference on*, (2017), pp. 1–6
2. S. Srivastava, N. Chaudhari, Appraising a decade of research in the field of big data "The next big thing", in *Computing for Sustainable Global Development (INDIACom) 2016 3rd International Conference on*, (2016), pp. 2171–2175
3. P. Sachar, V. Khullar, Genetic Algorithm Using MapReduce-A Critical Review. *i-manager's J. on Cloud Comp.* **2**, 35 (2015)
4. S. Sedkaoui, M. Khelfaoui, Understand, Develop and Enhance the Learning Process with Big Data, Information Discovery and Delivery (2018)

5. M. Recanatini, C. Cabrelle, Drug research meets network science: Where are we? J. Med. Chem. (2020)
6. F. Berman, Got data?: A guide to data preservation in the information age. Commun. ACM **51**, 50–56 (2008). [Online]. Available). https://doi.org/10.1145/1409360.1409376
7. J. Dean, S. Ghemawat, Mapreduce: Simplified data processing on large clusters. Commun. ACM **51**(1), 107–113 (2008)
8. E. Levy, A. Silberschatz, Distributed filesystems: Concepts and examples. ACM Comput. Surv. **22**(4), 321–374 (1990)
9. P. Bhatotia, A. Wieder, R. Rodrigues, U.A. Acar, R. Pasquin, Incoop: MapReduce for incremental computations, in *Proceedings of the 2nd ACM Symposium on Cloud Computing (SOCC '11)*, (ACM, New York, 2011)
10. J. Dean, S. Ghemawat, MapReduce: Simplified data processing on large clusters, in *Proceedings of the 6th Symposium on Operating Systems Design &Implementation-(OSDI'04)*, vol. 6, (USENIX Association, Berkeley, 2004)
11. J. Kour, N. Awasthy, Non minutiae based fingerprint matching. International Association of Computer Science and Information Technology – Spring Conference, Singapore, pp. 199–203 (2009)
12. Understanding Hadoop, http://docs.hortonworks.com/CURRENT/index.htm#About-Hortonworks-Data-Platform/Understanding-Hadoop-Ecosystem.htm
13. F. Chang, J. Dean, et al., Bigtable: A distributed storage system for structured data. ACM Trans. Comput. Syst. **26**(2), 205–218 (2008)
14. J. Petrovic, Using memory cached for data distribution in industrial environment, in *Third International Conference on Systems (ICONS08)*, Cancun, 2008, pp. 358–372
15. J. Shafer, S. Rixner, The Hadoop distributed file system: Balancing portability and performance, in *2010 IEEE International Symposium on Performance Analysis of Systems & Software (ISPASS2010)*, White Plains, March 2010, pp. 122–133
16. J. Xie, S. Yin, et al., Improving MapReduce performance through data placement in heterogeneous Hadoop clusters, in *2010 IEEE International Symposium on Parallel & Distributed*
17. S.C. John, J.A. Stankovic, J.F. Kurose, D. Towsley, Performance evaluation of two new disk scheduling algorithms for real-time systems. J. Real-Time Sys. **3**, 307–336 (1991)
18. M. Stonebraker, D. Abadi, D.J. DeWitt, S. Madden, E. Paulson, A. Pavlo, A. Rasin, Mapreduce and parallel DBMSs: Friends or foes? Commun. ACM **53**(1), 64–71 (2010)
19. K. Jaspreet, A. Neeta, Non minutiae based fingerprint matching. Computer Science and Information Technology, International Association of. 199–203 (2009). https://doi.org/10.1109/IACSIT-SC.2009.127
20. R.E. Bryant, Data-intensive scalable computing for scientific applications. Comput. Sci. Eng. **13**(6), 25–33 (2011)
21. D. Agrawal, et al., Challenges and opportunities with big data, *Cyber Center Technical Reports*, (2011)
22. A. Kumar, N. Awasthi, An efficient algorithm for text localization and extraction in complex video text images, in *2013 2nd International Conference on Information Management in the Knowledge Economy*, pp. 14–19 (2013)
23. V.B. Bobade, Survey paper on big data and hadoop. Int. Res. J. Eng. Technol. (IRJET) **03**(01) (2016)
24. V.K. Singh et al., A literature review on Hadoop ecosystem and various techniques of big data optimization, in *Advances in Data and Information Sciences*, (Springer, Singapore, 2018)., [online] Available: https://doi.org/10.1007/978-981-10-8360-0_22
25. M.K. Karthiban et al., Big data analytics for developing secure internet of everything. J. ISMAC **1**(02), 129–136 (2019)
26. Z. Liu, A. Zhang, Sampling for big data profiling: A survey. Access IEEE **8**, 72713–72726 (2020)
27. G. George, D. Lavie, Big data and data science methods for management research. Acad. Manag. J. **59**(5), 1493–1507 (2016)

28. D. Agarwal, S. Das, A. El Abbadi, *Big Data and Cloud Computing: Current State and Future Opportunities* (ACM, 2011)
29. A. Cuzzocrea, Privacy-preserving big data management: The case of OLAP, in *Big Data-Algorithms Analytics and Applications*, (2015), pp. 301–326
30. Z. Liu, K.R. Choo, M. Zhao, Practical-oriented protocols for privacy-preserving outsourced big data analysis: Challenges and future research directions. Comput. Secur. **69**, 97–113 (2017)
31. S. Wang, W. Li, F. Wang, Web-scale multidimensional visualization of big spatial data to support earth sciences-A case study with visualizing climate simulation data. Inform. **4**(3), 17 (2017)
32. J.K. Seng, K.L. Ang, Big feature data analytics: Split and combine linear discriminant analysis (SC-LDA) for integration towards decision-making analytics. IEEE Access **5**, 14056–14065 (2017)
33. A. Rani, N. Goyal, S.K. Gadia, Data provenance for historical queries in relational database, in *Proceedings of the 8th Annual ACM India Conference*, pp. 117–122, 29–31 October 2015
34. M. Sokolova, Big text advantages and challenges: Classification perspective. I. J. Data Sci. Anal. **5**(1), 1–10 (2018)
35. A. Cuzzocrea, Privacy-preserving big data management: The case of OLAP, in *Big Data – Algorithms Analytics and Applications*, (2015), pp. 301–326, [online] Available: https://doi.org/10.1201/b18050-21
36. S. Kung, Discriminant component analysis for privacy protection and visualization of big data. Multimed. Tools Appl. **76**(3), 3999–4034 (2017). [online] Available: https://doi.org/10.1007/s11042-015-2959-9
37. N. Victor, D. Lopez, J.H. Abawajy, Privacy models for big data: A survey. IJBDI **3**(1), 61–75 (2016). [online] Available: https://doi.org/10.1504/IJBDI.2016.073904

Experimental Investigation of Eco-enzyme and Its Application for Removal of Foul Odour and Organic Impurities

AnoopNarain Singh, Anil Kapoor, and Maaz Allah Khan

1 Introduction

Fifty years ago, the world was facing a shortage of food grains and people in several developing countries were dying. At that time, the need of the hour was to save the people dying of hunger. The green revolution came and saved the day. Today, the environment has occupied prime focus due to pollution, which leads to climate change. The root cause of this pollution – higher temperature and climate change – is the generation of gases like methane and other toxic material, leading to spread of foul odour. However, least attention is given to this challenge. Municipal Corporations and Civic Bodies are struggling to remove the foul odour in drains, garbage dump yards, water bodies, meat and fish markets, public toilets, STPs/ETPs, etc. Partly, it was ignored because no remedy was available until now. Sewage water flows into drains and merges into rivers, ponds, and lakes. Garbage is dumped into landfills, rivers, and ponds. As a result, methane and other toxic gases are formed. Methane in its pure form does not create any foul smell. However, when mixed with other gases, it emits foul odour. It is a silent killer, polluting the area. News reports appear every other day about the deaths of sweepers when they enter the sewage drains, due to the formation of H_2S, methane, and toxic gases. Even electric and electronic appliances are affected and costly bathroom fittings are corroded due to toxicity of the emanating gases.

A. Singh
Rajkiya Engineering College, Azamgarh, Uttar Pradesh, India

A. Kapoor
The Art of Living International Centre, Bengaluru, India
e-mail: anilkapoor@ssrdp.org

M. A. Khan (✉)
Maharishi University of Information Technology, Lucknow, Uttar Pradesh, India

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_7

In this paper, effort has been made to highlight the methodology to reduce foul odour all around. In the last 3 years, we had conducted several experiments to reduce foul odour and stench emanating from different places. The indigenously made eco-enzyme was applied in diluted form in different proportions and observed that foul smell was greatly reduced, giving relief to the local populace. In the experimental investigation, the eco-enzyme is used for bio-remediation. It is an enzyme, produced by fermentation of raw kitchen waste. It is the application which makes it innovative and effective to remove foul smell. Also some laboratory results are presented which show the ability of the eco-enzyme to reduce organic impurities at a faster rate. Microbes and enzymes have a very close relationship. Microbes such as bacteria, fungi, algae, protozoa, and viruses are living organisms essential for life on the planet. Most of the microbes are harmless. They are essential for the cycling of nutrients in the eco-systems of the planet. Microbial activity is used for the benefits of human kind such as production of medicines, food, enzymes, cleaning up of sewage and other waste, etc. There are about 4000 types of enzymes and not even 10% of these have been identified. The exact functioning of enzymes is still an area of ongoing research. Enzymes are catalysts and promote the activity. These are classified for food, pharma, textile, paper, wastewater industries, etc. Enzymes fall under six classes according to their functions:

1. Oxidoreductases – oxidation
2. Transferases – transfer groups
3. Hydrolases – hydrolysis
4. Lyases – hydrolysis and oxidation
5. Isomerases – conversion into single molecule
6. Ligases – joining the molecules

Enzymes are classified based on their target substrates:

1. Proteases split proteins into amino acids
2. Lipases split fats into fatty acids and glycerol molecules
3. Amylases split carbohydrates such as starch and sugar into simple sugar such as glucose
4. Celluloses break down cellulose

Pollutants consist of various combinations of carbon (C), nitrogen (N), hydrogen (H), oxygen (O), sulphur (S), and other elements like CH_3, C_2H_5, H_2S, SO_2, NO_2, and NO_3. These pollutants serve as food for the microbes that eat away the solid material in the polluted water or dump yards and break down the pollutants and discharge free radicals of C, N, H, O, S, etc. The results indicate that the garbage enzyme can remove ammonia nitrogen and phosphorus in wastewater dilutions [1–3]. The BOD and COD values of the grey water were reduced significantly [5]. It is effective in aerobic as well as anaerobic conditions.

Dr. Rosukon P [4] of Thailand has developed a very innovative enzyme called garbage enzyme or eco-enzyme, a complex eco-enzyme produced by fermentation of raw kitchen waste (fruit and vegetable peels), sugar (brown sugar, jaggery, or molasses sugar), and water. Since it is made of different vegetables and fruits, it contains different types of enzymes.

Generally, enzymes are available in the market cultured from one plant. Therefore, it can break down one type of pollutant, but this eco-enzyme has the property to break down pollutants of all types present in wastewater due to the presence of multiple types of enzymes. When it is diluted 1000 times in water (1 l of biosolution in 1000 l of water) and kept overnight, microbes multiply and become innumerable.

This eco-enzyme helps to produce a sustainable and pollution-free environment [5]. Microbes are used to meet effectively the crisis in both environment and energy sectors [6–15]. When such a large work force is employed, it speeds up the bioremediation process. Best results are visible from the 5th day onwards. The question that comes to mind is that such a large force may be harmful, wherever it goes. The answer is that the life span of these microbes is 5–6 days. If multiplication is there, depletion process is also at work. Nature balances the process. This is the cheapest technology available at the moment. One of the main concerns is whether this eco-enzyme is safe to use in water bodies. Bioassay test (Fish survival test) conducted by NABL-verified lab shows that it is safe to use this eco-enzyme and it does not harm aquatic life.

2 Materials and Methods

Eco-Enzyme was prepared by putting jaggery, kitchen trash (orange peels, green pea peels, etc.) and water in a weight ratio of 1:3:10 (100 gm jaggery, 300 gm peels, 1000 ml). The mixing was done in an expandable airtight plastic container. During the first month of the fermentation process, gases were emitted. To avoid bursting, the container's pressure was released on a daily basis. Every now and again, orange peels were pushed below. The container was stored somewhere cool, dry, and well ventilated. To manufacture enzyme, it was fermented for 3 months. The solid particles were removed from the brownish liquid produced by the fermentation. After 3 months, the fluid was filtered to get the enzyme solution. On the top surface of the solution, a white mould development was noticed. B complex yeast and vitamin C yeast are two examples. The enzyme solution obtained was pale brownish yellow in colour. The liquid was then poured into a plastic bottle. Enzymes have no expiration date. It gets stronger the longer you keep it. When water is introduced to the enzyme, its strength is increased. Garbage Enzyme can only be used externally. It should not be kept in the fridge. The features of pure Garbage Enzyme solution were investigated immediately after the enzyme solution was filtered (after 3 months of fermentation period).

3 Results

Results are presented in two sections, first section presents laboratory results which analysed the ability of eco-enzyme for reduction of solids and organic impurities and second presents field results which shows its ability in reducing foul smell and organic impurities from lakes, water bodies and dump yards.

3.1 Laboratory Results

The goal of this research is to use enzymes in wastewater treatment to reduce organic pollutants. As a result, processes for preparing fruit enzyme and conducting laboratory tests must be designed and carried out throughout the investigation. This research begins with the production of fruit enzyme, which takes 3 months. Following that, water samples are taken from various locations for laboratory testing and wastewater classification. The study's findings are based on laboratory testing of samples after enzyme treatment.

3.1.1 Sample Collection

Sewage was collected from the end of an open drain and from a manhole at Krishna Engineering College (KEC), Mohan Nagar, Ghaziabad, India. The samples were collected in the morning at around 10 a.m. (Fig. 1).

The water quality indicators were observed and the water quality condition was classified in a laboratory test. BOD, TDS, Ph, Conductivity, and Turbidity were measured in both raw and processed wastewater. The raw wastewater and the treated wastewater were compared.

Table 1 shows the parameters of pure garbage enzyme solution after 3 months of fermentation (immediately after enzyme solution filtration) and 60 days of filtration.

The characteristics of pure wastewater is given below in Table 2.

Characteristics of effluent following treatment with garbage enzyme solutions containing 5%, 10%, 20%, and 25% (60 days after filtration).

The wastewater was treated with garbage enzyme solutions containing 5%, 10%, 20%, and 25% garbage enzyme. Thereafter, they were left to digest. pH, TDS, BOD5, ammonia nitrogen, and phosphorus levels were all measured.

After 5 days of digestion, the effluent met the irrigation standards (as per E(P) guidelines). Table 3 shows the effluent properties following treatment with 5%, 10%, 20%, and 25% trash enzyme solution after 5 days of digestion. Table 4 shows the percentage reduction of several parameters following treatment with 5%, 10%, 20%, and 25% trash enzyme solution after 5 days of digestion.

According to the findings of the study, garbage enzyme solution is not suited for treating wastewater immediately after filtration. With time, the enzyme properties

Fig. 1 Sewage collection point at manhole, KEC

Table 1 Characteristics of pure garbage enzyme solution

Parameter	After immediate filtration	After 60 days of filtration
pH	2.85	3.8
TDS	2216	1130
BOD	1250	90.5
Ammonia	BDL[a]	BDL[a]
Phosphate	BDL[a]	BDL[a]

[a]Below detectable limit

Table 2 Characteristics of pure waste water

Parameters	Unit	Value
pH	–	6.15
TDS	mg/lit	525
BOD	mg/lit	192
Ammonia	mg/lit	9.6
Phosphate	mg/lit	115

Table 3 Effluent characteristics after 5 days of digestion

Parameters	5%	10%	20%	25%	Irrigation standards
pH	6.9	6.6	6.4	5.8	5.5–9.0
TDS (mg/lit)	258	233	410	532	2100
BOD (mg/lit)	78	68	91	96	100
Ammonia(mg/lit)	0	0	0	0	–
Phosphorus(mg/lit)	0	0	0	0	–

Table 4 Percentage reduction of various parameters

Parameters	5%	10%	20%	25%
pH	–	–	–	–
TDS (mg/lit)	54.55	58.80	27.19	5.51
BOD (mg/lit)	61.46	65.44	54.51	49.51
Ammonia(mg/lit)	100	100	100	100
Phosphorus(mg/lit)	100	100	100	100

changed. When the treatment was done 60 days after the enzyme solution was filtered, the treatment was effective.

3.2 Field Results

We have achieved unique success in removing foul smell from drains, ponds/lakes, STPs/ETPs, garbage dump yards, etc., in 3–20 days. In the last 3 years, we have applied it in drains, rivers, ponds, STPs/ETPs, public toilets, slaughter houses, garbage bins, etc. to eliminate foul smell, reduce BOD, COD, increase DO level, contain methane and other toxic gases. We have reclaimed land of dump yard at Ghaziabad, U.P., India. Leachate oozing out of dump yard is another problem since it emits terrible odour. Removed its foul smell and broke its combination. Results of the studies obtained at different location/site in India are as given below:

3.2.1 Barapullah Drain, Delhi (16 February to 9 March 2016)

The Art of Living organized World Cultural Festival near Sarai Kale Khan, Delhi in March, 2016. Barapullah Drain passes along the venue before merging into Yamuna River. It starts from Ashoka Hotel in New Delhi and the drain water is not very polluted. The drain passes through Lakshmi Bai Nagar, Defence Colony, Lajpat Nagar, Kale Khan, etc., and reached Sun Dial area before merging into Yamuna River. In between, there are several colonies and slum areas from which untreated sewage water and discharge from commercial establishments join in the drain. It contains harpic, cleaning agents, soap and detergents, grease and oil, etc., converting them into pollutants and toxic material. The drain is partially covered with concrete slabs, but when it opens up near Defence Colony area, it blows out methane and other toxic gases and emits a pungent odour. The problem continues further aggravating the situation until it merges into Yamuna River. Since a large number of dignitaries around the globe and foreigners were expected to attend the function, it was decided to remove foul smell through this eco-enzyme. It was poured into the drain for 15 days from 16 February 2016 at the rate of 200 l per day in concentrated form because it gets diluted into the drain immediately. From the 5th day onwards, foul smell started disappearing and by March 10, it completely disappeared and the

Fig. 2 Barapullah drain near Sun Dial, Kale Khan, Delhi – highly toxic

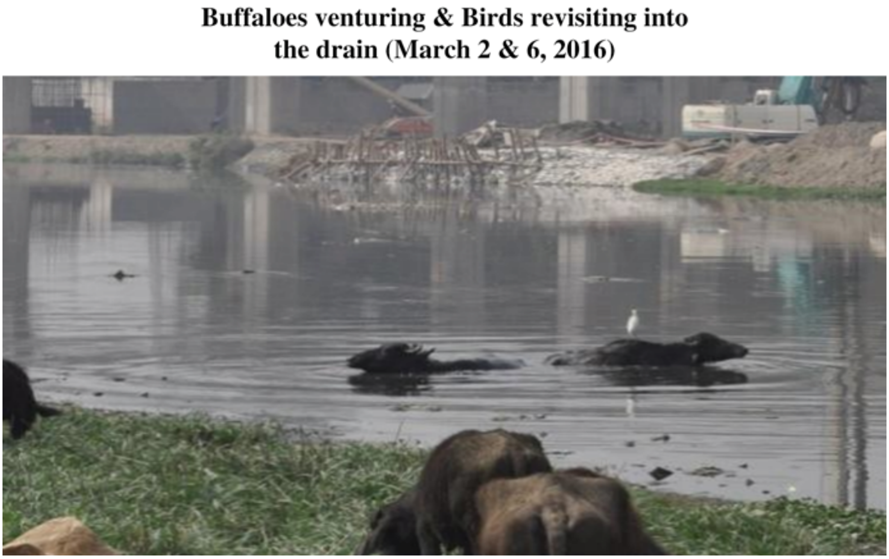


Fig. 3 View of drain before the treatment

event passed off well without any inconvenience to the foreigners, dignitaries, and audience (Figs. 2, 3, and 4).

3.2.2 Drains at Pune

Four rivers, namely, Mula, Mutha, Ram Nadi, and Pavana, flow through Pune city. Indrayani is another river flowing through Pimpri Chinchwad area and merges into Pavana River. Most of these rivers are now sewage drains as the discharge from

Fig. 4 View of drain after the treatment

households and industries goes into them. Pimpri Chinchwad is the industrial belt and big giants like Tata Motors, Mahindra & Mahindra, Bajaj Auto, Volkswagon, etc., are located there. There are 27 drains which carry polluted water and merge into Indrayani River. We decided to clean these drains and approached Pimpri Chinchwad Municipal Corporation (PCMC) with the request to allow us to conduct the treatment and clean the wastewater of these drains. PCMC allowed us to conduct trial on two drains carrying polluted water into Indrayani River. Three drums were installed on two drains with 5 l of eco-enzymes in the each drum diluted with about 1000 l of water and kept overnight. Next day, it was poured. Treatment was conducted from 2 to 19 May 2017. The eco-enzyme was poured for 12–14 h. Sampling was done by PCMC lab and it was found that the polluted water in the drain improved greatly with reduced COD and BOD levels after the treatment with the eco-enzyme within 17 days. COD content reduced from 272 mg/lit to 56 mg/lit in one drain and BOD from 105 mg/lit to 22 mg/lit. In the second drain, COD content reduced from 260 mg/lit to 40 mg/lit and BOD from 95 mg/lit to 20 mg/lit. Details are as given in Table 5.

3.2.3 Sanjay Park, East Delhi

Sanjay Lake is located in over 50 acres in Trilokpuri of East Delhi. It is a beautiful water body with beautiful pathways and GYM facilities. However, the administration is struggling to remove foul odour in the lake. It is under the control of Delhi Development Authority which allowed (August, 2017) us to treat with eco-enzyme. Central Pollution Control Board (CPCB) was also associated in the

Table 5 TSS, COD, BOD levels before and after treatment at PCMC drains (Pune)

		Parameters		
Location	Date	TSS mg/lit	COD mg/lit	BOD mg/lit
Near Symbiosis College, Kiwale	2 May 2017 (before treatment)	105	272	105
	19 May 2017 (after treatment)	5	56	22
Dehu Rear Pump House, Kiwale	2 May 2017 (before treatment)	28	260	95
	19 May 2017 (after treatment)	5	40	20

Table 6 COD, BOD levels before and after treatment at Sanjay Park, East Delhi

	Parameters					
Sampling date	Location	pH	COD mg/lit	BOD mg/lit	E-Coli	Total plate counts
07 Nov. 2017 (before treatment)	Near entry gate	7.84	83	25	Present	1.2 × 10 (5)
07 Nov. 2017 (before treatment)	Near tubewell	8.27	55	16	Present	1.4 × 10 (5)
23 Nov. 2017 (after treatment)	Sanjay Lake	7.68	78	21	Absent	1.6 × 10 (4)

project for conducting tests of water before and after the treatment. Two tankers of 30,000 l were engaged and 300 l of the eco-enzyme was added in each tanker and kept overnight to develop and multiply the microbes. Diluted 60,000 l of eco-enzyme was poured into the lake for 5 h on 8 Nov. 2017. CPCB and Sigma NABL Laboratory took the samples to measure the COD, BOD, and other parameters on 7 Nov. 2017. Again, the samples were drawn on 23 Nov. 2017 to compare the results of treatment.

The results of Sigma NABL Lab showed reduction in COD and BOD levels and increase in DO level. The most important change was elimination of E-Coli in the lake after treatment. Within a month, foul smell disappeared and the water became cleaner. Ducks and migratory birds, which hardly entered the lake water, are seen venturing into the lake more frequently.

Details of water quality parameters before and after treatment of sample at Sanjay Park are given in Table 6.

3.2.4 Ponds in Chennai

We are treating 108 ponds in Chennai. First pond was treated in Semmencherry Karunachaved, Chennai, on 10 June 2017. Before treatment, grey algae existed in the pond, which disappeared after 45 days as visible in Fig. 5.

Before Treatment with the Eco-enzyme (10 June 2017) and After Treatment (3 Oct. 2017) (Fig. 6).

Encouraged with the results, Greater Municipal Corporation sought our assistance to save fishes that were dying in Temple Pond at Sengeiamman Temple Pond,

Fig. 5 View of water body after the treatment at Sanjay Park

Fig. 6 View of pond water before and after the treatment at Semmencherry Karunachaved, Chennai

Table 7 DO, COD, BOD levels before and after treatment at Neelankarai, Chennai

Sampling date	Parameters		
	DO mg/lit	COD mg/lit	BOD mg/lit
28 September 2017 (before treatment)	0.13	195	57
4 October 2017 (after treatment)	2.80	72	19

Off East Coast Road, Neelankarai, Chennai, after they saw positive treatment results in other ponds in Chennai. Eco-enzymes solution (50 l) was poured in the water tanker (5000 l) and kept overnight. The treatment was done in the lake on 28 Sept. 2017. NABL tests were conducted on samples of pond water taken before treatment on 28 Sept., and after treatment on 4 Oct. 2017. Fishes had died in this pond because DO level reduced to 0.13 mg/lit and the water was polluted as shown by high COD and BOD contents. The results showed significant improvement in 6 days. DO level increased to 2.8%, which was the main reason for survival of fishes. The detail of DO, BOD, COD levels before and after treatment at Neelankarai, Chennai, is given in Table 7 (Fig. 7).

Fig. 7 View of pond water after treatment at Sengeiamman Temple, Neelankarai, Chennai dated: 2/10/2017

3.2.5 STP at Chandigarh

Eco-enzymes were tried on an STP in Trishala Residential Complex at Zirakpur near Chandigarh. 5 l of eco-enzymes were diluted in 100 l of water and poured daily for 20 days from 18 May to 5 June 2017. Foul smell disappeared from the discharge of STP water and significant fall was noticed in COD and BOD levels. COD level reduced from 404 mg/lit to 53 mg/lit and BOD from 149 mg/lit to 23 mg/lit.

3.2.6 Drains of Azamgarh City

Figure 8 shows the application of eco-enzyme trickling in a drain through a tank. We have achieved success in removing foul smell from the open drain of Azamgarh city, going through the main market of Mukeri ganj area and creating nuisance and discomfort for the people.

3.2.7 Garbage Dump Yard at Ghaziabad, UP

Foul smell, methane, and toxic gases are prevalent in garbage dump yards. It used to frequently catch fire and there was smoke from the dump coming out. A trial was conducted on a 4-acre dump yard near Haj House at Ghaziabad on 4 Jan. 2017 using

Fig. 8 Application of eco-enzyme at sewage drain (Mukeri ganj Nala, Azamgarh)

diluted bio-solution eco-enzymes in the ratio 1:1000 and kept overnight. Foul smell went off the same day and bio-degradable material vanished within 15 days. Only polythene, plastic, and cloth were left which were also removed later on. The land of the dump yard has been reclaimed (Figs. 9 and 10).

Fig. 9 Mayor inspecting the site during treatment with eco-enzyme, paper, cloth, and polythene left out

Fig. 10 View of reclaimed land

3.2.8 Two Garbage Dump Yards at Dehradun (Uttarakhand)

Foul smell was a problem in the Trenching Ground dump yard at Dehradun and agitation was going on by the residents surrounding the dump yard. Eco-enzyme (10 l) was kept in 10,000 l (Dilution ratio 1:1000) water tanker overnight and treatment was conducted on 6 April 2017. The foul smell disappeared within no time and those agitating were satisfied and testified to continue the treatment (Fig. 11).

Fig. 11 Eco-enzyme treatment at dump yard, Dehradun

Fig. 12 Akola dump yard before treatment (15 Aug. 2017)

3.2.9 Garbage Dump Yard at Akola (Maharashtra)

Foul smell, mosquitoes, and flies in Akola Dump Yard were also a problem as the population residing in the surrounding areas was protesting. A section of the dump yard was treated with the eco-enzyme (dilution ratio 1:1000, kept overnight in the tanker) on 15 August 2017. In 15 days, the size of the dump yard has reduced by 3.5 ft and foul smell and mosquitoes and flies have reduced. Now Akola Municipal Corporation has given us permission to manage the whole dump yard spread over eight acres (Figs. 12 and 13).

The results of experiments are based on trials conducted on ponds, drains, STP, and garbage dump yards. Methodology is dilution of eco-enzyme in the ratio 1:100–1000 and process is bio-remediation. The effectiveness of the process is with correct

Fig. 13 Reduced to height by 3.5 ft (30 Aug. 2017)

application, proper dilution and pouring or spray. In ponds, when it is poured once in a quarter (3 months), it is found that grey algae disappear in 45 days, DO level increases and COD/BOD reduces in the polluted water in 5–20 days.

In drains, the treatment is required to be done continuously because the polluted water is always coming from upstream. The enzyme flows down and its effectiveness is visible 3–5 km downstream. Foul smell and COD/BOD levels are reduced. The longer the treatment in the drains, the better the results obtained. Same is the case in STPs because of the continuous inflow of sewage water and sludge into the STP tank.

As regards garbage dump yards, methane and toxic gases are formed which produce foul smell due to the presence of bio-degradable material. It often catches fire. The spray of bio-solution prevents the formation of methane and other toxic gases. Like water, bio-remediation process takes place and the microbes eat away the solid waste. It results in reduction in size of the dump yard. The process is more effective in dump yards which have more bio-degradable material and slow where inert material (construction waste) is more.

4 Conclusions

According to the findings of the study, garbage enzyme solution was not adequate for treating wastewater immediately after filtering. The features of the enzymes changed over time. The enzyme solution is more concentrated than previous concentrations. The treatment time was also cut to 5 days. It's important to keep track of how enzyme properties change over time. Only once the BOD levels of the enzyme solution are reduced is wastewater treatment with garbage enzyme solution shown to be effective. More research is needed to determine the effects of appropriate additives or activators on enzyme function. For reduction of high initial BOD and COD, studies on pre-treatment approaches prior to enzyme action should be investigated. More importantly, determining the components of the trash enzyme is a crucial step.

Our studies are based on ground-level work and trials conducted. The technology is very effective in removing foul odour, reducing BOD, COD levels, increasing DO level in polluted water of drains, rivers, and ponds, improving air quality and overall improvement in the environment. In dump yards also it is equally effective in removing foul smell, curtailing the formation of methane and other toxic gases, reducing bio-degradable material. Results are achieved in shortest possible time of 5–25 days. It is the cheapest solution available cost-wise. 1 l of bio-solution can clean one lakh litres of polluted water. 3 l of eco-enzyme in 3000 l can treat one acre of dump yard of 1 m height. Bio-remediation and use of enzymes are not fully explored and more research work needs to be done.

Acknowledgements We drew our inspiration from H.H. Sri Sri Ravi Shankar Ji and express our gratitude to Him for guiding us in our work and giving us strength and confidence to continue with our mission. Also we convey our sincere thanks to Dr. Rosukon P of Thailand for guiding us on the technology and its application.

References

1. Manual of Municipal Solid Waste Management, (2016), Central public health and engineering organization, Swachh Bharat Mission, Ministry of Urban Development, Govt of India, www.moud.com. Part II, p. 126
2. F.E. Tang, C.W. Tong, A study of the garbage enzyme's effects in domestic wastewater, World Academy of Science, Engineering and Technology. Int. J. Environ. Chem. Ecol. Geol. Geophys. **5**(12), 892 (2011)
3. C. Arun, P. Sivashanmugam, *Waste management, Journalhomepage*: www.elsevier.com/locate/wasman 44, p. 28–33 (2015)
4. www.enzymesos.com
5. F. Nazim, V. Meera, Bonfring, Treatment of synthetic greywater using 5% and 10% garbage enzyme solution. Int. J. Ind. Eng. Manag. Sci. **3**(4), 111 (2013)
6. P.K. Gupta, *Elements of Biotechnology* (Rastogi Publications, Meerut, 2000), p. 4. (ISBN 0-471-65071-4)
7. Q. Yang et al., Improving disintegration and acidification of waste activated sludge by combined alkaline and microwave pretreatment. Process Saf. Environ. Protect **91**(6), 521–526 (2013)
8. R. Uma Rani et al., Low temperature thermo-chemical pretreatment of dairy waste activated sludge for anaerobic digestion process. Bioresour. Technol. **103**(1), 415–424 (2012)
9. C.H. Shu et al., Purification and partial characterization of a lipase from Antrodia cinnamomea. Process Biochem. **41**(3), 734–738 (2006)

10. M. Perle et al., Some biochemical aspects of the anaerobic degradation of dairy wastewater. Water Res. **29**(6), 1549–1554 (1995)
11. O.H. Lowry et al., Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193**, 265–275 (1951)
12. A. Ayol et al., Enzymatic treatment effects on dewater ability of anaerobically digested biosocial-II; laboratory characterization of drain ability and filterability. Process Biochem. **40**(7), 2427–2434 (2005)
13. H. Abbasnejad, et al., Anaerobic treatment of dairy wastewater in a hybrid UASB and bio filter reactor
14. American Public Health Association (APHA) Standard Methods for Examination of Water and Waste Water (2005)
15. B. Prakash, How to make and use garbage enzyme (2011). http://www.ecowalkthetalk.com

Sign Language Recognition Using AI

Laxmi Kant Sagar, Kartik Kumar, Akshit Goyal, Riya Singh, and Anubhaw Kumar Soni

1 Introduction

According to the World Federation report [1] for Deaf and dumb, over 300 sign languages are used by over 70 million deaf people worldwide. Sign language mainly involves the movement of fingers, hands, eyes, and other parts of the body. Deaf and dumb people always use hand gestures to communicate with normal people, as this is the only way of communicating with normal people. However, normal people often cannot understand the meaning of hand gestures; therefore, communication cannot be fully understood. Thus a need for an application that can detect predefined sign language by recognizing hand gestures and can also provide a facility for users to store the recognized character in text form. In this study, a solution is proposed for hearing impaired people using Artificial intelligence, i.e., deep learning and image processing technologies. The proposed solution aims to improve effective communication with people with hearing difficulties who use any form of sign language to express themselves. The objective of this proposed solution is to decrease the communication gap faced between the mute and deaf and the people by creating a program that can take in speech and convert it into Indian sign language.

2 Literature Survey

Sign language is not a new computer vision. Over the last two decades, researchers have addressed this in various ways. They can be broadly classified into sensor-based, appearance-based, glove-based or vision-based methods. Glove-based or

L. K. Sagar (✉) · K. Kumar · A. Goyal · R. Singh · A. K. Soni
G.L. Bajaj Institute of Technology & Management, Greater Noida, India

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_8

sensor-based approaches are more accurate than vision-based because these gadgets collect data directly from the signer, but they have the disadvantage of signing and carrying them along with the external device. Vision-based methods can be classified into 2-D and 3-D vision-based methods. There are two main methods that are used for making gestures from the user. The first method is vision-based, which provides more freedom to the user to perform natural actions and the second is sensor-based. In paper [2], experiments prove that Hidden Markov models can be used for handwriting recognition. Two real-time experiments are performed for recognizing American sign language, one with hands wearing colored gloves and the other without gloves. But experiments were based on an HMM-based recognition system. The results show the first experiment attains an accuracy of 99%, while the second experiment attains an accuracy of 92%. Paper [3] compares sign language recognition systems using Microsoft Kinect, convolutional neural networks, and GPU acceleration. The result shows that convolutional neural networks outperform other methods with an accuracy of 91.7% by recognizing 20 gestures accurately. CNNs can also be used to automate feature extraction. In paper [4], three different approaches for sub-unit feature extraction are discussed. The first approach is based on appearance only. The advantage of using this approach is that it doesn't depend on high-quality tracking and also includes information that is not trajectory, thus encoding information within moment-based classifiers. The second approach uses 2D tracking with an appearance-based hand-shaped classifier. The advantage of the second approach is that it is less ambiguous and achieves a recognition rate of 73% over large datasets. The third approach translates sub-unit data into 3D. It then combines this 3D data with SP-Boosting, thereby achieving a recognition rate of up to 99.9% on 20 sign data. Paper [5] examines multiple-sensor data with the aim of improving recognition accuracy by disambiguation of noisy data. The experiments are performed on noisy vision data, accelerometer data, and a combination of both. Results show remarkable improvement from previous results, vision data 52.38%, accelerometers 65.87%, and the combined set 90.48% improvement. In paper [6], Kalman filter is proposed to track the locations of hands. In paper [7], noble design is proposed for recognizing American sign language. The design is based on parallel HMMs, which are capable of handling each parallel process independently. Further, it can also be trained independently. A recognition algorithm is also designed for a parallel Hidden Markov Model that runs in time polynomial for several states. In paper [8], automatic American sign language recognition system is presented. The system segments the hands and face of the signer for feature extraction. Experiments show that using a known grammar, the system achieved a 97% recognition rate and a 99% success rate. In paper [9], an algorithm for Hand Gesture Recognition using Dynamic Time Warping methodology is presented. The experimental results showed the recognition rate to be around 90%. In paper [10], a 3D application is implemented by using the hand as an input device. The application is designed using 2D graphical-based interfaces, to provide full-fledged multi-participant Virtual Environment (VE) systems.

3 Proposed Methodology

The solution of vision-based Indian sign language translator was initially started by [11]. An Inception V3 Convolutional Neural Network model was trained with a dataset of 7800 images divided into 26 classes, one for each alphabet. The training process ran for 2000 epochs. The frames get extracted from the input video (from the webcam) which is fed to the newly trained model, which further classifies the input sign in the image. A label is extracted corresponding to this classified image, and if it is an alphabet, it is printed to the terminal. It also incorporates signs for delete and space operations on the output sequence. This process is continued frame by frame till an escape command is hit. It gave a 99% training accuracy and a 98% test accuracy. Different letters gave different accuracies overall (e.g., O gave 99.1%, B gave 77%, etc.).

However, there were a few limitations reported regarding this program: (1) It required the input frames to be free of noise (e.g., shadow, poor illumination, or any other object). (2) It was highly dependent on a white and clear background (especially for a live video). (3) The dataset consisted purely of symbols made by the right hand, so it didn't work with the left-hand symbols. (4) For video input, if the frame rate was slow, the symbols would get repeated. (5) For a live video, there was a fixed box on the screen, and the signer was restricted within its boundaries to show its symbols. Later it was discovered that there were a couple more specific technical limitations and concerns for this project: (1) The dataset of 7800 images for 26 classes (essentially 300 images per class) was too small for such a large classification problem. As transfer learning was used, it magnified this problem of the small dataset as it didn't satisfy the minimum requirements to apply this method. (2). As 2000 iterations were performed over this limited dataset, it potentially could have skewed the performance of this model by overfitting to this dataset, which might also be the cause for the earlier limitations. (3) The code was difficult to interpret and improve upon directly.

Sign language is the only available method of communication for deaf and dumb people as it helps in communicating by the use of gestures instead of speech. A sign language interpreter has to be designed in case a vocal person does not know the sign language. The sign language interpreter makes communication possible between hearing-impaired people and vocal people.

With recent advanced developments in the field of computers and Artificial Intelligence technology, an automatic interpreter has become a possibility, which can convert Speech to Text, which can be further translated to sign language gestures. The proposed software can thus be a good initiative for communication with the hearing impaired and the other people in our society, leading to more opportunities and exposure for the deaf. Systems analysis decomposes a system into its component pieces for studying how well those parts work and interact to accomplish their purpose. It is the main step in any business as firstly, we determine the requirements, available resources, system parts, and how each part interacts with another part. System analysis in information technology is widely used as we

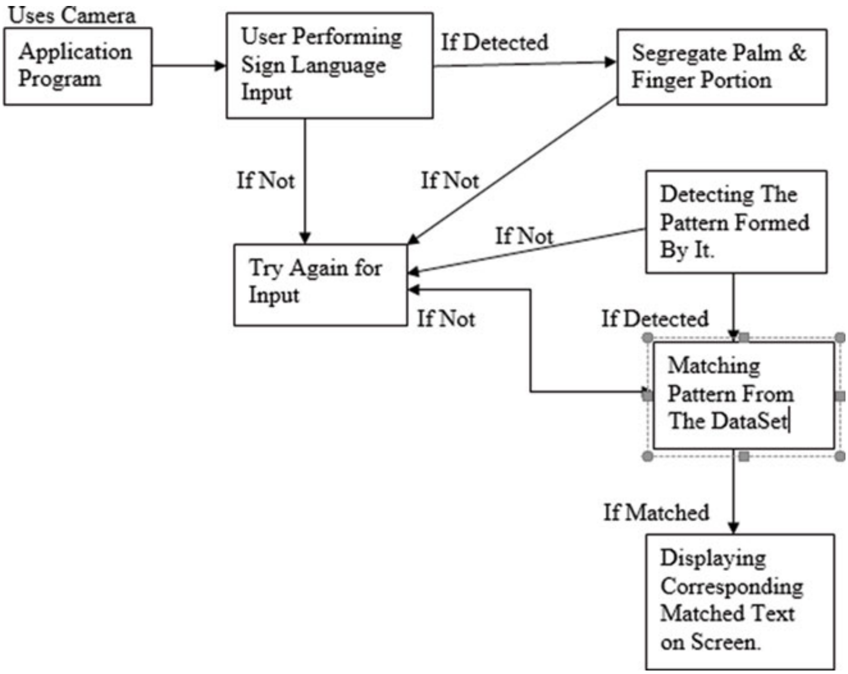


Fig. 1 Proposed software

identify the system requirements, model the system, and test it to repair errors before the final deployment of the whole system (Figs. 1 and 2).

3.1 Implementation Model

Each gesture is decomposed into three frames: contours, defects, and binary images (Figs. 3, 4, 5, and 6).

4 Results

The coding and simulations were done in a python environment, i.e., anaconda and open CV python. Once the webcam starts, the webcam feed was recorded for a fixed number of seconds. During this duration, the user made the symbols corresponding to the alphabet or numbers they wanted to sign. In this instance, the user is making the gesture for the number "1" (Fig. 7).

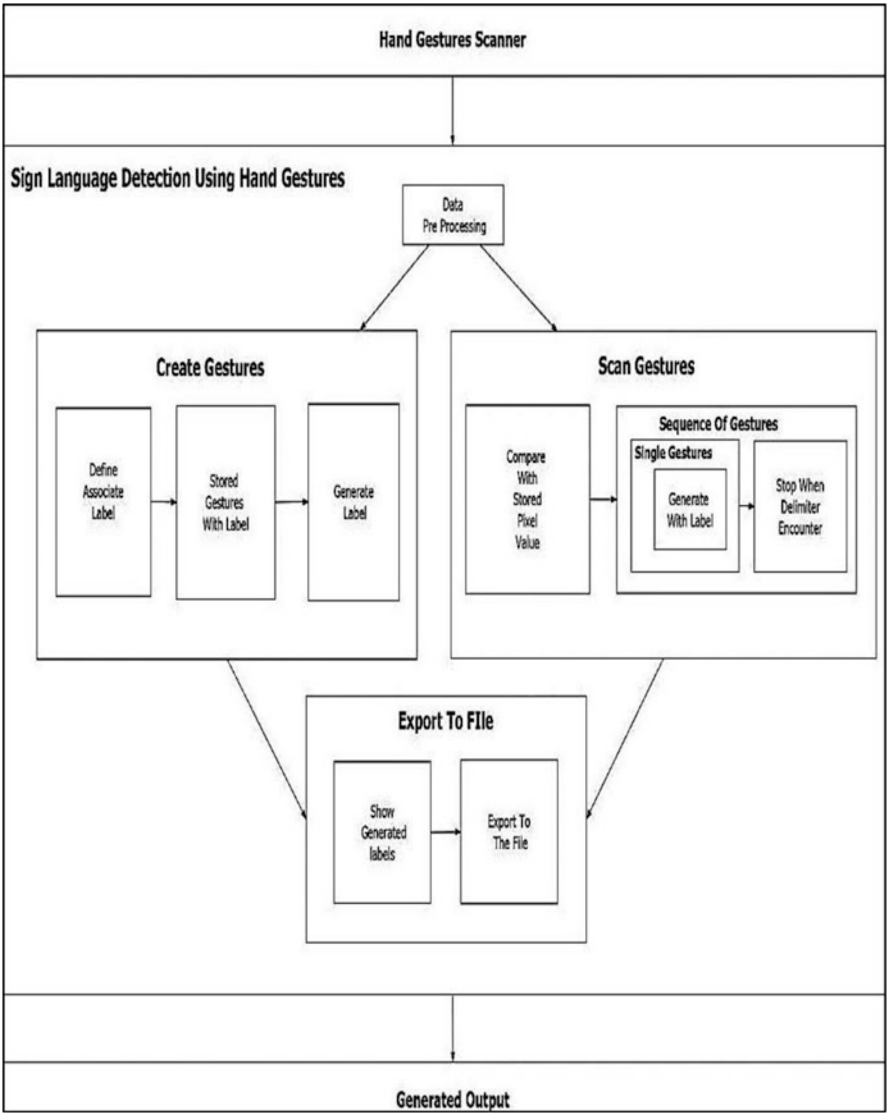


Fig. 2 System architecture

After this, the frames were extracted from this recorded clip, which went through the preprocessing techniques. The following three images represent the extracted frame, the bounded box frame, and the skin-segmented frame corresponding to the same moment (Figs. 8, 9, and 10).

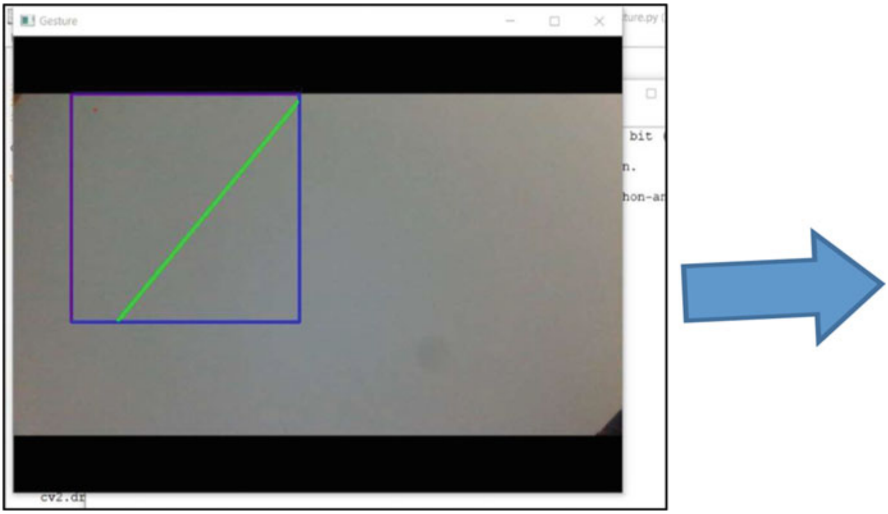


Fig. 3 Original frame

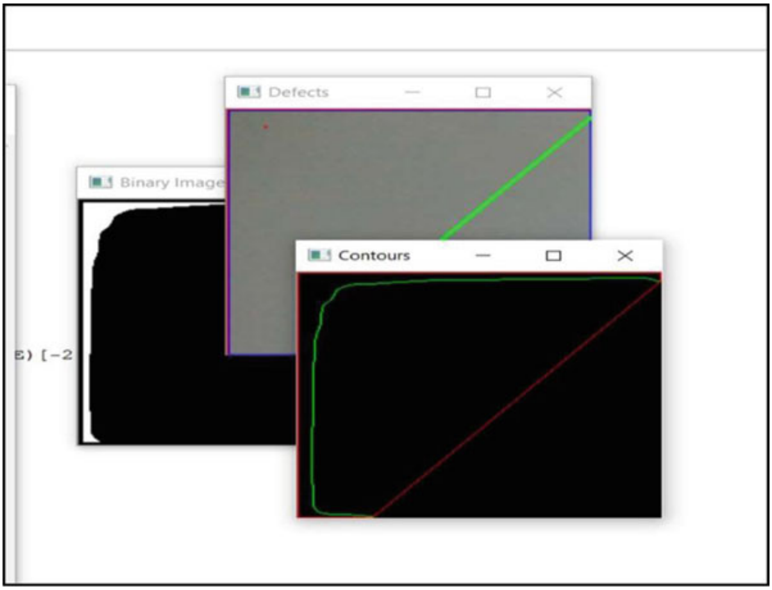


Fig. 4 Showing decomposition of frames

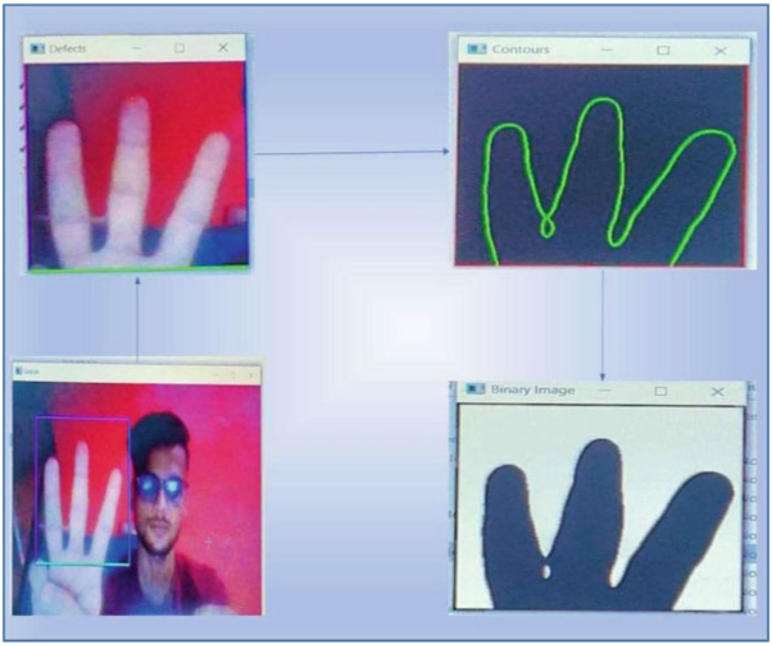


Fig. 5 Training datasets

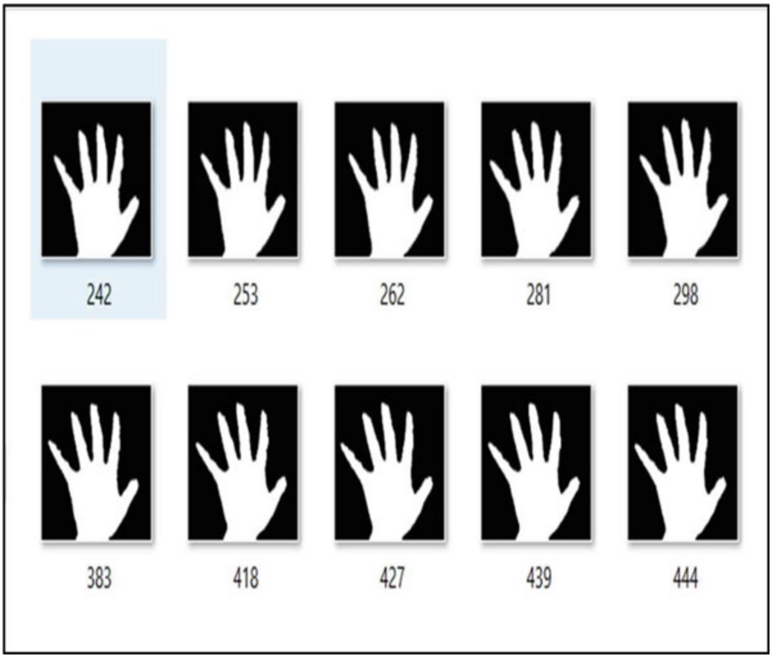


Fig. 6 Showing some datasets

Fig. 7 Gesture for number "1"

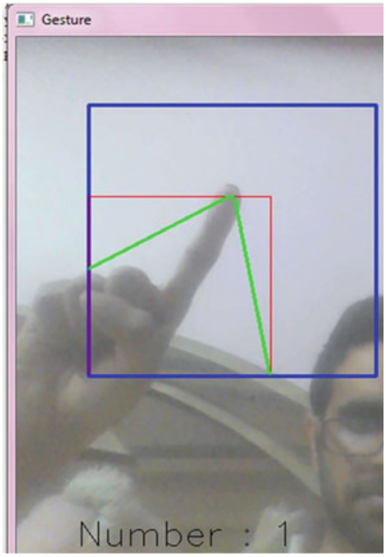


Fig. 8 Extracted Frame for gesture "1"

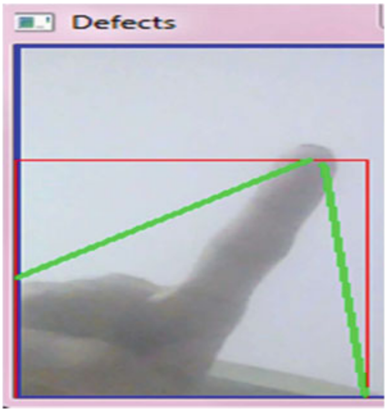
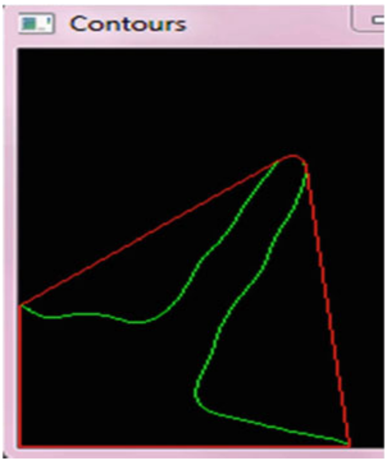


Fig. 9 Screenshot bounding box



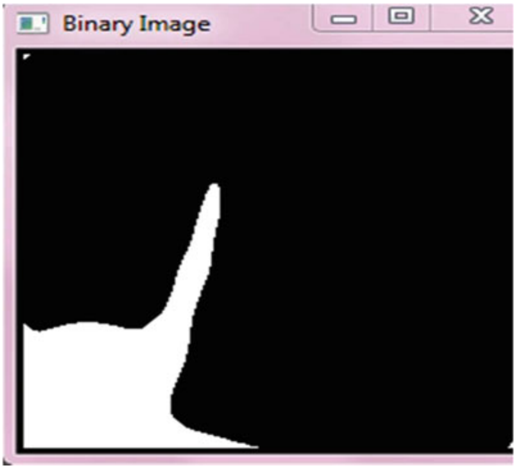


Fig. 10 Skin-generated frame

5 Conclusion, Limitations, and Future Scope

5.1 Conclusion

In this proposed solution, work has been done to improve the previous implementation of the vision-based sign language translator and tried to solve the majority of issues present in the earlier version. The proposed algorithm is a bit uneconomical but increases the adaptability of a hand gesture recognition system in near real-time. By implementing hand/skin segmentation and face detection, the technique aims to reduce the time complexity and provide better results under various degrees of scene background, complexity, and illumination conditions. This proposed solution can be used as a blueprint for achieving a more efficient algorithm that is implemented in real-time expeditiously and optimistically, helping the community of the hearing and speech impaired.

5.2 Limitations

The first and the most crucial drawback of this implementation is that the execution time of a single run is higher than the ideal time. It took approximately 41 s from start to end and a response time of 20 s (i.e., after a user has finished recording the video), as given in the example run. This period is greater than our expectations. The two primary concerns with this are setting up part of face detection and webcam feed recording phases and in the preprocessing phase. Additionally, the current implementation is a single run process, which means it runs once from start to end and then stops. However, in a practical scenario, this program should run continuously. That would also create the need for a cleanup process to manage the storage resources of the system running this program. The unconditional stop of the camera leads to the clean-up process, which, in turn, will decrease the execution time of the algorithm. Due to poor execution time, this algorithm cannot be used in applications where fast outputs are required.

5.3 Future Scope

Poor execution time can be improved by making the setup phases of the program a one-time thing to be done at the start of its execution. As for the preprocessing phase, this can be solved by using libraries like a pillow- SIMD, which can make the preprocessing steps be executed parallelly for multiple images, thereby significantly reducing the execution time and achieving a real-time or near-real-time execution. A continuously running implementation with a proper cleanup process may be incorporated to make it faster. Moreover, the whole proposed solution can also be made angle-independent. This would open up the potential for implementing this algorithm with a CCTV camera. Finally, as mentioned earlier, the idea behind this proposed solution is to use these symbols as not individual letters, but as gestures signifying phrases, terms, or messages. Hence, the current dataset which contains alphabet notational hand gestures can be converted seamlessly to ideographic notational hand gestures, thereby achieving one of the goals of this proposed solution.

References

1. J. Murray, *World Federation of the deaf. Rome, Italy* (2018). Retrieved from http://wfdeaf.org/our-work/. (Accessed 30 January 2020). MXNET (2020). MXNET. Available online: Accessed date: June 2020
2. T. Starner, A. Pentland, Real-time American sign language recognition from video using hidden Markov models, in *Motion-Based Recognition. Computational Imaging and Vision*, ed. by M. Shah, R. Jain, vol. 9, (Springer, Dordrecht, 1997). https://doi.org/10.1007/978-94-015-8935-2_10
3. L. Pigou, S. Dieleman, P.J. Kindermans, B. Schrauwen, Sign language recognition using convolutional neural networks, in *Computer Vision – ECCV 2014 Workshops. ECCV 2014*, Lecture Notes in Computer Science, ed. by L. Agapito, M. Bronstein, C. Rother, (2015), p. 8925
4. H.M. Cooper, E.J. Ong, N. Pugeault, R. Bowden, Sign language recognition using sub-units. J. Mach. Learn. Res. **13**, 2205–2231 (2012)
5. H. Brashear, T. Starner, P. Lukowicz, H. Junker, *Using Multiple Sensors for Mobile Sign Language Recognition* (Georgia Institute of Technology, Atlanta, 2003)
6. K. Imagawa, S. Lu, S. Igi, Color-based hands tracking system for sign language recognition, in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, (IEEE, New Jersey, 1998), pp. 462–467
7. C. Vogler, D. Metaxas, Parallel hidden Markov models for American sign language recognition, in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, (IEEE, New Jersey, 1999), pp. 116–122

8. E.J. Holden, G. Lee, R. Owens, Australian sign language recognition. Mach. Vis. Appl. **16**(5), 312–320 (2005)
9. NAD, *American sign language-community and culture frequently asked questions* (2017) [Online]. Available: https://www.nad.org/resources/american-sign-language/community-and-culturefrequently-asked-questions/. Accessed 6 Apr 2019
10. S.S. Rautaray, Real time hand gesture recognition system for dynamic applications. Int. J. UbiComp (IJU) **3**(1) (2012)
11. S. Jain, K.V.S. Raja, *Indian Sign Language Character Recognition*, [Online]. Available: https://cse.iitk.ac.in/users/cs365/2015/_submissions/vinsam/ report.pdf. Accessed 6 Apr 2019

Data Leakage Detection and Prevention Using Cloud Computing

Vanshika Singh, Manish Raj, Indrajeet Gupta, and Mohd Abuzar Sayeed

1 Introduction

Data leakage refers to the accidental disclosure of critical data that can cause it to be exposed to cyber-attackers and criminals which can cause huge damage to the organization and data subjects. Intellectual property (IP), monetary details, healthcare documents, credit report data, and other details related to the organization and market are examples of critical data held by corporations and businesses. Furthermore, the critical data is frequently shared across a range of consumers, including workers operating from outside the corporation (e.g., on computers), corporate associates, and consumers. Any unauthorized access to the organization-specific data can result in leak of sensitive information, which can cause certain implications for the business model itself. Software as a Service (SaaS) is defined in cloud computing as technology that acts as a middleman among both the consumer and the web and deploys via the internet. Where a provider authorizes a software to consumers as a service on request, through a membership, in a "pay-as-you-go" arrangement, or (progressively) for no cost when there are opportunities to make money from sources other than the customer, it is known as SaaS [1]. Because of this tremendous expansion of SaaS, it will soon be ubiquitous within every business establishment. It is critical that software purchasers and consumers comprehend the applicability of SaaS. In information security and computer security, cloud computing security is one of the major concerns.

It is usually recommended that Information Security Controls (ISCs) be established and deployed in accordance with and in proportion to the vulnerabilities, and repercussions [1]. Cloud computing is provisioning and assigning IT Services

V. Singh · M. Raj (✉) · I. Gupta · M. A. Sayeed
Bennett University, Greater Noida, Uttar Pradesh, India
e-mail: e19cse013@bennett.edu.in; manish.raj@bennett.edu.in; indrajeet.gupta@bennett.edu.in; abuzar.sayeed@beneett.edu.in

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_9

and resources over the internet to individuals and enterprises. The main advantage of cloud computing is you only pay for the resources that you use. You do not need to set up datacenters and operate them, which results in reducing the capital expenditure. Due to the growing popularity of cloud computing, we can find examples of many tasks that require high computing power. Examples may include Research Labs, Artificial and Machine Learning use cases, and Complex System Simulations [1]. These tasks may lead to cloud resources being unutilized, which in turn creates cloud wastage.

We observed that many people who are new to cloud computing are not handling there resources properly, which eventually leads to wastage of lots of resources and many resources being unutilized. First of all, our innovation is predicting cloud resource utilization using a machine learning model, using the compute resources dataset of specific tasks. Even the cloud platforms are currently not exploring this problem statement which makes it more unique. Apart from that, we will also predict cloud wastage utilization which will help individuals, companies, and other institutions to manage their cloud utilization systematically and cost-efficiently. we will create a new model from scratch which will help our solution be revolutionary in this sector.

The billed amount for clouds is frequently more than expected. It's because people don't understand how to handle cloud resources. People have yet to fully grasp cloud computing, and their inexperience shows. Several corporations have clouds all over the world but fail to make optimum use of the cheapest ones. It contributes to cloud waste and latency difficulties. You will suffer a greater fee if you use an Indian server for any operation in the United States. If we don't act quickly, the prices of these unutilized units will rise over time. Because of shifting demand, the cloud manager must be able to utilize resources by providing and de-provisioning resources to match the need. Inadequate resource provisioning results in Service Level Agreement (SLA) violations, poor Quality of Service (QoS), and performance deterioration, which leads to customer dissatisfaction [1]. Overprovisioning, on the other hand, results in resource waste, which raises the cost and energy.

A careful examination of the dynamic and precise resource provisioning is required for the system's smooth operation [2]. Accurate prediction should be used to determine the proper quantity of resource to meet the needs. To accurately anticipate future workload, it is necessary to use a trustworthy and exact prediction model. Typically, in Cloud Data Centers, user tasks occur in an irregular schedule with varying resource requirements. This scenario makes predicting the specific workload extremely difficult. Cloud computing is provisioning and assigning IT Services and resources over the internet to individuals and enterprises. Due to the growing popularity of cloud computing, we can find examples of many tasks that require high computing power. Examples may include Research Labs, Artificial and Machine Learning use cases, and Complex System Simulations.

These tasks may lead to cloud resources being unutilized which in turn creates cloud wastage. Accurate prediction should be used to determine the proper quantity of resources to meet the needs. Hence, we have decided to use cloud computing data to create a model which can accurately predict the usage and wastage of computing resources to demonstrate a system for predicting resource utilization in real time.

The system monitors resource consumption in real time and caters utilization value to various buffers based on the type of resource and time interval. The data in these buffers is checked to see if it follows a Gaussian distribution. Autoregressive Integrated Moving Average is used when the distribution is Gaussian; or else, Autoregressive Neural Network is used. A model is chosen in the ARIMA process based on small Akaike Information Criterion value. On the other hand, if Network Information Criterion value is low, then the Autoregressive Neural Network method is used. We tested our approach using real-time CPU utilization data from an IaaS cloud.

1.1 Motivation

Today's world is heavily reliant on data transmission, or the flow of information from one individual to another. The information sent by the supplier should be safe, secret, and unique, as the information exchanged with authorized third parties is sensitive and proprietary. Leakage occurs whenever information is accessed, viewed, or tampered with, while being uploaded to the cloud [2].

Based on the Ponemon Institute's Cost of a Data Breach Report, a yearly compilation of data leak patterns that has become something of a touchstone for the data protection sector over the years, security breaches have caused approximately 3.86 million dollars of loss in 2020. Healthcare industries have experienced the greatest expenditures connected with a data leak for the tenth year in a row. According to an IBM report, healthcare data theft ended up costing the companies $7.1 million on aggregate, up a smidgeon from previous year's figure ($6.45 million).

The energy sector, which was the second most expensive, caused businesses an average of 6.39 million dollars. According to Ponemon, sectors with more stringent legal requirements had greater information theft expenses this year. The more serious a data breach is, the more likely a company would lose revenue, which might illustrate why the healthcare, energy, finance, and pharmaceutical enterprises were among the most impacted [3].

2 Data Loss Prevention

In a cloud setting, a virtual machine may be deployed to operate a security engine that manages all the remaining virtual machines on a defined range of virtual servers using virtual machine management (VMM) architecture. Client software with a DLP engine may then be launched on virtual computers, which will scan, detect, and stop the flow of confidential data. The VMM may combine them into a single virtual

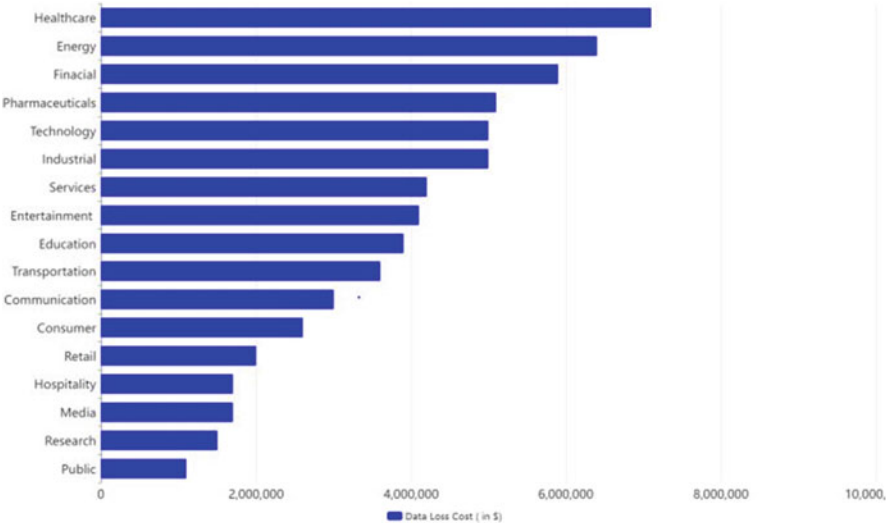


Fig. 1 Data breach cost 2020

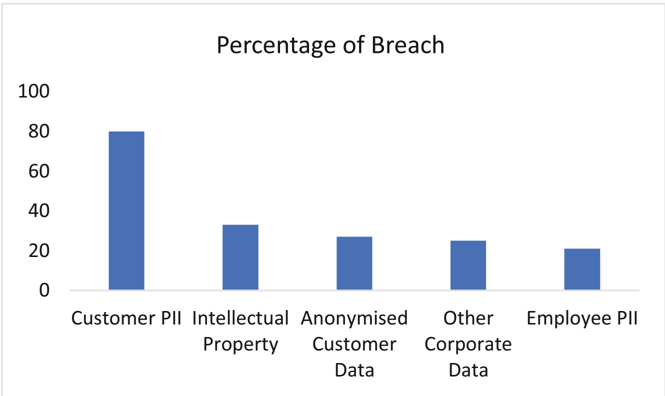


Fig. 2 Percentage of breach involving data loss

machine, allowing the DLP engine to monitor and control all the virtual machines that operate a customer, as well as observe information in transit [4] (Figs. 1 and 2).

This expands the potential for confidential information protection standards such as PCI DSS, PII, and others. DLP is a technology that may be turned on or off for virtual machines in the cloud data center. A DLP system, like a cloud infrastructure, is dynamic, since it can be extended and operated [5]. APIs can be used to orchestrate controls in a DLP solution, such as creating a regulation that moves a virtual machine with confidential information behind a firewall or puts it on shutdown. Some of the key advantages of major cloud DLP solutions are listed below:

1. Develop with cloud storage services to inspect systems for confidential information, detect it, and encrypt it before it is transmitted in the cloud
2. Analyze and verify information that has previously been saved in the cloud at any moment
3. Retrieve confidential information from the cloud with accuracy. Inspect transmitted documents on a regular basis
4. Instantly deploy restrictions to confidential information in line with corporate policy (prompt, block, or encrypt)
5. Whenever the information is exposed or an unauthorized access is detected, immediately notify the relevant authorities and data proprietors
6. Retain the transparency and authority required to adhere to confidentiality and information security laws

2.1 Limitations of Cloud DLP

If the cloud infrastructure is public, each instance may only have one network connection, necessitating the use of a virtual DLP version that may detect, transmit, or prohibit traffic with restrictions. While utilizing DLP to detect data migration to the cloud and for content discovery on cloud-based storage has a lot of value, adopting DLP in a public cloud may not. The application architecture, which relies more on application security and encryption, is perhaps the most secure part of any cloud implementation in accordance with DLP. DLP is a fantastic tool for improving information security in the cloud. Given that it is adjusted appropriately, it may be utilized to monitor information transferring to the cloud, find critical data stored on the cloud, and safeguard cloud-based applications.

3 Related Work

3.1 Watermarking Technique

Watermarking, in which a distinctive code is included in each disseminated duplicate, has historically been used to identify breaches. The leaker can be determined if that duplicate is subsequently acquired in the possession of an unauthorized person. Watermarks are effective in some situations, although they do need some change of the initial material. Moreover, if the data consumer is hostile, watermarks can be removed [6]. A healthcare institution, for example, may provide patient details to researchers who may develop new medicines. Likewise, a business may form alliances with other businesses that need the exchange of client information. Because other company may subcontract data processing, data must be handed to a number of different firms. The data proprietor is referred to as the distributor, while the ostensibly trustworthy third parties are referred to as agents. Watermarking Technique [7].

3.2 Fake Objects Method

Fake objects may create less issues in some applications than genuine things. Consider health documents as dispersed data objects, and hospitals as the agent. In this situation, even slight alterations to the data of genuine patients might be hazardous [8]. The inclusion of certain fictitious health documents, on the other hand, may be allowed because no patient resembles this information, and so no one will ever be diagnosed relying on fictitious information. In this scenario, corporation A provides a mailing list to corporation B that will only be utilized once (e.g., to send advertisements). Corporation A adds trace data with corporation A's addresses. As a result, whenever corporation B utilizes the acquired mailing list, duplicates of the mailing are sent to A. These records are a form of fictitious object, that aids in the detection of data misuse [9].

3.3 Adversary Model

This approach captures all types of information security concerns, and cloud data is not designated at the cloud customer level, but rather at the cloud service level [10]. There are two forms of attacks capable of causing this:

1. Internal Attack: The cloud service provider might not be trustworthy
2. External Attack: The attacks launched by unauthorized users

4 Secure Hashing Algorithm (SHA)

SHA is an abbreviation for Secure Hash Algorithm [11]. Hashing techniques are computationally intensive routines that squeeze, encode, and encrypt the ingress data and generate hash or hash values, which appear to be random. These seemingly random numbers are really the input data in an encrypted or coded format. Hash values of data are relatively easy for computers to deploy than the actual data because hashing allows the system to execute numerous procedures or calculations over directories and data chunks. A hashing algorithm must be predestined, which means it must return symmetrical results for each ingress value. The SHA secure algorithm provides a hash function during data transfer with a reference to a third-party agent and the actual cloud; this authentication mechanism is introduced to help secure the transfer of essential information, over the internet. The SHA algorithm is utilized because message digests are needed to ensure data security and authenticity. In this method, the data is secured using SHA-2 (Secure Hash Algorithm 2), a subset of the cryptographic primitive family Keccak [1] SHA-2 was designed with an internal block size of 1024 and can perform And, Xor, Rot, Add (mod 264), Or, Shr, and contains 256–128 security bits [12].

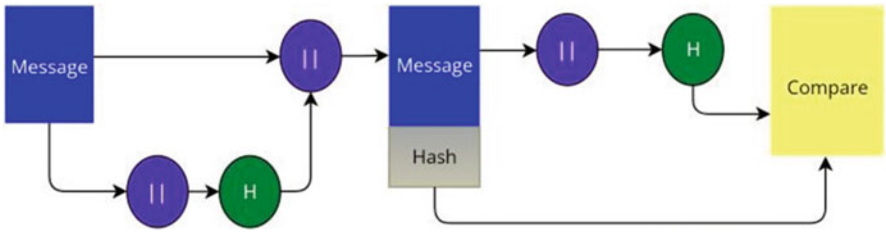


Fig. 3 Hash function working

The Third-Party Auditor (TPA) ensures performance, integrity, and impartiality in necessary auditing operations and acts as a negotiator between CSUs and CSPs [13]. Corporations and enterprises do not have their own storage places for their databases and other confidential information in today's reality. They procure disk space from other entities on bill per use basis. TPA serves as an administrator in this case, governing user access control and other administrative rights such as uploading incoming data or modifying current data to entities in the data communication network (Fig. 3).

5 Proposed Model

The proposed methodology addresses the persistent problem of data leakage by providing a conceivable attestation against the rogue operator and the unauthorized data released by the same. The modules included in the proposed model are listed below.

5.1 Data Allocation

This approach emphasizes on a secure and ingenious methodology for a distributor to distribute data to its stakeholders to identify the "guilty" workforce. With administrative credentials, a checked-in user can rescript and alter their files and upload it to the cloud. Admin sends the necessary info to the user and the authentication data is sent between agents and users through email.

5.2 Fake Object Module

To discover the "site of leakage," the data distribution mechanism will append redundant information to the credible data stream. Fake or redundant objects are objects that appear to be credible data being delivered, but are different and differentiable from real records. The data leaker believes he has the actual record due to a false item. It will show the bogus record and deliver the message if any record is downloaded using counterfeit measures using the "wrong key." The term "wrong key" relates directly to the duplicate key known as Public Key, which is shown as the primary key. It safeguards the private key (the primary key).

5.3 Optimization Module

This is a mandatory module for the data distribution entity. The entity has binary objectives that it must achieve. These are:

1. To give User-requested data and to conduct the User-requested changes
2. To recognize information breach and discover the root of the incident

5.4 Data Distribution Module

A data distributor is tasked with transmitting confidential information to one or more of his entrusted stakeholders or workers. He must guarantee that none of this information gets disclosed. However, if it is released, it is rumored that admin will find out about the particular files which are being disclosed. Also, for picking out the guilty employee distributor, data must be analyzed and scrutinized, assuming that it has been disclosed by one or more of its own employees. The information that is transmitted might be in any form or magnitude.

The entities and flow of data for constituting the cloud computing environment, as well as the data flow paradigms, are depicted in the architectural Fig. 4.

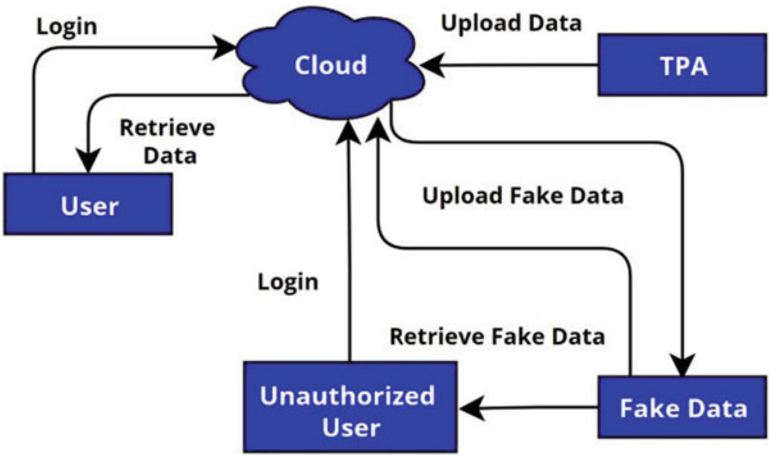


Fig. 4 Proposed architectural diagram of data leakage detection

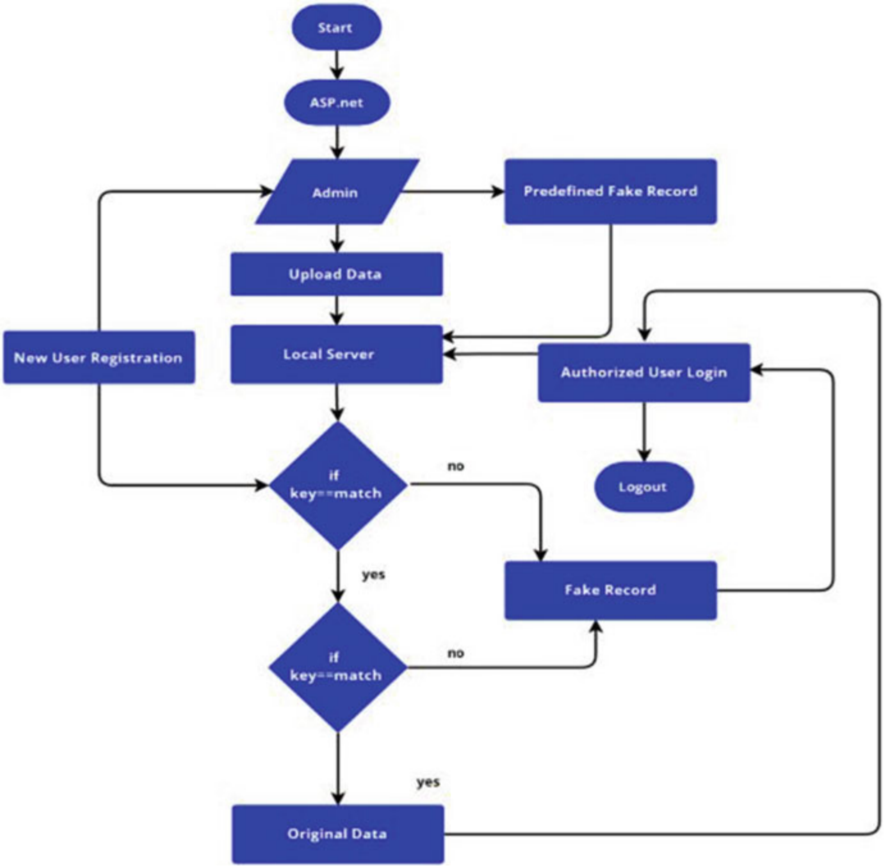


Fig. 5 Workflow of proposed diagram

The suggested model's control flow is depicted in the flow diagram in Fig. 5. The new customer creates an account and then uploads their information to the local server with administrative capabilities granted by admin. For safety reasons, the admin concurrently regulates the placement of pre-defined bogus information on the local server. Any information delivered to a receiver now includes the phony redundant information that was appended or concatenated to the authentic data using Wrong Key. If someone gains access to the transmission apparatus or channel and holds the correct key to decipher the transmission (the recipient), will be facilitated with access privileges to the original transmission. However, if an unauthorized access is detected, the information sent to the customer will be bogus owing to a key match error.

The proposed methodology can be accomplished by asymmetric encryption techniques. Asymmetric cryptography facilitates the encryption and decryption of information using two keys: a private key and an original key [3]. The public key, on the other hand, is utilized to mislead unlawful access into assuming they have the original key. This approach is also faster in terms of computing. Information integrity is improved using hash functions and hashing algorithms.

6 Conclusion

The presented paper may be ended with the observation that the supplied model includes an approach to the concerns of information leakage and information transmission safety. The employment of data encrypting methods results in data being more reliably encoded, as well as the keys included inside it. Through the assistance of redundant objects functioning as outlandish marking over the authentic data, the culpable employee inside the company or any trustworthy entity of the corporation may be in the information communication network of the corporation. Also the corporation must investigate what information has been released by the responsible party. As we've seen, this suggested approach improves and extends information leakage tracking and protection, and even considers the potential of what information has been compromised. This methodology also solves the challenge of locating the data source and perpetrator of the leaks in the institution's information transmission network. Because hashing methods like SHA-2 are used in the suggested paradigm, the computational performance for encryption and decryption of information to be transferred or acquired is also faster. Though the presented model has solved the fundamental elements of the problem, more advanced and efficient algorithms and approaches can be formulated in the future. For example, in this methodology, we used the SHA-2 method, which is more efficient and safer than the SHA-1 algorithm, but there is also a variant called SHA-3 that may be employed in other methodological approaches.

References

1. R.G. Pearson et al., SPECIES: A spatial evaluation of climate impact on the envelope of species. Ecol. Model. **154**(3), 289–300 (2002)
2. M.S. Darms et al., Obstacle detection and tracking for the urban challenge. IEEE Trans. Intell. Transp. Syst. **10**(3), 475–485 (2009)
3. C. Tan et al., Understanding the nature of first-person videos: Characterization and classification using low-level features, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (2014), pp. 535–542
4. K.V. Wong, Research and development of drones for peace—High power high energy supply required. J. Energy Resour. Technol. **137**, 3 (2015)
5. H.U. Zaman et al., A novel design of line following robot with multifarious function ability, in *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*, (IEEE, New Jersey, 2016), pp. 1–5
6. O. Shrit et al., A new approach to realize drone swarm using ad-hoc network, in *2017 16th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, (IEEE, 2017), pp. 1–5

7. M. Bennis, M. Debbah, H.V. Poor, Ultrareliable and low-latency wireless communication: Tail, risk, and scale, in *Proceedings of the IEEE* 106(10) (2018). Accessed on 16 Nov 2021, pp. 1834–1853
8. Rotor Riot LeDrib. *What is FPV Freestyle* (2018). https://rotorriot.com/pages/beginners-guide. Accessed on 16 Nov 2021
9. Liftoff Drone Simulator (2018) https://www.liftoff-game.com/liftoff-fpv-drone-racing. Accessed on 17 Nov 2021.
10. F. Naujoks et al., From partial and high automation to manual driving: Relationship between non-driving related tasks, drowsiness and take-over performance, in *Accident Analysis & Prevention*, vol. 121, (2018), Accessed on 16 Nov 2021), pp. 28–42
11. Get FPV Aaron Ziemann. *How to Find the Perfect Drone Racing Line* (2019). https://www.getfpv.com/learn/fpv-essentials/how-to-find- the-perfect-drone-racing-line/. Accessed on 16 Nov 2021
12. J. Delmerico et al., Are we ready for autonomous drone racing? The UZH-FPV drone rac- ing dataset, in *2019 International Conference on Robotics and Automation (ICRA)*, (IEEE, New Jersey, 2019), pp. 6713–6719
13. A. Loquercio et al., Deep drone racing: From simulation to reality with domain randomization. IEEE Trans. Robot. **36**(1), 1–14 (2019)

To Brace Society 5.0: Enhanced Reliability with a Cost-Effective Protocol for Underwater Wireless Sensor Network

Ankur Sisodia, Swati Vishnoi, and Ajay Kumar Yadav

1 Introduction

Basically, the purpose of Society 5.0 is to enhance the capability of devices developed in Industry 4.0. During the revolution of Industry 4.0, many devices were developed that would be helpful in the development of the life of human beings. The main objective of Society 5.0 is to make the world an innovation hub so that researchers contribute to the development of human society through their different researches. Society 5.0 makes the world economy way stronger by its innovation. Society 5.0 works to optimize the applications developed in Industry 4.0 by making space so that the user can use these devices in an efficient manner which is developed with the help of different computer science techniques like Artificial Intelligence, IoT, Blockchain, etc. [1]

Society 5.0 focuses on enhancing the quality of human life by implementing new technologies. There are many areas in the field of underwater wireless sensor network like underwater wireless communication using wireless sensors. For the efficient usage of underwater sensor networks, effective protocol is needed which must be cost-effective. Every human being wants a cost-effective solution in society 5.0. Here in this paper we are providing a solution to this problem with reliability enhancement.

There has been a growing rapid transmission of the interest for Wireless Sensor Networks (WSNs) into embedded electronics and wireless technologies. The WSN consists of nodes with limited computer capabilities and power because of intelligent sensors and built-in CPUs. Sensor nodes track humidity, pressure, heat,

A. Sisodia (✉) · A. K. Yadav
CSE Department, Banasthali Vidyapith, Aliyabad, Rajasthan, India

S. Vishnoi
CSE Department, Sanskriti University, Mathura, India

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_10

vibration, and many other factors. Wireless sensor network nodes have basically three units: a computing unit, a sensor interface, a transceiver unit. More critical tasks are performed by the sensor nodes, which makes communication between the nodes to pass the data between the sensors. The communication is mostly carried out in a distributed setting, and therefore a central network is required to improve the message passing among the nodes in the wireless sensor network. WSN with IoT provides faster access and more feasible data of surroundings. This might increase the effectiveness and efficiency of WSN.

The main component in IoT is WSN and it assists with offering normal types of assistance and teams up with it when it's anything required in a distant access heterogeneous framework in an ad hoc climate. In any case, coordination of IoT with WSN isn't only a theory, yet it is upheld by global organizations and guarantees a better turn of events and testability of the framework. It assists with connecting the information created by the sensor nodes with informing highlights, interpersonal organizations, and online administrations.

Routing [2–5] is a significant factor of WSN correspondence, which empowers a viable correspondence to divide the data among the nodes in the UWSN. The primary focal point of the paper is to give a compromise between the network lifetime and energy utilization, which is assessed as far as Signal-to-noise-ratio, bit-error-rate, and packet-delivery ratio.

- In this paper, an adaptive routing protocol is proposed which is intended to keep up with the trade-off between the dependability of packets in UWSN and organization lifetime (SNR, BER) utilizing an ideal adaptive routing protocol.
- The whole system worked in three distinct stages, in particular, initialization, dynamic directing, and transmission.
- Further, the routing is set up utilizing an ideal coordinated non-cyclic diagram adaptive routing protocol that assists with choosing the route among neighbor and replacement nodes.
- Finally, the expense work with coordinated non-cyclic charts is used to discover the nodes with higher remaining energy for better transmission of bundles.

2 Literature Review

Numerous underwater WSN conventions have been proposed in ordinary strategies, some of which are examined in this part.

In [6], they raised the issue regarding topology control in UWSN; for this, they proposed an algorithm that is based on a topological model with three layers as follows: partition unit, cluster, and temporary control node. The algorithm is designed to take into consideration the vulnerability of sensor hubs due to their outer components. The disappointment hub, inclusion weakness, inclusion network, key position, and valuable hub are characterized in the proposed calculation. The calculation likewise utilizes the inclusion network and the weakness edge hubs to decide if the overlay weakness requires a fix.

In [7], they propose a steering convention called the Layer-based Energy-Efficient Routing (LEER) convention to address the course disappointment issue with avaricious directing, just as the long start to finish deferral and high energy utilization issues. In LEER, every hub extricates the layer field data from high parcels and refreshes its own layer to stay away from the issue of steering a bundle to avoid region. All hubs with the LEER convention forward parcels to a sink hub without the requirement for any area data.

In [8], they give a thorough review on the difficulties, advances, and prospects of underwater optical remote organizations from a layer-by-layer viewpoint which incorporates the following: (1) physical layer issues including spread qualities, channel demonstrating, and balance methods; (2) data interface layer issues covering join setups, interface financial plans, execution measurements, and various access plans; (3) network layer themes containing transferring strategies and potential directing calculations; (4) transport layer subjects like availability, dependability, stream and clog control; (5) application layer objectives; and (6) Localization and its effects on layers.

In [9], another clog control calculation for Wireless Sensor Networks is proposed. The current calculations for this issue have high intricacy and force utilization because of retransmission with clog control being done by tracking down the ideal rate through a straightforward Poisson measure. Retransmission of the impacting parcels causes wastage of energy since the remote sensor network has a restricted battery. The new calculation is executed in MATLAB R2016a and thought about against the current Cuckoo Search (CS) and Adaptive Cuckoo Search (ACS) calculations. Reproduction results demonstrate that the proposed system has preferable outcomes over the current methodologies.

In [10], subsequent to introducing an outline of potential UASN applications, a review of the arrangement methods and confinement calculations for UASNs has been introduced, depending on their significant benefits and weaknesses. At long last, research difficulties and open examination issues of UASNs have been talked about to give an understanding into future exploration openings.

In [11], an energy-productive versatile grouping steering calculation for underwater acoustic sensor networks is proposed (ACUN). The calculation utilizes a staggered progressive organization structure dependent on the distance between group heads and the sink hub and the remaining energy of a bunch head to decide the size of the opposition span. It can keep away from the early passing of a group head away from the sink hub due to extreme contest range prompting over-the-top energy troubles. While choosing a group head, thinking about the hub-remaining energy and the energy loss of a transmission way, the calculation can choose a hub with bigger leftover energy as a bunch head while improving organization energy utilization.

In [12], they characterize the new coordinated idea of the remote sensor, actuator, and robot organizations, study the present status-of-craftsmanship in the field, and present plan necessities and open examination issues.

In [13], they attempt to broaden the exhibition of Wireless Sensor Network Routing Protocols by presenting new component data compression into it and we break down this proclamation by contemplating these performance parameters like throughput, routing overhead, and end-to-end delay with data compression as a new component.

In [14], they attempt to concentrate some non-anecdotal applications so we can dissect in future how these applications work with wireless sensor networks; for this, they briefly characterize Wireless Sensor Network to bring everything together and for the most part center around the investigation of non-anecdotal applications like patient data in hospital, tracking: looking and deciding area, Context-Aware, and Retailing: deals and administration support. Take considerations about these utilizations of Wireless Sensor Network so they can be utilized in a productive way by both client and engineer.

Versatile hop-by-hop vector-based sending [15] takes out invalid focuses, fixated on the followed field execution sub-block decoration, and a straight framework-blended number advancement. It limits the measure of submarine sensors mounted inside the objective field, ensuring the essential control dependability and admittance to the organization.

3D UWSN Deployment with heuristic-based blended whole number direct program improvement [16] guarantees, as far as possible, for stream harmful substances. This ensures full recognizable proof of compound starting points for poisons, while diminishing expenses of conveyance. 2D-UBDA picks, by mulling over the cross-area between the plausible utilization zones of all discharge beginning downstream of the objective field arrangement, various subzones in which substance sensors are being introduced dependent on the whole number straight planning calculation.

A 2-Dimensional Underwater Barrier Deployment Algorithm is demonstrated in [17] where the underwater source isotropically transmits acoustic impedance and radiation and utilizes the sign gathering data from the diverse UWSN sensors to effectively estimate the beginning area. Then, we pick the node whose force is the most elevated reference node as per our previous work, which parts the energy yield of the sensor into sets. Second, the most starting point step approach is confirmed, and another definition focused on the position assessor for the most elevated likelihood beginning is given. Second, to cover the non-raised issue into the arched streamlining issue, a semi-clear programming approach is created.

A curved improvement issue with a one-venture least-square strategy [18] is working to broaden UWSNs' lives. The BEAR convention recommended is in three stages: (1) statement, (2) working of the rundown, and (3) move of data. The two nodes share data with respect to their leftover energy level and position during the instruction cycle. During the tree development stage, our arranged BEAR exploits the area data to pick adjoining nodes and to pick supporting and resulting nodes relying upon the money-saving advantage.

A versatile bounce-by-jump vector-based sending [19] prevents invalid nodes from utilizing the middle way and weighted the RLS, empowering every heading to be the immediate course for UWSNs.

A middle and turned around switch limitation conspire [20] joining direct and hand-off sending designs to move bundles from beginning to the objective. The strategy for handing-off requires choosing the right transfer in a progression of hand-off nodes. The ideal hand-off measure is a hand-off node that is less distant from the beginning to the objective and has a base number of adjoining nodes.

Energy proficient, obstruction, and course mindful convention (EEDBR) [21] decrease the quantity of sending nodes and delays the presence and use of the organization.

An Energy-Efficient Depth Based Routing [22] further develops the UWSN dependability, network presence, and execution. The framework remembers versatile sink adaptability for a way that falls into the densest piece of the organization. The movement of the sink to a high-thickness area permits ideal information assortment.

Dynamic Sink Mobility Equipped Depth-Based Routing [23] builds the parcel conveyance proportion just as it saves restricted-energy by the ideal task of part nodes with GN. Conversely, the AUV movable stay stretch diminishes the bundle fall proportion and in this way augments network effectiveness [24].

ARCUN [25] expands single-receiving wire node cooperation effectiveness. This uses the remote medium's communicated nature and sends the general sensor nodes as transfers together.

A blunder control and change strategy [26] works on the precision of the sensor position. We use distance information between nodes to make a model of the organization mistake change that builds limitation exactness and ensures a steady position blunder.

Energy-effective chain-based directing convention [27] deals with the heap on nodes. The connection dependent on distance is centered on position-cognizant nodes and can be utilized in checking spaces in a consistent state yet in a liquid framework and accordingly free correspondence of area is required.

An open CSMA-based support learning approach [28] utilizes a preparation calculation and a various access framework open transporter sensor. The proposed framework will refine its boundaries to adjust to the underwater condition whenever it has been carried out, because of its further development and preparation calculation.

A flaw versatile limitation plot [29] offers great position accuracy and diminished overhead communication. A sensor node predicts its area, by concentrating on how its neighbors are portable through numerous straight relapses, to keep away from an anchor node blunder.

The nodes in UWSNs are furnished with limited battery limits, as indicated by [30, 31], battery replacement isn't possible as a result of the cruel underwater environment. Accordingly, to broaden network's lifetime in a savvy climate, the directing convention ought to be energy effective [1, 32–40], as in shrewd urban areas.

All the above techniques give information dependability, energy productivity, and get directed in underwater sensor organization. Notwithstanding, as far as we could possibly know, the utilization of IoT to work on the correspondence between the sensor nodes isn't tended to, which is the center point of the proposed strategy.

3 System Model

The unwavering quality of connections on underwater IoT correspondence is assessed utilizing a channel model, which is put in an underwater situation. The fundamental point of the model is to gauge the effective conveyance of packets over each UWSN connection. The effective conveyance of packets to the objective node gauges the parcel's unwavering quality and this is considered as an invaluable one for displaying a dependable convention for correspondence utilizing our model. There are a few components in the underwater channel model, where the channel is partitioned into two principal parts. At first, the connection between sending force and SNR is displayed and afterward, the connection between Packet Delivery Ratio (PDR) and SNR is demonstrated. This assists with figuring the unwavering quality of packets sent over IoUT links.

3.1 Transmission-Power and SNR

The fundamental goal is to find out the relationship between the Signal-to-noise ratio and TP. The signal-to-noise ratio depends on four parts noise-level, directivity-index, transmission-loss, source, which is given by

$$\gamma = \mathrm{Source}_{\mathrm{level,dB}} - \mathrm{TransLoss}_{\mathrm{loss,dB}} - \mathrm{Noise}_{\mathrm{level,dB}} + \mathrm{DirIndex}_{\mathrm{index,dB}} \tag{1}$$

SNR equation is as follows:

$$\gamma = \mathrm{Source}_{\mathrm{level}} - \mathrm{TransLoss}_{\mathrm{loss,}} - \mathrm{Noise}_{\mathrm{level}} + \mathrm{DirIndex}_{\mathrm{index}} \tag{2}$$

Equation for SNR and transmission-power is as follows:

$$\gamma = 10\ \mathrm{loglog}\ (\mathrm{Power}) - \mathrm{loglog}\ 4\pi r^2 - \mathrm{loglog}\ 0.67 \times 10^{-18} - 20\mathrm{loglog}\ d + 18\mathrm{loglog}\ f \tag{3}$$

3.2 Signal-to-Noise Ratio and PDR

Another goal of the channel model is to assess the PDR over IoUT. For this, BPSK modulation and Rayleigh fading is utilized and that incorporates multipath effects in shallow and deep water.

Hence, Bit-Error-Ratio is defined as

$$\mathrm{BER}\,(\gamma) = 0.5\left\{1 - \sqrt{\frac{10^{0.1\gamma}}{1 + 10^{0.1\gamma}}}\right\} \tag{4}$$

PDR with size m bits is calculated as

$$P_{\mathrm{success}^m}\,(\gamma) = \{1 - \mathrm{BER}\,(\gamma)\} \tag{5}$$

$$P_{\mathrm{success}^m}\,(\gamma) = \left\lfloor 0.5 + 0.5\sqrt{\frac{10^{0.1\gamma}}{1 + 10^{0.1\gamma}}}\right\rfloor^m \tag{6}$$

4 Proposed Model

The proposed model determines three steps for the packet transmission, which are named as, Initial Phase, Graph Building, and transmission. All three steps are given below.

4.1 Initial Phase

Firstly, the sink node communicates the packets to imply different nodes with respect to the complete accessible nodes, and then it advises the beginning and stopping time of the initial phase. The nodes further gauge the accompanying

- Comparative area
- Comparative separation from the sink node
- Identification of the area where the nodes are found

4.2 Graph Building

The development of the graph is done in two phases. At first, the whole nodes discover the neighbor node and lastly, it picks its replacement to work with the nodes from its local nodes.

Find Neighbor Nodes The lower profundity node is looked inside its area and the root node is set up towards the sink. The nodes are designated with a scheduled

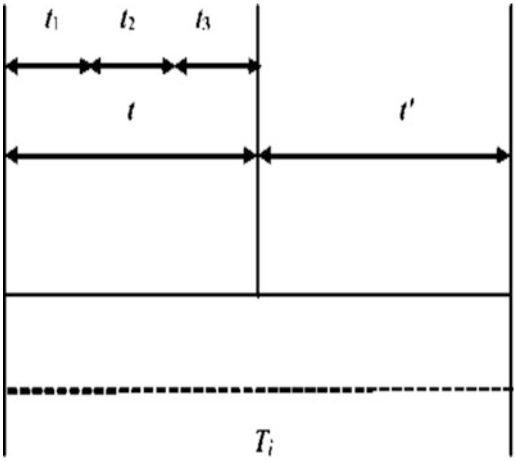


Fig. 1 4-way handshake time slot

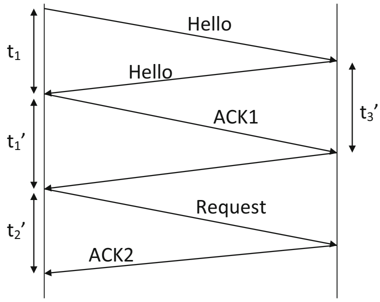


Fig. 2 Mechanism of handshaking method

opening dependent on the sink for better correspondence, for example, 4-way handshake time allotment (Fig. 1).

The scheduled opening (Ti) is separated into two portions with a most extreme length equivalent to four-way handshakes. Here, the placement of nodes is done in an irregular manner and henceforth the node doesn't have a neighbor and during that time, a second time allotment (t') is utilized for disclosure of neighbors.

At first, a transmission hello packet is communicated with the transmission range inside the area. The nodes within a range answer with an ACK1 message and arrangement for a node is made to choose just one neighbor from the area of next-hop that lies at lower profundity (Fig. 3). When the ACK1 is received, the node I communicated the neighbor demand to nodes, which has answered with an ACK1 message. Further, the planned nodes were acknowledged by sending ACK2 packet (Fig. 2).

This is done while the character of the sensor node is recorded inside its area. When the ACK2 is obtained, the characteristic of the adjoining node is stored, this reacts with the assistance of the ACK2 message, and it acknowledges the sink node with a handshake.

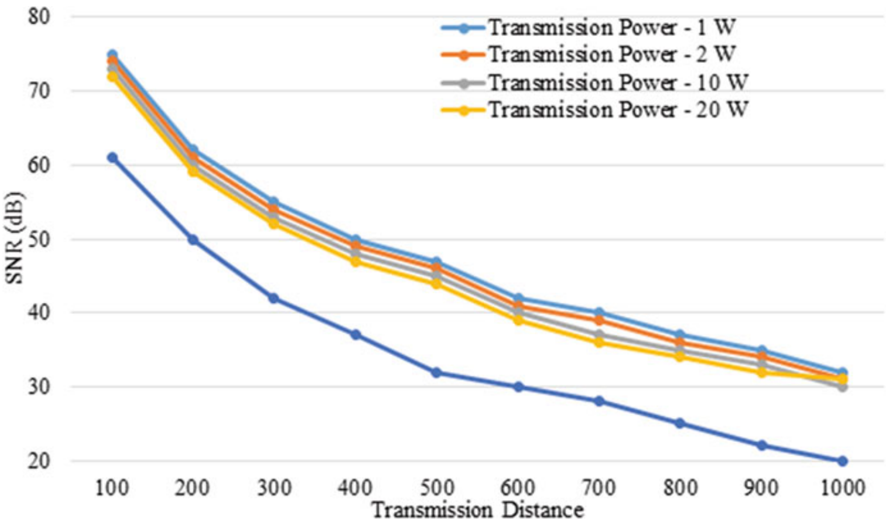


Fig. 3 Results of SNR by applying channel models with different transmitter powers and distances

Energy-Efficient Path In DAR, the whole correspondence is directed by the relative and aiding nodes. Consequently, the primary undertaking is to choose the relative and aiding nodes, since it devours less energy for transmission. Thinking about a cost function (Q_{ij}) when a node i communicates with node j, and α is the used energy of a node then, at that point, the cost function is given by

$$Q_{ij} = \alpha \times d\,(i, j) + (1 - \alpha) \times d\,(j, \sin\ k) \tag{7}$$

The cost function is determined by the broadcasting work for the adjoining nodes. After the assessment of cost function, the cost function is figured out in an ascender request, where the node id having a place with the first worth is considered as the descendent node and the node id having a place with second worth is considered as aiding node.

At each round, the transmission node gauges the separation from the sink and when the distance is not exactly the transmission range, the packet is communicated straightforwardly to the sink. On the opposite, when the node lies outside the sectoring range, then, at that point, the relative energy of the relative node is checked with average comparative energy and the cost of near energy is more prominent than average comparative energy, then, at that point, the bundle is communicated to the relative node. In the event that the second checking of distance is fizzled, the relative energy of the aiding node is checked by the transmission node and it communicates the information parcels to the aiding node. In any case, if the node lies inside the reach, it is communicated straightforwardly.

4.3 Transmission Phase

The distance of the transmission node is checked at standard stretches from the sink and when it is less than the edge range, the immediate transmission is conceivable towards the sink. Then again, when the distance is more prominent than the edge range, then, at that point, the similar energy of the next node is checked. At the point when the energy of the neighbor node is more noteworthy than normal relative energy, the replacement gets the parcel. Notwithstanding, when there is a disappointment in the second check, the transmission nodes gauge the relative energy and after gathering the necessary condition, the bundle is communicated towards that node or it is communicated straight over the sink.

5 Results

The performance assessment is led utilizing C++ and the underwater sensor is reproduced over Rayleigh Fading channel utilizing BPSK adjustment. The boundaries are set, the transmission power is set between 1 and 30 W, the recurrence is set as 10 kHz, and the parcel size is set as 3 KB. At last, the presentation measurements are estimated as far as SNR, BER, and PDR.

5.1 Evaluation of SNR

The after-effects of SNR are displayed in Fig. 3, which plots the BER as a component of distance that changes between 100 and 1000 m and the transmitter powers are set somewhere in the range of 1–30 W.

The outcome shows that at whatever point there is an expansion in transmission distance, the SNR decreases. This is predominantly because of misfortune in transmission, which is considered a significant fundamental of the SNR. With expanding distance, there is an aggregation of transmission misfortunes and this makes an adverse consequence on SNR. Henceforth, in underwater WSN, the transmission distance is conversely relative to the SNR and this is valid if there should be an occurrence of different transmission powers. This affirms that the proposed channel model for assessing the SNR is legitimate for different transmission powers.

Also, the increment in transmitter power adequately expands the SNR esteem. This case is valid as the distance of transmission is the same, the force of the transmitter from low to high SNR is 1–30 W. Since, source-level force is considered as the main consideration because of SNR, the source level force may turn out to be high when the transmitter power is high. Henceforth, it very well may be respected that the sent force is straightforwardly related with SNR, since the SNR esteem is predictable for all transmission distances and it at long last affirms that the proposed model is material for different distances.

5.2 Evaluation of BER

The consequences of BER are displayed in Fig. 4, which plots the BER as a component of distance that differs between 100 and 1000 m and the transmitter powers are set somewhere in the range of 1–30 W.

The outcome shows that transmission distance and the BER are directly proportional to each other. This is because of misfortune in transmission, which is considered a significant fundamental of the SNR, which builds the BER. With expanding distance, there is a gathering of transmission misfortunes and this makes an adverse consequence on SNR, which prompts an expansion in BER. Subsequently, in underwater WSN, the transmission distance is conversely corresponding to the SNR and this is valid in the event of different transmission powers. This affirms that the proposed channel model for assessing the BER is substantial for different transmission powers.

Furthermore, the increment in transmitter power adequately expands the SNR esteem with lessening BER. This case is valid as the distance of transmission is the same, the force of the transmitter from high to low BER are 1–30 W. This is explained dependent on the connection between the BER and SNR. From Fig. 3, it is reasoned that as the force of transmission is high, SNR increases and henceforth BER lessens. Thus, the force of transmission is contrarily corresponding to BER. Since, source level force is considered as the central point because of BER, the source level force may turn out to be low when the transmitter power is high. Henceforth, it very well may be respected that the communicated power

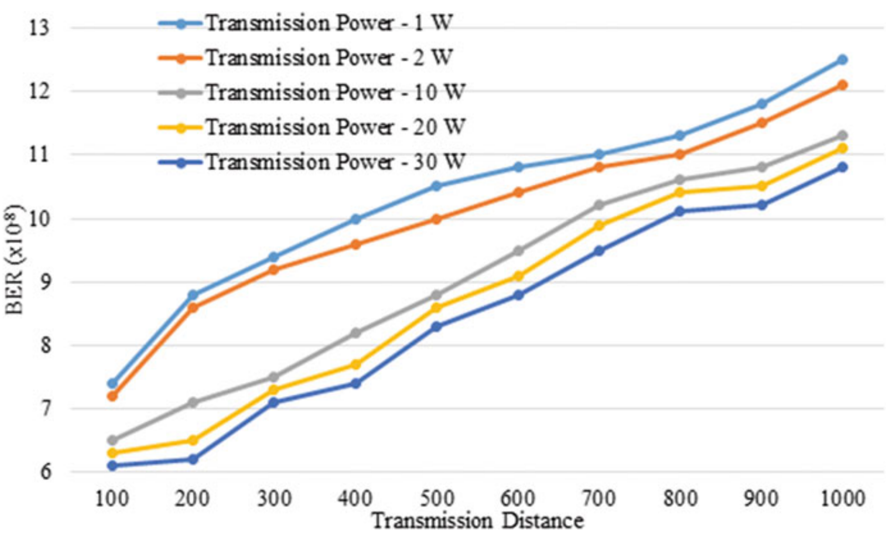


Fig. 4 Results of BER by applying channel models with different transmitter powers (×10–8) and distances

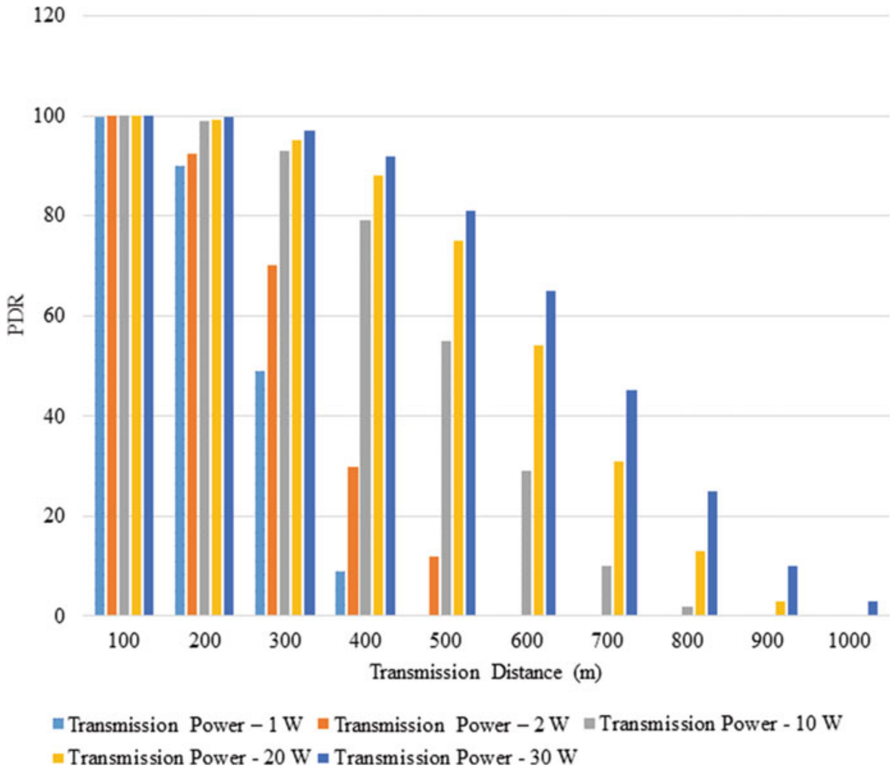


Fig. 5 Results of PDR with different transmitter powers and distances (a) 1 W (b) 2 W (c) 10 W (d) 20 W (e) 30 W

is straightforwardly related with BER, since the BER esteem is steady for all transmission distances and it at last affirms that the proposed model is material for different distances.

Further, it may very well be seen that the transmission distance of 1 W and 2 W prompts higher BER than other transmission distances. Notwithstanding, it's anything but a computational weight with more costly executions contrasted and light weight execution of other transmission distances.

5.3 *Evaluation of PDR*

The aftereffects of BER are displayed in Fig. 5, which plots the PDR as an element of distance that differs between 100 and 1000 m and the transmitter powers are set somewhere in the range of 1–30 W.

It is obvious from the figure that as the distance of transmission expands, the PDR lessens. This is because of the reality as the transmission distance expands, there is a decrease in SNR and expansion in BER. The expansion in BER is because more pieces are modified because of obstruction, commotion, and synchronization blunders and twisting. Additionally, if an entire parcel is being communicated, high BER prompts less PDR. So, as the distance of transmission expands, SNR diminishes with expansion in BER and consequently PDR lessens, which is valid for all transmitter powers. As the transmission power builds, it very well may be seen that the PDR increments with higher SNR and lower BER, which guarantees low likelihood of deceiving parcels during transmission. This prompts higher conveyance of bundles towards the objective node. It is affirmed from the outcomes that the proposed station model is appropriate for different transmission power.

6 Conclusion

This paper gives another class of IoUT model in Underwater WSN with acyclic graph directing convention. The proposed steering convention in underwater WSN demonstrates that the station model is functional and can be made appropriate for shifting communicating force and distances. As the force of transmission is high, the SNR and BER are high, resulting in effective PDR. In any case, as the distance in SNR and PDR lessens, there in an increment in BER. The outcomes affirm that the channel model planned is sensible and assists the future examination with testing the IoT model in different other underwater situations. In future, we will in general work on the steering in UWSN utilizing computerized reasoning methods. For the development of Society 5.0, we require devices that provide better results. As the working of IoT is a lot similar to WSN and so we have to work in optimizing the working of WSN. With this thought, this proposed scheme would be very beneficial in the development of IoUT, i.e., IoT underwater and it would be very helpful to develop devices for underwater and to study its different aspects.

References

1. A.G. Pereira, T.M. Lima, F. Charrua-Santos, Industry 4.0 and society 5.0: Opportunities and threats. Int. J. Recent Technol. Eng. **8**(5) (2020). ISSN: 2277-3878
2. S. Sharma, S. Verma, K. Jyoti, A new bat algorithm with distance computation capability and its applicability in routing for WSN, in *Proceedings of Soft Computing and Signal Processing*, (Springer, Singapore, 2019), pp. 163–171
3. D. Chuang, Y.Y. Chu, Y.K. Tzeng, Y. Chen, W.M. Chen, Secure routing for WSN-based tactical-level intelligent transportation systems. J. Internet Technol. **20**(4), 1013–1026 (2019)
4. N. Mahesh, S. Vijayachitra, DECSA: Hybrid dolphin echolocation and crow search optimization for cluster-based energy-aware routing in WSN. Neural Comput. & Applic. **31**(1), 47–62 (2019)

5. W. Zhang, G. Han, Y. Liu, et al., A coverage vulnerability repair algorithm based on clustering in underwater wireless sensor networks. Mob. Netw. Appl. (2020). https://doi.org/10.1007/s11036-020-01621-4
6. J. Zhu, X. Du, D. Han, et al., LEER: Layer-based energy-efficient routing protocol for underwater sensor networks. Mob. Netw. Appl. (2020). https://doi.org/10.1007/s11036-020-01693-2
7. N. Saeed, A.K. Celik, T.Y. Al-Naffouri, M.-S. Alouini, Underwater optical wireless communications, networking, and localization: A survey, 101935. Ad Hoc Netw. **94** (2019). ISSN 1570-8705, https://doi.org/10.1016/j.adhoc.2019.101935
8. K. Singh, K. Singh, L.H. Son, A. Aziz, Congestion control in wireless sensor networks by hybrid multi-objective optimization algorithm. Comput. Netw. **138**, 90–107 (2018). ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2018.03.023
9. G. Tuna, V.C. Gungor, A survey on deployment techniques, localization algorithms, and research challenges for underwater acoustic sensor networks. Int. J. Commun. Syst. **30**, e3350 (2017)
10. Z. Wan, S. Liu, W. Ni, Z. Xu, An energy-efficient multi-level adaptive clustering routing algorithm for underwater wireless sensor networks. Clust. Comput., 1–10 (2018)
11. D.-I. Curiac, Towards wireless sensor, actuator and robot networks: Conceptual framework, challenges and perspectives. J. Netw. Comput. Appl. **63**, 14–23 (2016). ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2016.01.013
12. A. Sisodia, S. Kundu, Enrichment of performance of operation based routing protocols of WSN using data compression, in *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, (IEEE, 2019), pp. 193–199
13. A. Sisodia, H.H. Swati, Incorporation of non-fictional applications in wireless sensor networks. Int. J. Innov. Technol. Explor. Eng. **9**(11), 42–49 (2020). ISSN: 2278–3075 (Online), https://doi.org/10.35940/IJITEE.K7673.0991120
14. N. Javaid, T. Hafeez, Z. Wadud, N. Alrajeh, M.S. Alabed, N. Guizani, Establishing a cooperation-based and void node avoiding energy-efficient underwater WSNs for a cloud. IEEE Access **5**(1), 11582–11593 (2017)
15. Z. Khalfallah, I. Fajjari, N. Aitsaadi, P. Rubin, G. Pujolle, A novel 3D underwater WSN deployment strategy for full-coverage and connectivity in rivers, in *Proceedings of IEEE ICC*, (2016), pp. 1–7
16. Z. Khalfallah, I. Fajjariz, N. Aitsaadi, R. Langar, G. Pujolle, 2D-UBDA: A novel 2-dimensional underwater WSN barrier deployment algorithm, in *Proceedings of IEEE IFIP Networking Conference (IFIP Networking)*, (2015), pp. 1–8
17. Y.S. Yan, H.Y. Wang, X.H. Shen, F.Z. Yang, Z. Chen, Efficient convex optimization method for underwater passive source localization based on RSS with WSN, in *Proceedings of IEEE Signal Processing, Communication and Computing (ICSPCC 2012)*, (2012), pp. 171–174
18. N. Javaid, S. Cheema, M. Akbar, N.A. Alrajeh, M.S. Alabed, N. Guizani, Balanced energy consumption based adaptive routing for IoT enabling underwater WSNs. IEEE Access **5**(1), 10040–10051 (2017)
19. Y. Dong, R. Wang, Z. Li, C. Cheng, K. Zhang, Improved reverse localization schemes for underwater wireless sensor networks, in *Proceedings of ACM/IEEE International Conference on Information Processing in Sensor Networks, IPSN '17*, 2017, pp. 323–324
20. A. Khan, N. Javaid, H. Mahmood, Z.A. Khan, U. Qasim, EEIRA: An energy efficient interference and route aware protocol for underwater WSNs, in *Proceedings of IEEE International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, (2016), pp. 264–270
21. S. Mahmood, H. Nasir, S. Tariq, H. Ashraf, M. Pervaiz, Z.A. Khan, N. Javaid, Forwarding nodes constraint based DBR (CDBR) and EEDBR (CEEDBR) in underwater WSN. Procedia Comput. Sci. **34**, 228–235 (2014)
22. A.H. Khan, M.R. Jafri, N. Javaid, Z.A. Khan, U. Qasim, M. Imran, DSM: Dynamic sink mobility equipped DBR for underwater WSNs. Procedia Comput. Sci. **52**, 560–567 (2015)

23. N. Ilyas, M. Akbar, R. Ullah, M. Khalid, A. Arif, A. Hafeez, N. Javaid, SEDG: Scalable and efficient data gathering routing protocol for underwater WSNs. Procedia Comput. Sci. **52**, 584–591 (2015)
24. S. Ahmed, M. Akbar, R. Ullah, S. Ahmed, M. Raza, Z.A. Khan, N. Javaid, ARCUN: Analytical approach towards reliability with cooperation for underwater WSNs. Procedia Comput. Sci. **52**, 576–583 (2015)
25. Y. Han, J. Zhang, D. Sun, Error control and adjustment method for underwater wireless sensor network localization. Appl. Acoust. **130**, 293–299 (2018)
26. S. Rani, S.H. Ahmed, J. Malhotra, R. Talwar, Energy efficient chain based routing protocol for underwater wireless sensor networks. J. Netw. Comput. Appl. **92**, 42–50 (2017)
27. L. Jin, D.D. Huang, A slotted CSMA based reinforcement learning approach for extending the lifetime of underwater acoustic wireless sensor networks. Comput. Commun. **36**(9), 1094–1099 (2013)
28. A.P. Das, S.M. Thampi, Fault-resilient localization for underwater sensor networks. Ad Hoc Netw. **55**, 132–142 (2017)
29. J. Poncela, M.C. Aguayo, P. Otero, Wireless underwater communications. Wirel. Pers. Commun. **64**, 547–560 (2012). [Google Scholar] [CrossRef]
30. B.B. Das, S.K. Ram Localization using beacon in wireless sensor networks to detect faulty nodes and accuracy improvement through DV-Hop algorithm, in *Proceedings of the Inventive Computation Technologies (ICICT)*, International Conference, Coimbatore, India, 26–27 August 2016, vol. 1 pp. 1–5. [Google Scholar]
31. J. Yu, Q. Zhang, D. Yu, C. Chen, G. Wang, Domatic partition in homogeneous wireless sensor networks. J. Netw. Comput. Appl. **37**, 186–193 (2014). [Google Scholar] [CrossRef]
32. S. Ullah, J.J. Rodrigues, F.A. Khan, C. Verikoukis, Z. Zhu, Protocols and architectures for next-generation wireless sensor networks. Int. J. Distrib. Sens. Netw. **2014**, 705470 (2014). [Google Scholar] [CrossRef]
33. A.F. Salami, E.A. Adedokun, F. Al-Turjman, H. Bello-Salau, B.O. Sadiq, E.M. Dogo, Explorative analysis of AUV-aided cluster-based routing protocols for internet of intelligent underwater sensors, in *Drones in Smart-Cities 2020 Jan 1*, (Elsevier, Amsterdam, 2020), pp. 143–187. [Google Scholar]
34. T. Al-Subhi, B. Arafeh, N. Alzeidi, K. Day, A. Touzene, A void avoidance scheme for grid-based multipath routing in underwater wireless sensor networks. Wirel. Sens. Netw. **10**, 131–156 (2018). [Google Scholar] [CrossRef]
35. A. Khan, K. Aurangzeb, E.U. Qazi, A. Ur Rahman, Energy-aware scalable reliable and void-hole mitigation routing for sparsely deployed underwater acoustic networks. Appl. Sci. **10**, 177 (2020). [Google Scholar] [CrossRef]
36. S. John, V.G. Menon, A. Nayyar, Simulation-based performance analysis of location-based opportunistic routing protocols in underwater sensor networks having communication voids, in *Data Management, Analytics and Innovation*, (Springer, Singapore, 2020), pp. 697–711. [Google Scholar]
37. M. Awais, J. Nadeem, N. Nidal, I. Muhammad, Exploiting energy efficient routing protocols for void hole alleviation in IoT enabled underwater WSN, in *Proceedings of the 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Tangier, Morocco, 24–28 June 2019, pp. 1797–1802. [Google Scholar]
38. M. Awais, A.k. Zahoor, T. Nadeem, M. Abdul, R. Aymen, H. Farooq, Cluster-based routing protocols with adaptive transmission range adjustment in UWSNs, in *Proceedings of the International Conference on Emerging Internetworking, Data & Web Technologies*, Fujairah, UAE, 26–28 February 2019, (Springer, Cham, 2019), pp. 528–539. [Google Scholar]
39. M. Awais, M. Abdul, A.K. Zahoor, J. Nadeem, A survey on multi-hop routing protocols for efficient resource allocation in IoTs, in *Proceedings of the 2018 Fifth HCT Information Technology Trends (ITT)*, Dubai, UAE, 28–29 November 2018, pp. 239–244. [Google Scholar]
40. M. Li, X. Du, X. Liu, C. Li, Shortest path routing protocol based on the vertical angle for underwater acoustic networks. J. Sens. **2019**, 9145675 (2019). [Google Scholar] [CrossRef]

A Novel Algorithm for Reconfigurable Architecture for Software-Defined Radio Receiver on Baseband Processor for Demodulation

H. D. Nataraj Urs, R. Venkata Siva Reddy, Raveendra Gudodagi, K. M. Sudharshan, and B. N. Aravind

1 Introduction

The wireless communication industry is facing new challenges due to constant evolution of new standards (2.5G, 3G, 4G and 5G), existence of incompatible wireless network technologies in different countries inhibiting deployment of global roaming facilities and problems in rolling out new services/features due to widespread presence of legacy subscriber handsets. Software-defined radio (SDR) technology promises to solve these problems by implementing the radio functionality on a generic hardware platform. Further, multiple modules, implementing different standards, can be present in the radio system, and the system can take up different personalities depending on the module being used [2] (Fig. 1).

2 Background

2.1 Baseband Demodulation

In the current implementation, algorithms for demodulation are implemented on GPP (general purpose processor) hardware [3], and the analog front end of SDR is already made to be flexible and reconfigurable [4]. This work focuses

H. D. Nataraj Urs (✉) · R. Venkata Siva Reddy · R. Gudodagi · K. M. Sudharshan
School of Electronics & Communication Engineering, REVA University, Bengaluru, India
e-mail: natarajurs.hd@reva.edu.in; venkatasivareddy@reva.edu.in; raveendra.g@reva.edu.in; sudharshankm@reva.edu.in

B. N. Aravind
Department of Electronics and Communication Engineering, Rajeev Institute of Technology, Hassan, India

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_11

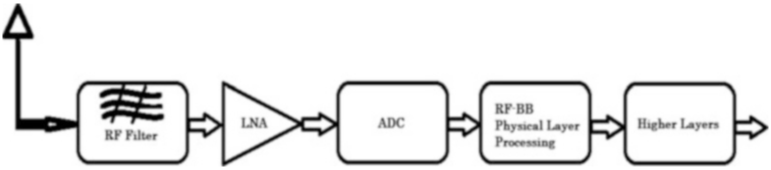


Fig. 1 SDR architecture

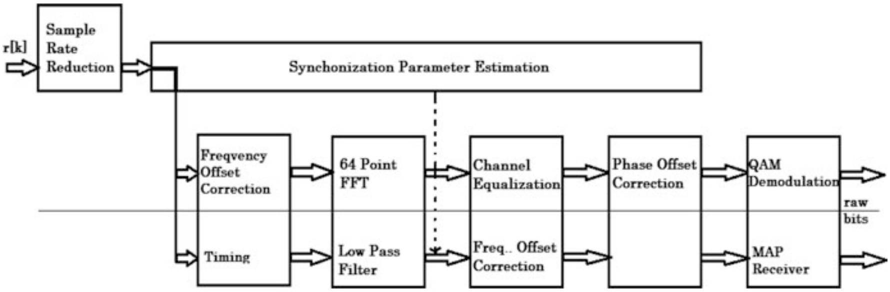


Fig. 2 Functional architecture of the receiver

on the hardware implementation of digital baseband part of the receiver (PHY (physical layer) only). Input data is coming in the BB receiver after the analog front end (including ADC (analog-to-digital converter)) at the rate of 80 MSPS (megasamples-per-second). The digital baseband part consists of a sample rate reduction block followed by digital demodulator block. The output from sample rate reduction block is fed to the digital demodulator part which demodulates the data stream digitally.

2.2 *OFDM*

After frequency offset correction, the first step is the inverse OFDM as shown in Fig. 2. The inverse OFDM is same as fast Fourier transform (FFT) operation. An OFDM symbol has a duration of 80 complex samples. Only 64 samples of them are needed for the FFT operation. The remaining 16 samples are used as cyclic prefix to reduce inter-symbol interference (ISI) and synchronization. So the first step in the receiver is to pass the data through 64-point FFT block. After examining various FFT algorithms [1, 5], we chose to use radix-2 FFT in our implementation. Radix-2 FFT is performed using radix-2 butterflies and requires $64 \times \log_2(64)$ complex multiplications.

2.3 Channel Equalization

After FFT, the channel equalizer block has to compensate the channel for the carriers. The estimation of the channel is done by comparing the known preamble and the received subcarrier values. This equalization should be done for 52 subcarriers. So it will require 52 complex multiplications per OFDM symbol [6].

2.4 Phase Offset Correction

At the front end of the receiver, frequency offset correction is implemented by calculating only the values of the frequency offset for the first symbol, and these values are subsequently reused for other symbols. This saves (computational-intensive) instructions (cos and sin) but also introduces a phase offset. This phase offset can be corrected by using the pilot carriers in the OFDM symbol. This requires 48 complex multiplications.

2.5 QAM (Quadrature Amplitude Modulation) Demapping

The final step in demodulation is demapping. There are four constellations available: BPSK (binary phase-shift keying), QPSK (quadrature phase-shift keying), 16-QAM and 64-QAM. Each of these constellations has a different number of bits per complex symbol. Demapping can be done using lookup table. In the lookup table, all possible subcarrier values for a certain mapping scheme are defined [7]. For BPSK, two subcarrier values are stored in the lookup table; for QPSK, 16-QAM and 64-QAM, there are 4, 16 and 64 subcarrier values stored, respectively. The largest constellation used is 64-QAM (Table 1).

A 64-QAM symbol has $2^3 = 8$ possible values for both the real and imaginary parts. Demapping can be implemented by generating an index for a table. So demapping requires two comparisons (border checking), one addition, one multiplication and one lookup table.

Table 1 Computational requirements for receiver

Function	Data rate	Number of multiplications	Number of additions
64-point FFT	16	153.6e6	76.8e6
Channel equalization	13	20.8e6	10.4e5
Phase shift correction	12	19.2e6	10.4e6
64-QAM demapping	12	9.6e6	9.6e6

3 Algorithm Analysis

The algorithm domain of the SDR includes baseband demodulation algorithms. In this work, we are dealing with the hardware implementation of the channel selection block of receiver and OFDM block. (The halfband filter block and matched filter block are combined together into one channel selection block in the receiver.) For this purpose, our first step is to perform the dataflow analysis in various computations of these algorithms.

3.1 Dataflow for Channel Selection/FFT

The first block in the baseband demodulation receiver is a 64-point FFT block. This block is used for OFDM demodulation. The data from the sample rate reduction block is coming at 20 MSPS. This data is arranged in blocks of 80 samples each. Due to OFDM scheme, last 16 samples are same as the first 16 samples in each block. So we need to take 64 samples out of these 80 samples.

The first block in the baseband demodulation receiver is a channel selector/low-pass filter (LPF). This is required to select the desired 1 MHz bandwidth (BW) channel. As we analysed, the complexity and data computation unit of FFT block are similar to LPF section [8]. So in our implementation, we propose to combine FFT with LPF. But direct implementation of LPF is computationally intensive. The input data is first passed through two linear phase halfband filters. Each halfband filter decimates data by factor 2. These halfband filters help in reducing the order of matched filter. Also matched filter can be designed to be linear phase. In this way, the number of computations can be reduced further. A simple schematic for channel selector section is shown in Fig. 3.

3.2 Signal Flow Graph for FIR/FFT

The signal flow graphs and basic building blocks corresponding to halfband filter, matched filter and FFT (butterfly) are described below.

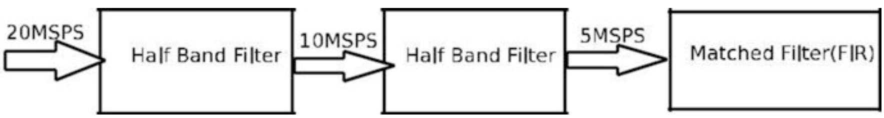


Fig. 3 Channel selector section of Bluetooth

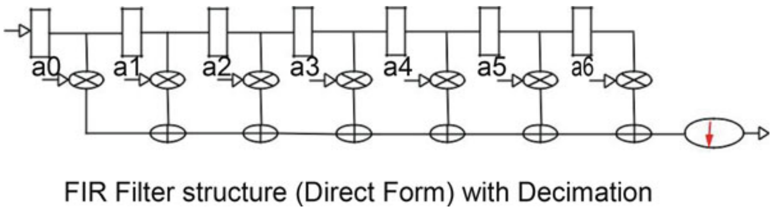


Fig. 4 Direct form FIR filter

Halfband Filter

Input data stream is filtered through halfband filters before doing low-pass filtering. There are two halfband filters. Each halfband filter is of seventh order. To simplify the computations, main points to remember about this building block are linear phase, halfband and decimation. By using linear phase property, we can reduce the number of multiplications by a factor 2. Halfband property means that the number of multiplications (corresponding to the amount of zeros in filter coefficient) can be reduced further. Also, using a polyphase representation, decimation can be used to reduce the speed of computation. A basic seventh-order FIR filter can be represented as in equation:

$$H(z) = a_0 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} + a_4 z^{-4} + a_5 z^{-5} + a_6 z^{-6} \tag{1}$$

Its critical path contains one multiplier and six adders. A direct form implementation of such filter is shown in Fig. 4.

The transposed form of above filter is shown in Fig. 4. Its critical path contains one multiplier and one adder only.

The halfband property of the filter implies that a_1 and a_5 have zero value and can be omitted to reduce the number of multiplications required. Also, the linear phase property implies that $a_2 = a_4$ and $a_0 = a_6$. So the multiplications in first half of the filter are identical to the multiplications in other half. Thus, Eq. 1 can be rewritten as follows:

$$H_{(z)} = a_0 + a_2\, z^{-2} + a_3\, z^{-3} + a_2\, z^{-4} + a_0\, z^{-6} \tag{2}$$

By using polyphase representation, decimation by 2 can be used to reduce the speed of computations (if needed). Thus, Eq. 2 can be written in polyphase form as follows:

$$H_{(z)} = \left(a_0 + a_2\, z^{-2} + a_2\, z^{-4} + a_0\, z^{-6}\right) + Z^{-1}\left(a_3\, z^{-2}\right) \tag{3}$$

The simplified structure, which is computationally most efficient in terms of speed of operation and in terms of the amount of data path computations, is shown in Fig. 5.

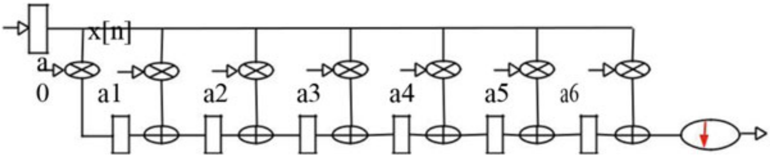


Fig. 5 Transposed form FIR filter

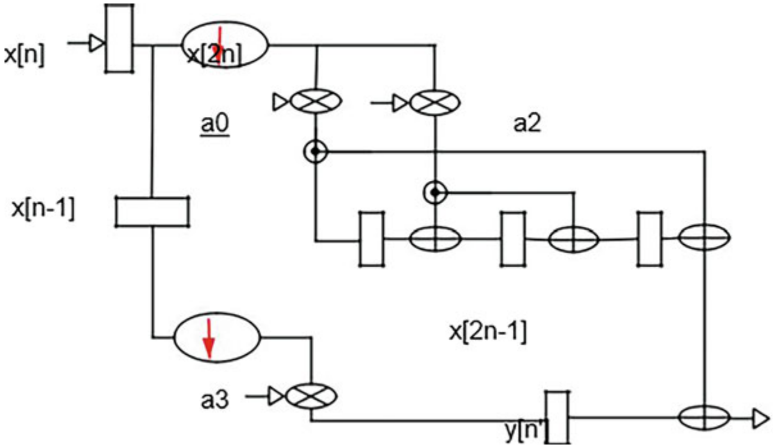


Fig. 6 Filter structure simplification

In this way, the number of multiplications can be reduced by a factor of 3/7 from direct form halfband filter. Also, each computation unit can work at half of the incoming data rates.

Moreover, it is important to notice that the filter structure above has a basic computation unit (shown in Fig. 6). The repetitive use of this unit realizes the filter. The basic operation can be described as multiply and add (Fig. 7).

FIR (Matched Filter)

After halfband filtering, the input data (decimated by 4) is fed to matched filter block. The output of this block is the data corresponding to desired channel. The matched filter used here is of 17th order. The transposed form representation is shown in Fig. 8. The basic computation unit is the same as the one for halfband filters. Polyphase decomposition for efficient decimation and halfband properties are not applicable for this stage. So filter structure is corresponding to transposed form structure with linear phase. This means that the number of multiplications can be reduced by 2.

FFT

An OFDM demodulator consists of a FFT block. An FFT represents set of algorithms to compute discrete Fourier transform (DFT) of a signal efficiently. An

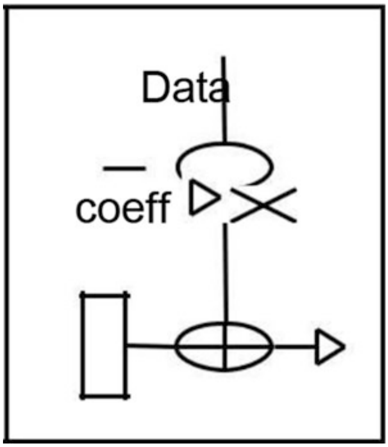


Fig. 7 Filter calculation unit

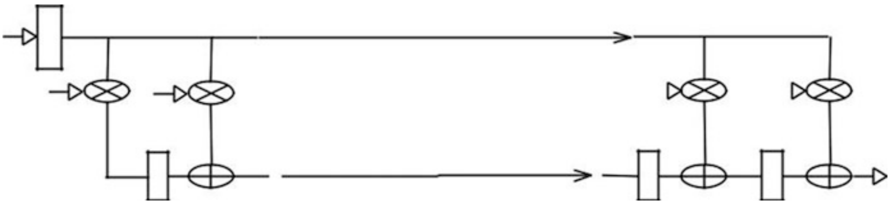


Fig. 8 Transposed form LPF for matched filtering

N-point DFT corresponds to the computation of N samples of the Fourier transform at N equally spaced frequencies, $\omega_k = 2\pi k/N$, that is, at N-points on the unit circle in the z-plane. The DFT of a finite-length sequence of length N is

$$X[k] = \sum_{n=0}^{N-1} x[n]\, {W_N}^{kn} \cdots \forall \ k \in \{0, 1, \ldots N - 1\} \tag{4}$$

where ${W_N}^{kn} = \mathrm{e}^{-j2\pi/N}$. The idea behind almost all FFT algorithms is based upon divide and conquer strategy and establishes the solution of a problem by working with a group of subproblems of the same type and smaller size.

An objective choice for the best DFT algorithm cannot be made without knowing the constraints imposed by the environment in which it has to operate. The main criteria for choosing the most suitable algorithm are the amount of required arithmetic operations (costs) and regularity of structure. Several other criteria (e.g. latency, throughput, scalability, control) also play a major role in choosing a particular FFT algorithm. We have chosen radix-2 DIF FFT implementation for our system because it has advantages in terms of regularity of hardware, ease of computation and number of processing elements. Also, the basic butterfly corresponding to radix-2 can be combined easily with filter processing element (of our implementation). This facilitates the similar data path computations in two receivers and simple control structure for receiver.

4 Architecture Design Approach

In this work, we propose a solution which is optimized for our specific algorithmic domain. Our algorithm domain is limited to the DSP (digital signal processing) algorithms for each stage of SDR receiver. In the proposed architecture, the basic approach is to limit the flexibility of design to the algorithms of interest (OFDM and channel selection) [9]. This limited flexibility requirement will result in only moderate degradation of the ASIC performance. This is in contrast to various designs discussed in previous chapter, where the approach is to incorporate the sufficient flexibility to support the application domain. So our approach is to design a flexible ASIC-like system for specific algorithms only. Our design approach has four main steps as follows:

(a) In the first step, we are identifying the dominant kernels of our algorithm domain. This step is similar to any domain-specific design mentioned previously and requires careful reviewing of the tailored application's area requirements.
(b) In the second step, we have designed the optimal control hardware for our algorithm domain. This is in contrary to various regular available hardware design approaches that put their attention towards the dominant data processing operations only.
(c) In the third step, we have identified the communication patterns in our algorithm domain as recommended. This has helped us in designing the optimal communication network in the system. Only those parts of communication are programmable which are really needed. As far as possible, global buses are minimized to reduce capacitance and crosstalk effects. So point-to-point and local communication is preferred in our proposed architecture.
(d) In the fourth step, we have identified the memory requirements for our systems. In this step, we have identified things like how much RAM (random-access memory) and ROM (read-only memory) are needed, what are the memory bandwidth requirements, is it better to reuse the memory by using in-place computations, etc.

The proposed architecture comprises of nine homogenous data processing tiles, two 128 $\times$ 16-bit memory (RAM) tiles, one 64 $\times$ 16-bit ROM, a configuration unit to configure the data and communication network and a control section in the form of a state machine to execute algorithm steps sequentially. The control section also controls the data transfer from data path elements to memories through the communication network. It also generates the control signals for the configuration unit. The architecture view of the system is shown in Fig. 9.

The proposed design is based on tiled architecture. A tiled architecture in which various tiles are connected by an on-chip network has a very modular design [10]. The design of a single processing tile is relatively simple and allows extra effort for power optimizations at physical level. To increase or decrease processing power of

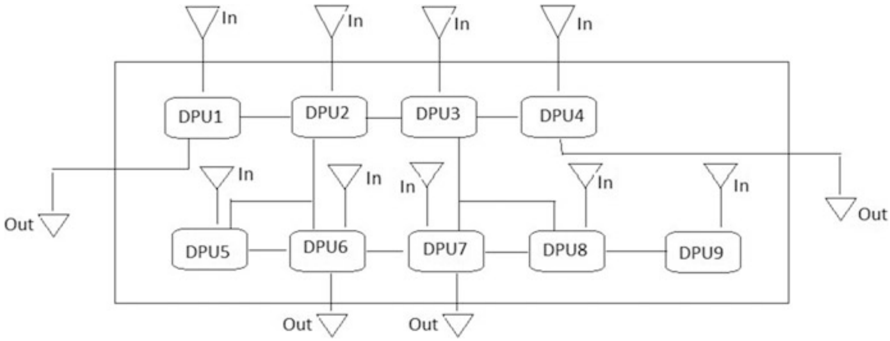


Fig. 9 Tiled architecture

our system, we can easily add or remove tiles. A simplified view of our tiled network is shown in Fig. 9.

4.1 Reconfigurability

The proposed design is reconfigurable within one clock cycle and supports the chosen subset of the SDR algorithms. So the algorithm domain of our design includes FIR filter, halfband filters and radix-2 FFT. These algorithms are also the most common algorithms used to benchmark a DSP system [9]. The dynamic reconfigurability allows time-sharing of hardware resources by pipelining the algorithms. This minimizes the total hardware resources required to implement the complete system. Also, almost all of the WLAN (wireless local area network) systems use either phase modulation or OFDM-based modulation [11]. So the suitability of our system for phase-modulated and OFDM-based receivers implies that our design can be used in number of WLAN systems.

4.2 Data Path

In the proposed design, data path consists of nine homogeneous 16-bit data processing tiles called data processing units (DPUs). The detailed view of our data path is shown in Fig. 10. A single DPU is depicted in Fig. 10. The design of a DPU can be divided into four parts: the processing part, the storage part, the configuration part and the communication interface. These parts are shown as arithmetic unit, registers, configuration part and various input/output ports, respectively, in Fig. 10.

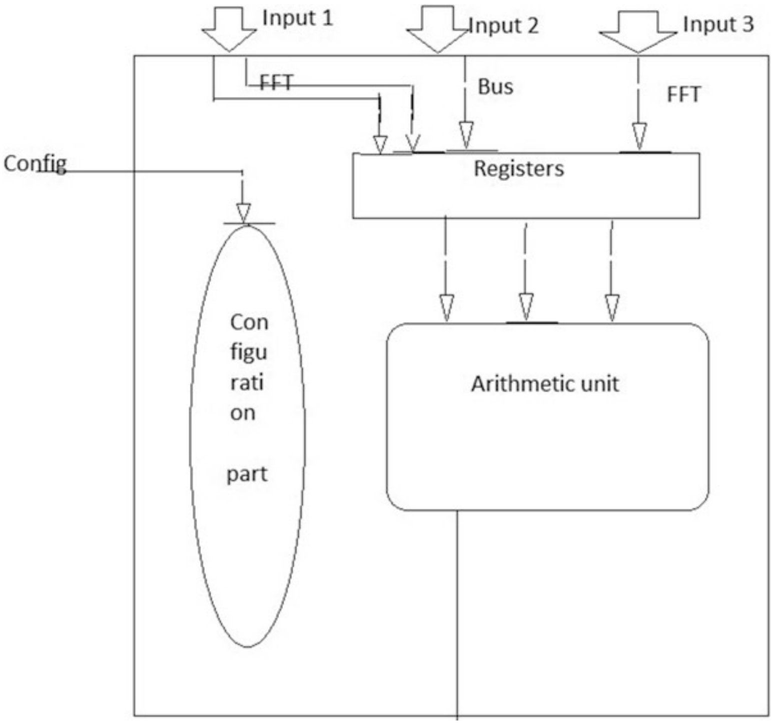


Fig. 10 A data processing unit (DPU)

4.3 The Communication Interface

The communication interface of each DPU supports the use of heterogenous processing occupying one or more tiles. This interface manages the communication through each tile and synchronizes the global communication. Each DPU has three sets of 16-bit inputs:

- Input 1 set is used to read data either from left or from right neighbour into the registers. The ports corresponding to these inputs are named as 'LHS' and 'RHS'.
- Input 2 set (bus 2) is used to read data from global bus of the system. There are two global buses in our system. Each global bus is providing the input to one row of DPUs.
- Input 3 set is connected to two point-to-point buses of the system. The ports corresponding to these inputs are named as 'FFT bus' and 'global bus'.

Each DPU has the following two 16-bit outputs:

- First output ('sideout') is used to communicate with the adjacent left and right side neighbours.

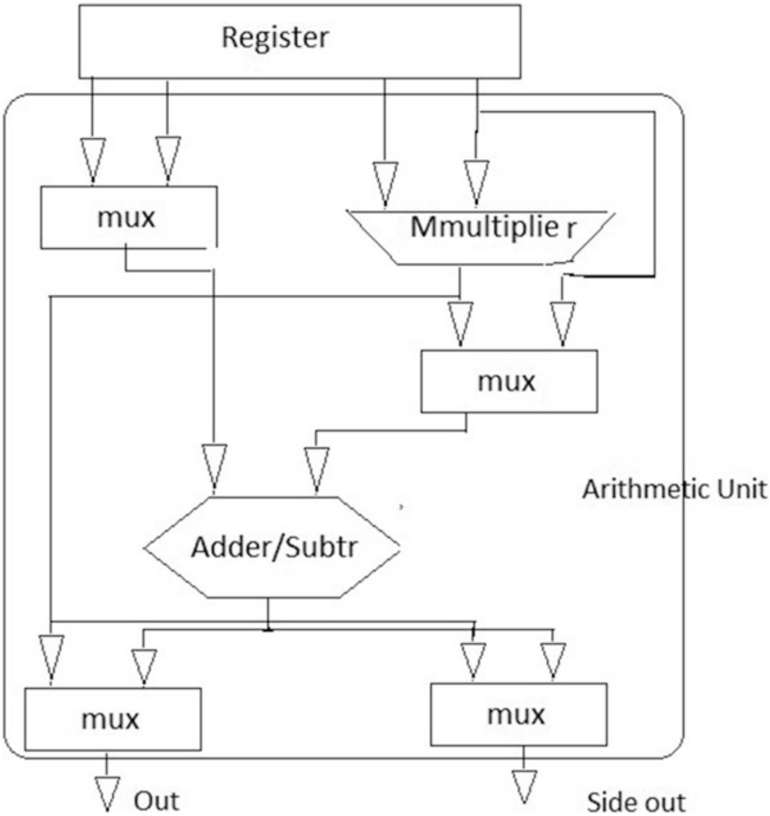


Fig. 11 Arithmetic unit (AU) of DPU

- Second output ('out') is used to communicate data over the system communication buses. To avoid bus arbitration output, 'out' is a tri-state output.

4.4 *The Processing Part*

The data processing capabilities of DPU are attributed to a 16-bit arithmetic unit (AU). A functional representation of the AU is shown in Fig. 11. An AU is purely combinational and is capable of doing the basic 16-bit arithmetic operations, namely, add, subtract, multiply, multiply and add, and multiply and subtract. The input to AU is from internal registers, and outputs are provided on the output ports.

4.5 The Storage Part

Each DPU comprises a set of 11 local data registers of 16 bits each. These registers can be used to store intermediate data variables as required in FIR data structure. This way of having local registers is far more efficient than one centralized set of registers [15]. These registers are used to read data from input ports and to provide data to ALU. In this way, inputs are always registered, thus minimizing the excessive glitches. Another reason for having registered inputs is to allow pipelining between various data path units. This not only allows the reduction of critical path delay but also allows a straightforward implementation of transposed form FIRs.

4.6 The Configuration Part

Each DPU has a local configuration section called 'configuration part', which provides the configuration signals to various entities within the DPU. This configuration section is part of the control hierarchy of the system to reduce the control overhead significantly [12]. The input to this section comes from the main configuration unit of the architecture.

4.7 Control Section

In the proposed architecture, the control section is implemented as a state machine corresponding to each algorithm. This is motivated by the fact that dataflow is determined at the design time itself. In the normal operation, the control system loops through the set of algorithms steps called a schedule. To compute an algorithm, first the control section is activated with the corresponding wake-up call. The control section responds by generating the series of control signals to memory and to the configuration part, thus controlling the data operations in the system. In this way, we avoid the common bottleneck (correspondingly to fetch and decode an instruction before execution) found in normal processor-like architecture. This scheme has obvious disadvantage that each new algorithm needs to be implemented separately. So if algorithm is subject to change, one should incorporate the programming facility in the control.

4.8 Configuration Unit

In the proposed architecture, reconfigurability is achieved by reconfiguration of the data path and reconfiguration of the communication network.

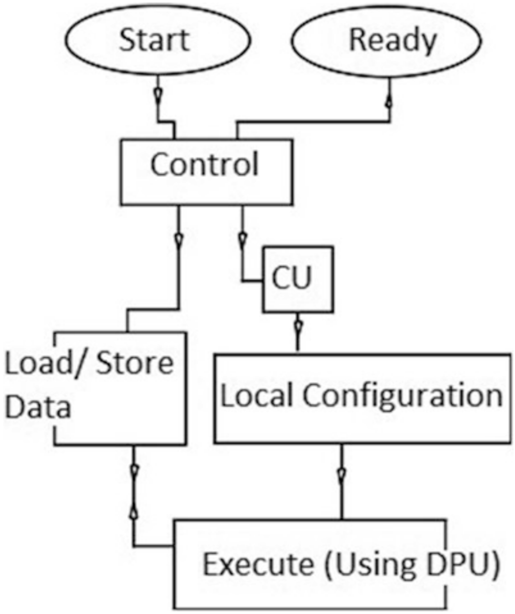


Fig. 12 Configuration unit block diagram

These configuration signals are generated in the configuration unit (CU). The input of the CU comes from control section in the form of control signals. The CU decodes these control signals and provides input to local configuration sections of various DPUs. The configuration of the data path and communication network is achieved within one clock cycle. This allows dynamic and static reconfigurations in the proposed architecture. To compute an algorithm, the first step is to activate the centralized control section. This control section then activates the CU on a per-clock-cycle basis. The CU provides the input to local configuration of each DPU. Each local configuration part responds by configuring the corresponding subsection of data path. This way, distributed control is achieved in the proposed architecture.

This is shown in Fig. 12. This facilitates high operating speeds and time-sharing of data and communication network. The low-overhead and dynamic reconfiguration allows time multiplexing of the processing part.

5 Algorithm Mapping

5.1 Mapping of Matched FIR Filter

The input data after halfband filtering and decimation is processed into 17th-order matched FIR filter. This means that we need 17 basic computations equivalent to a MAC operation (shown in Fig. 13). For each sample, our implementation can range

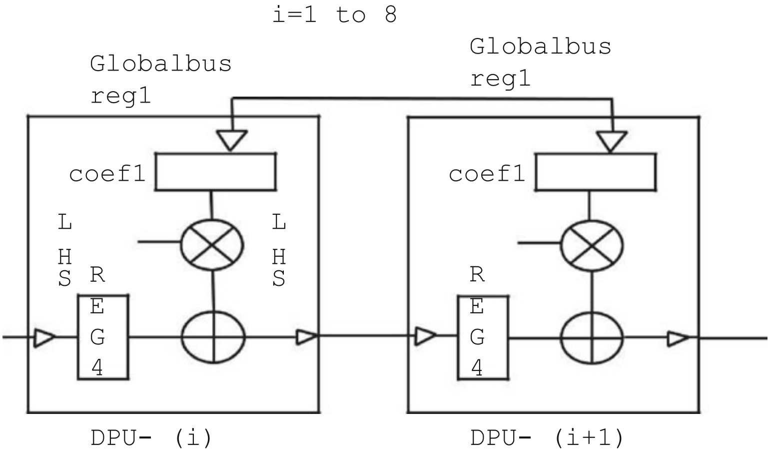


Fig. 13 First clock cycle in FIR mapping

from using 1 DPU, that is, 17 clock cycles for 1 computation, to 17 DPUs, that is, 1 clock cycle computation. We propose to use an intermediate solution which uses two clock cycles for one computation of real or imaginary data. Data processing of real and imaginary parts is done in alternate cycles. This means that there will be four clock cycles of computation for each data input. For this solution, we need nine DPUs. This decision is the main determining factor for choosing nine DPUs in the proposed architecture. Scheduling corresponding to real part is discussed in next few lines. Imaginary part will be calculated in the same way:

- Load data sample from memory into the global bus connecting DPU1–DPU4 and into global bus connecting DPU5–DPU9.
- Each AU is configured for multiply and add.
- Read data from global bus input into a data register.
- Read intermediate data value from LHS input into a data register.
- Configure multiplier inputs of the AU: input 1 is from stored data input corresponding to global bus input and input 2 is from 'coef1' value stored in another register within the DPU.
- Configure adder inputs of the AU: input 1 is from multiplier output and input 2 is from intermediate value corresponding to LHS input stored in a data register.
- Put adder output into 'sideout' output (data is flowing from left to right).
- Tri-state the main output of each DPU.

Dataflow in this clock period is shown in Fig. 13.

Similar to the halfband filtering step, in the second clock cycle, only the following steps are different:

- Read intermediate data value from RHS input into a data register. In all operations in the first clock cycle, LHS is replaced by RHS.

- In DPU1, put adder output onto the main output. This is the filtered output from FIR filter. Store this output into memory for the next stage.

Dataflow in this clock period is shown in Fig. 13. This implementation allows us to use linear phase property, and hence, a number of multipliers in hardware are reduced by half. Also, the speed of multiplication and addition in the AU is corresponding to the critical path delay of the system.

5.2 Mapping of FFT

The heart of the FFT is the butterfly computation. As already discussed, we use radix-2 butterfly for regularity and ease of computation. This means that we will have 32 butterflies and 6 stages of computation. The basic butterfly was shown in Fig. 14. From the figure, it is clear that the real and the imaginary parts of a butterfly have a similar structure. For hardware mapping, we need two ROMs for storing real and imaginary parts of twiddle factors ($= e^{-j2\pi k/N}$). There are two memory (RAM) units required for storing real and imaginary parts of data of one stage. In the next few lines, we will discuss the mapping corresponding to real part of butterfly. This mapping needs four DPUs each for real and imaginary part of butterfly [12]. So we will need to use DPU1–DPU8. This means that throughput of our design will be one butterfly per clock cycle. Therefore, we will need 32 clocks to compute one stage of FFT. In total, we will need $32 \times 6 = 192$ clocks of computations. Configuration of each DPU is described below and is also shown in Fig. 14:

- Configure DPU1 for addition; read data from FFTbus input and bus2 input; put the AU output into the A_{re} memory.
- Configure DPU2 for subtraction; read data from FFTbus input and bus2 input; and put the AU output onto the FFTbus input of DPU5 and DPU7.
- Configure DPU3 for addition; read data from FFTbus input and bus2 input. Put the AU output into the A_{im} memory.
- Configure DPU4 for subtraction; read data from FFTbus input and bus2 input; and put the AU output onto the FFTbus input of DPU6 and DPU8.
- Configure DPU5 for multiplication; read data from FFTbus input and bus2 input; and put the AU output onto the sideout.
- Configure DPU6 for multiply and subtract; read data from FFTbus input and bus2 input into multiplier; and put the multiplier output and LHS input into the subtractor. Put the AU output into the B_{re} memory.
- Configure DPU7 for multiplication; read data from FFTbus input and bus2 input; put the AU output onto the sideout.
- Configure DPU8 for multiply and add; read data from FFTbus input and bus2 input into multiplier; and put the multiplier output and LHS input into the adder. Put the AU output into the B_{im} memory.
- Configure DPU9 for sleep mode.

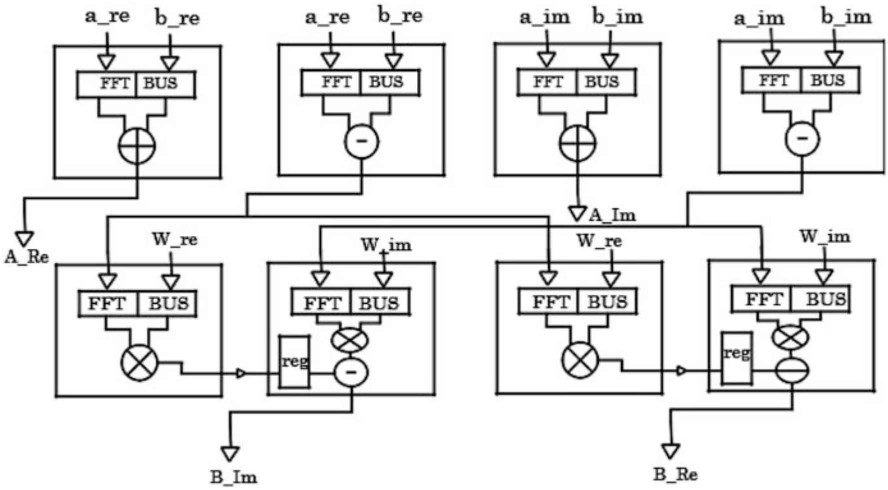


Fig. 14 One butterfly mapping

This implementation is slightly different from basic butterfly computation. This is because we are registering the data output of DPU2. This will cause one clock latency.

6 Synthesis and Evaluation

This section elaborates the synthesis results and evaluates the design after hardware realization. The control section discusses the minimum speed requirements that the design must fulfil to meet the SDR receiver requirements [13]. The synthesis results for the proposed design are presented, and it summarizes the performance of Montium TP, when the chosen SDR algorithms are mapped onto it. The performance of the proposed system is compared with the performance of the Montium TP system for the chosen SDR algorithms.

6.1 Synthesis Results for the SDR Receiver

The results of synthesis are shown in Table 2. These results indicate that the proposed system approximately requires 0.6 mm^2 of silicon area and has a critical path length of 5.3 ns. Thus, the maximum operating frequency of the system is 188 MHz, which is well above the minimum operating frequency estimated in the previous section. This gives us enough room to play with the latency requirements of the overall system.

Table 2 Synthesis results for SDR receiver

Component	Area [μm^2]	Critical path [ns]
DPU (x9)	510,000	5.3
Control	26,000	3.8
CU	1300	–
Wiring	62,700	–
Resultant	600,000	5.3

The results of synthesis are used as an indicator to evaluate the performance of our system. It is important to note that we have not included the area required due to various memories (RAM, ROM, buffer) in the system [7, 14]. In the proposed design, we need two RAMs of 128 × 16 size each and one ROM of 64 × 16 size. From the above results, it is clear that the majority of area is consumed by the data path of the system. The control part consumes less than 5% of the total area.

6.2 *Comparison of Proposed Design with Montium TP*

It is clear from the previous section that it will be very difficult for a single Montium TP to satisfy the real-time requirement of the parts of HiperLAN2 receiver we chose to implement. In the case of Bluetooth receiver, even if we use the Montium TP with maximum operating frequency, still we will need two TPs to realize the various filter stages [15]. It is very difficult to exploit the linear phase property of the filters because FIR matched filter requires four clock cycles. Also, the more general bus network in Montium TP implies more energy wastage in charging and discharging of redundant capacitances. The configuration time of a Montium TP varies depending on the algorithm, for example, a 64-point FFT needs 473 clock cycles and an FIR filter of 20th order needs 270 clock cycles.

The Montium TP occupies 2 mm^2 area in CMOS12 process from Philips. The maximum clock frequency for Montium TP is according to the synthesis tool, about 40 MHz. It is estimated that the Montium TP ASIC realization can implement an FIR filter at about 140 MHz and an FFT at about 100 MHz. The CMOS12 process has a gate density of 200 kgate/mm^2. So if we normalize our synthesis results to this process, our implementation will need 0.24 mm^2 area (approximately eight times smaller than one Montium TP). But it is important to notice that in the Montium TP, approximately 0.5 mm^2 area is occupied by RAM memory. In our system, we need a RAM of 256 × 16 size and a ROM of 64 × 16 size, which will occupy an additional area of approximately 30,000 μm^2 in our system.

On the other hand, the Montium TP has much more flexibility and is suitable to implement a number of DSP algorithms [10]. In the design space, our system is closer to the ASIC implementation than the Montium TP (which is a domain-specific reconfigurable accelerator for the chameleon SoC).

Table 3 Comparison of different architectures for butterfly computation

Architecture Type	Architecture name	Area [mm^2]	Speed [MHz]
ASIP	FASRA (ASIC)	0.63	120
DSRA	Avispa	6.5	150
DSRA	Montium TP	2	100
GPP	ARM920T	4.7	12
Reconfigurable ASIC	**Our design**	0.24	188
	(excl. RAM and ROM)		

6.3 Comparison of Different Implementations

Table 3 depicts a quick comparison (for butterfly computation) of different designs mentioned above. We have chosen to compare FFT (butterfly), because we know that in FFT, we have about 50% of redundant hardware in our implementation.

It is important to note that all these designs, except ours, have lot of data memories to store data operands and intermediate and final results. For example, in the FASRA-ASIC, approximately 0.5 mm^2 area is occupied by the RAM memory. For our system, we need an additional area corresponding to RAM (256 $\times$ 16) and ROM (64 $\times$ 16). This area will approximately be equal to 0.03 mm^2 in CMOS12 Philips process.

7 Conclusions

This section concludes the work and summarizes the achievements and lessons learnt through this project:

- In our SDR receiver, the Bluetooth channel selection algorithm requires more data path resources than the HiperLAN2 OFDM demodulation. On the other hand, HiperLAN2 demodulation needs more memory and memory bandwidth.
- By incorporating limited flexibility in our system, we are able to reduce the total hardware required to implement the SDR receiver compared to the implementation in which each receiver is implemented individually. This is shown in Table 3. It can be concluded that an area reduction of about 25–30% can be made in the combined implementation compared to the individual implementations of the two receivers.
- Dynamic reconfiguration in our system allows time-sharing of hardware resources by pipelining algorithms, thus increasing the performance of overall system at the cost of some latency.
- For state-of-the-art designs, an ASIC implementation with minimal flexibility can easily outperform the flexible implementation. The results of our ASIC-

Table 4 Area requirements of SDR receiver

Component	Sum of separate implementations	Combined implementation
Computation area [μm^2]	840,000	600,000
RAM	352x16	256×16
ROM	64x16	64×16

like implementation were shown to be superior to the implementation on more flexible systems.

- A GPP (ARM920T)-based implementation requires 20 times more area and computes 15 time slower than our ASIC-like implementation. A domain-specific processor like Montium TP requires 15 times more area than our implementation to meet the SDR computational requirements.
- On the other hand, flexible solutions like the Montium TP and GPP are superior to our design in terms of suitability for different algorithms and ease of implementation.
- So a design decision based on the performance requirements and implementation costs needs to be taken before deciding on the platform and methods for the final implementation of a DSP system.
- It can be concluded that the performance of ASIC > ASIP, ASIP > DSRA and DSRA > GPP, while the flexibility of ASIC < ASIP, ASIP < DSR and DSRA < GPP.
- By introducing pipelining in the data path, we are able to perform computations at higher speed than a non-pipelined data path (Table 4).

The 16-bit data path performs satisfactorily for the chosen SDR algorithms.

A high-level description language, like SystemC, can be used to design VLSI (very large-scale integration) systems. The benefits are in timely and easily realization of a design. The main drawback is that efficiency of synthesized code is largely dependent on the tools.

Almost all of the systems use either phase modulation or OFDM modulation [5]. So the suitability of our system for phase-modulated and OFDM receivers implies that our design can be used in a number of systems.

8 Future Work

In our FFT implementation, we have not performed the bit reversing operation on the output. This should be taken into consideration in the next stage of the receiver implementation while reading the data from the memory. Also, the data path may be changed to heterogenous DPUs to reduce the area. The control section can be optimized further. The butterfly computations in the last stage of FFT can be simplified to simple addition-subtraction operations. The overflow and underflow conditions need to be incorporated in the complex multiplication and addition functions. Also, extensive power consumption analysis in the system still needs to be done.

The computational complexity of receiver can be simplified by reducing the order of filters or increasing the decimation. Currently, the decimation factor is 4, which gives data rate of 5 MSPS for 1 MHz Bluetooth channel. If we change the decimation factor to 6, the data rate will be 3.33 MSPS for 1 MHz channel (a theoretically sufficient number). Also, the sample rate reduction block after ADC block may also be modified.

In the broader context, the design was made as a subsystem of SDR transceiver system. Also, the other blocks of the SDR receiver need to be implemented in hardware. The SDR transmitter needs to be designed and implemented as well.

References

1. A. Kapoor, A Reconfigurable Architecture of Software-Defined-Radio for Wireless Local Area Networks, 2005
2. A. Kumar Kaushik, A Comparative Study of Software Defined Radio and Cognitive Radio Network Technology Security, pp. 104–110
3. M.B. Blanton, An FPGA software-defined ultra wideband transceiver. Master Sci. (2006)
4. X. Zhang, J. Ansari, M. Arya, P. Mähönen, Exploring parallelization for medium access schemes on many-core software defined radio architecture. Proc. Second Work. Softw. Radio Implement. Forum - SRIF **13**, 37 (2013)
5. F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, W. Idler, Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration. J. Lightwave Technol. **34**(7), 1599–1609 (2016)
6. C. Zhang, Dynamically reconfigurable architectures for real-time baseband processing, no. May. 2014
7. L. Zhao, H. Shankar, A. Nachum, 40G QPSK and DQPSK modulation, Inphi Corporation, 2008
8. P. Dong, C. Xie, L. Chen, L.L. Buhl, Y.-K. Chen, 112-Gb/s monolithic PDM-QPSK modulator in silicon. Opt. Express **20**(26), B624–B629 (2012)
9. H.D. Nataraj Urs, V.S. Reddy, Implementation and analysis of low frequency transceiver for SDR platforms. J. Adv. Res. Dyn. Control Syst. **10**(04–Special Issue), 1–9 (2018)
10. C.Y. Chen, F.H. Tseng, K. Di Chang, H.C. Chao, J.L. Chen, Reconfigurable software defined radio and its applications. Tamkang J. Sci. Eng. **13**(1), 29–38 (2010)
11. P. Suarez-Casal, A. Carro-Lagoa, J. A. Garćia-Naya, L. Castedo, A multicore SDR architecture for reconfigurable WiMAX downlink, in Proceedings of 13th Euromicro Conference on Digital System Design: Architectures, Methods and Tools (DSD) 2010, no. Cc, pp. 801–804, 2010
12. H. Goyal, J. Saxena, S. Dewra, Performance evaluation of OWC using different modulation techniques. J. Opt. Commun. **37**(4), 33–35 (2016)
13. V. Kumar, H.D. Nataraj Urs, R.V.S. Reddy, Software defined radio: Advancement to cognitive radio and basic challenges in spectrum sensing. Asian J. Eng. Technol. Innov. **4**(7), 166–169 (2016)
14. C. Chaitra, H.D. Nataraj Urs, Performance of SDR transceiver using different modulation techniques. Int. J. Adv. Eng. Res. Sci. **3**(5) (2016). ISSN: 2349-6495
15. M.S. Karpe, A.M. Lalge, S.U. Bhandari, Reconfiguration challenges & design techniques in software defined radio. Int. J. Adv. Comput. Res. (2013)

Methods and Application of 3D Printing in Implantable Medical Devices

Vijay Kumar, Sambhrant Srivastava, Saurabh Kumar Singh, Saroj Kumar Sarangi, and Ambrish Maurya

1 Introduction

Medical devices and implants are nowadays becoming the need of society. These devices are used to replace or support the biological structures of humans. Some of the implants are used for orthopedics like knee replacement, dental implants, cardiovascular valves, stents to maintain the flow of blood, etc. [1]. But to substitute or support the biological structures of the body, an exact replacement is necessary. To bridge the gap, additive manufacturing is the technology that is commonly used. This is because 3D printing can make implantable medical prosthesis complex shapes in a well-organized manner. These characteristics of regenerative technology are helping to make complex medical device processing easy. This makes 3D printing of complex shapes easy and significant [2]. In printing medical parts, a high precision is required. Thus 3D printing is useful for high precision and customized printing of medical implants [3]. A model which is to be printed similar to complex shapes needs to follow five simple steps. Firstly, a body part is targeted for which a 3D model has to be generated, then a picture of the body part is generated by using CT/MRI scan, and then the file is formatted as per the printing requirement, also selecting suitable material for manufacturing (Fig. 1). The 3D printing file is monitored while its printed in three dimension form. The medical implants are printed using layered by layered process of materials. The layered design form also known as "sliced" design is then sent to a 3D printer, where it firstly creates a foundation layer, and then the object is created slice by slice on the top. Finally,

V. Kumar
Rajkiya Engineering College Azamgarh, Deogaon, Uttar Pradesh, India

S. Srivastava (✉) · S. K. Singh · S. K. Sarangi · A. Maurya
National Institute of Technology, Patna, Bihar, India
e-mail: sarojsarangi@nitp.ac.in; amaurya.me@nitp.ac.in

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_12

Fig. 1 A workflow in 3D printing

the process stops as the customized design is achieved. This not only increases the research knowledge and ability of the next-generation surgeons but also maintains a specified connection with the patients [4]. The customized design of implantable devices is increasing because it is less time-consuming and economical [5–8].

To better represent the patient's anatomy, many printing techniques and materials are now accessible. Unlike biological tissue, most accessible printing materials are stiff and thus unsuitable for flexibility and elasticity [9]. As a result, there are current materials that can bridge the gap between genuine and recreated anatomies, particularly when it comes to soft tissue [10, 11]. An overview of 3D printing's applicability in the medical profession is offered in this report, with emphasis on its utility and limitations, as well as how it could benefit surgeons.

2 3D Printing Manufacturing Process

3D printing is a technology that enables the construction of complex and diverse physical entities based on a 3D model [12]. Since 1986, with the first stereolithographic (SLA) systems, 3D printing techniques have advanced significantly. ASTM (American Society for Testing and Materials) internationally defines seven 3D printing technology processes, each of which is represented by one or more commercial technologies [13].

2.1 Stereolithography

The first feasible fast prototyping technology in 3D printing is stereolithography equipment. Photo-induced polymerization is used to construct multilayer structures employing highly cross-linked polymers in this rapid manufacturing approach [14]. Printing with this technology comprises three primary phases, independent of the category: exposure to light or laser, platform movement, and resin replenishment.

The control system monitors the laser beams' scan on the powder layer so that they can be connected and raise the powder temperature to its melting point. The workbench decreases by one-layer thickness; after the first layer is finished, a new layer of dense powder is spread uniformly and sintered to create the full model. Variables like the intensity of the light source, speed of scanning, amount

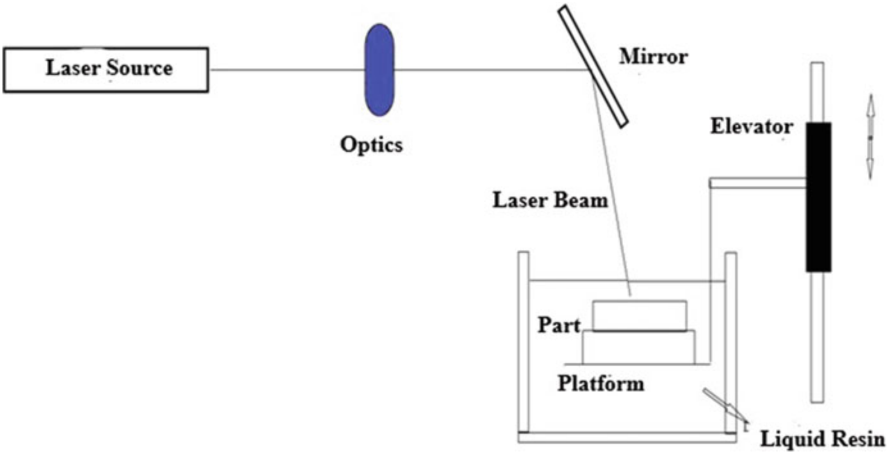


Fig. 2 Stereolithography apparatus

of the resin monomers, and number of photo-initiators can be regulated to get the good qualities of the final product [15]. Figure 2 shows its basic components: an ultraviolet laser, a lifting platform, a scraper, a fluid horizon, and a photosensitive resin. SLA (stereolithography apparatus) is frequently employed in tissue and orthopedic repair as well as for printing skull and hip bones. In the fabrication of dental stone casts, stereolithography revealed more clinical accuracy than other digital/analogue methods [16]. High moldability and stable working, high precision, and high finishing of the printed parts are the results of stereolithography. But the device is relatively large, and the size of optical pixels still has limitations, limiting the microstructure in the aircraft [17]. The addition of nanoparticles to a polymeric matrix increases mechanical characteristics [18]. Other coupling agents like ceramic fillers improve stress distribution and protect the printed structure against fracture. To deal with the problem of microbial colonization of oral implants, antimicrobial compounds have been added to resins [19, 20]. The use of a low-cost laser semiconductor in this 3D printer approach allows for the efficient production of large structures.

2.2 Laminated Object Manufacturing

Laminated object manufacturing was founded by the Helissy Company in 1986 and has grown substantially since its inception in 1991. Laminated item manufacturing is among the most advanced 3D technologies. The laminated object production operating principle is shown in Fig. 3. Thin materials, such as paper and plastic film surfaces, are pre-coated with a heat-melting coating in this process. A heated roller is run over a sheet of material, melting the adhesive and pressing it onto

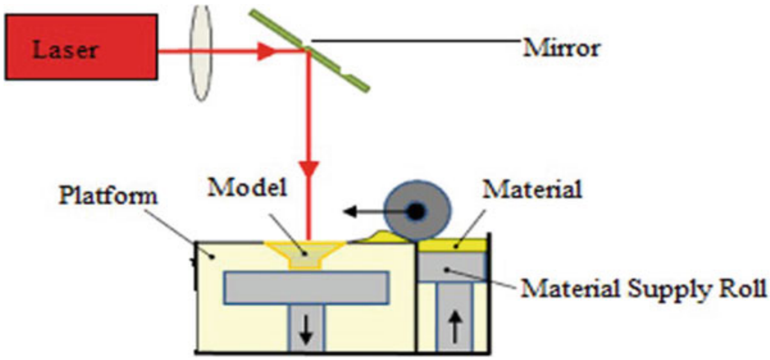


Fig. 3 Laminated object manufacturing

the platform to make an item. The desired design is then cut using a computer-controlled laser or blade. Excess material is also removed in a crosshatch pattern by the laser, making it easier to remove once the object is fully printed. The build platform is lowered by around one-sixteenth of an inch, the normal thickness of one layer, once one layer of the object is built. After that, new material is pulled across the platform to form the next layer. This procedure is repeated until the complete object is created. The processes of solid layered production processes are graphical processing, the manufacture of substrates, the fabrication of prototypes, waste elimination, and post-processing. Laminated object manufacturing includes various components like a laser, rolling material, heated roller, layered part with support material, rolling platform, waste and rolling outline, and crosshatch layer. It may be used for orthopedics such as the temporal bone, dental jaw, and jaw. Since the processing materials using the layered solid forming process are easily accessible, process raw material costs are low, and the accuracy of the produced workpieces, which is appropriate for the mounting of big workpieces, is generally excellent. The mechanical features of the technology, however, are quite low, and the efficiency of the multilayer solid molding process is weak compared to other printing techniques [21].

2.3 Selective Laser Sintering (SLS)

SLS is a process in which a laser with a high-energy beam is used to cause the fusion of powdered source materials. The laser fuses the powder into a solid layer. To allow the laser to sinter the next layer of powder, the platform will be lowered. This type of structure formation eliminates the need for additional material support during printing, as the powder surrounding the structure provides support [22]. As the polymers' absorption at the wavelength is substantially more efficient, a CO_2 laser is usually selected. The principle of selective laser sintering is to install two-

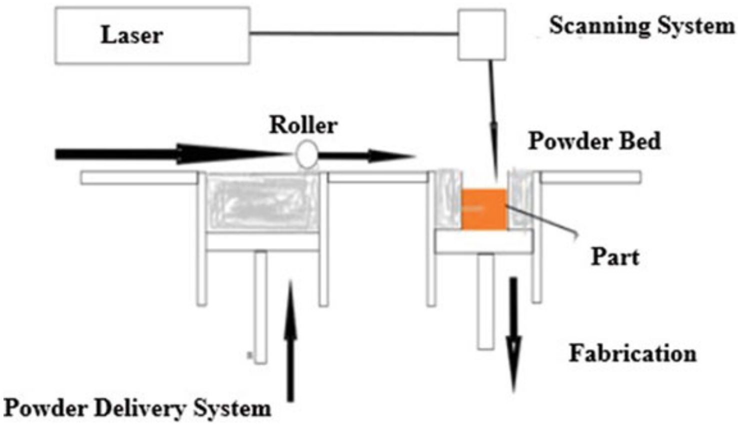


Fig. 4 Working principle of selective laser sintering

piston cylinders (one for powder and the other for molding) in a locked molding chamber (Fig. 4). The laser's job is to raise the temperature of the powder to near but not quite at its sintering point. The solid power is converted to a semiliquid condition during the sintering process. The platform on which the initial layer is placed lowers by 0.1 mm, allowing the laser to sinter a new layer of powder. This sintering and fusion process is repeated until the object is finished printing. The object is allowed to cool after printing is completed [23]. The non-sintered powder serves as the foundation for the first layer of the part. The laser beam is scanned after the second layer of information, and the second laying made on the first layer is also sintered and solidified. As a result of stacking layer by layer, a 3D solid part is produced [24]. Because of the non-requirement of structural support during printing in this technology, it is substantially faster than SLA and FDM (fused deposition modeling) printers [25–27]. Metal-based SLS can use a variety of metal powders, including stainless steel alloys and titanium alloys. Metal-based SLS, unlike polymer-based SLS, requires the flooding of protective gas into the printing chamber to keep the metal powder from oxidizing when heated to a high sintering temperature.

Scaffolds can be used to encourage bone recycling with this technology. Selective laser sintering has some of the most prominent advantages of a range of materials and a fast manufacturing cycle. The inconvenience of this technology is that the created parts are low in strength and have poor surface quality [28].

2.4 Fused Deposition Modeling

In 1988, Scott Crump invented the production of fused deposition for the first time. Fused deposition modeling (FDM) is a softening technique for polymers,

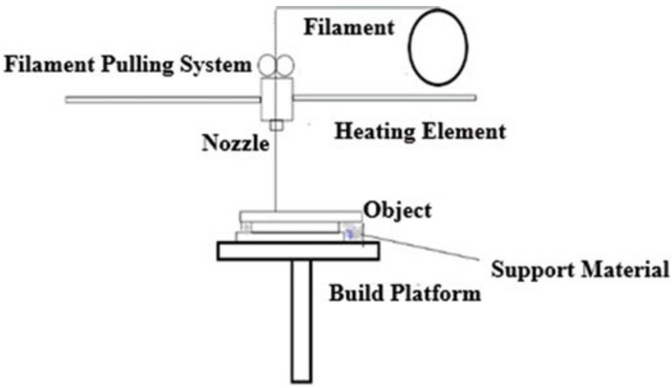


Fig. 5 Working principle of fused deposition manufacturing

composites, and metal alloys. It is the second most popular 3D printing technology after SLA [29] and significantly less expensive than other AM procedures [30].

The production of fused deposition is employed largely for the molding of thermoplastic polymers. In the FDM process (Fig. 5), the thermoplastic polymer wire is fed into a heating extruder over the nozzle by a feeding unit and onto the forming platform, laying and solidifying depending on the specified process heat. When one cross-sectional layer is finished, the worktable falls a given height, and on the formed section, another layer is machined to repeat until a workpiece is produced which is particularly common to bone techniques [31]. It can be machined to a significant degree to lower the cost of forming, with a range of materials and high material use. However, this technology has a low forming rate and a strip sense on the surface of the piece and a low production accuracy [32].

2.5 3D Printing

In the 3D printing operating concept (Fig. 6), powder is deposited flat on the groove. With the computer control, the nozzle selectively sprays the cut piece information so that a powder section is attached and a cross-section is progressively formed. The workbench descends a particular height when the section is printed and the following layer is connected until the component is completed fully. The prototype should be placed in a heating oven for further treatment or synchronization. This technology can all be printed on cheekbones, mandibles, and skulls. A large range of materials may be processed through the 3D printing process with rapid forming speed, which is employed for the printing of workpieces with difficult structures. However, surface quality and the precision of the printed product are poor, and the prototype component has a low strength which needs post-processing [33–35]. This technique is becoming ever more and more attractive by the day due

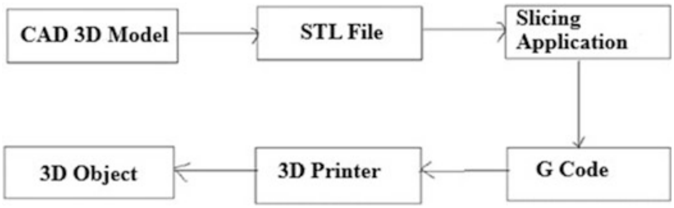


Fig. 6 Three-dimensional printing

to its low cost, small size, high molding speed, and printing of color products, among other advantages [36, 37]. Furthermore, the production of 3D printing differs from the standard coating technique. It is layered on top of the layer called "coating technology." As a new way of printing, the development of free 3D printing represents a significant achievement in biomedical engineering, since softer substances are mostly employed in the biomedical field, such as biomaterials, live cells, biopolymers, and silicones. The free-style 3D printer method enables scientists to create these materials with low viscosity and examine a broader variety of soft materials. There are an increasing number of articles in the fields of frequency engineering, vascularization, and 4D printing as 3D free-form printing methods gain prominence. This technique has more advantages than a standard type of implant in that complex body parts for each patient can be easily designed and fabricated.

2.6 Photopolymer Jetting

A dynamic printing head and a photo-polymerizable material are combined in this method. Jets of the light-sensitive polymer from the inkjet-type printing head are projected onto a building platform. Curing is done on the descending platform [38]. For easy removal, soft material is used for printing of supporting structure. This technology allows for the printing of a variety of resins and waxes for casting. Complex and highly intricate items can also be printed with a resolution of 16 μm using silicone-like rubber materials. Crowns or anatomical study models can be made from printed items. Another benefit of this technology is that implant drill guides can be made quickly and inexpensively while maintaining high quality. Multiple printing heads are used in 3D jet printers [39, 40].

The printing head and the working platform both move independently in opposite directions. An ultraviolet light source is used to harden each blasted layer of resin or wax. Rapid manufacture, surface smoothness, and cost-effectiveness are all advantages of this technology. The disadvantages include difficulty in entirely removing the support material, skin discomfort, inability to be heat-sterilized, and the high cost [41, 42].

2.7 Powder Binder Printer

A modified inkjet head is used in this AM method to create liquid adhesive droplets. These droplets are released by the inkjet head to infiltrate a layer of powder beneath it. This technique is repeated until the final product is made by the addition of a new layer of powder. This technology is most commonly used in dentistry to produce study casts or prototypes. Printed products, on the other hand, are flimsy and inaccurate. The manufacturing method, however, is clumsy due to the usage of powder [43, 44].

3 Importance of 3D Printing (Medical Field)

Implantable medical gadgets that improve patients' quality of life are becoming more popular as people's health awareness grows [25]. Processing can be problematic when implantable medical devices have a complicated form. Implantable medical devices may be made with any complex shape using 3D printing technology without having to deal with any processing concerns. It can also manage sophisticated implantable medical device design and production issues in 3D printing technology. It is briefly explained how 3D printing can be used to manufacture implantable medical devices.

3.1 3D Printing Materials (Implantable Medical Device)

Implantable medical devices are strongly connected to medical and material science research, notably the development of biomaterials. Creation of implantable medical devices, biological materials are substances that are capable of diagnosing, treating, repairing, replacing, or improving the function of tissues and organs in the body. They can be both natural and synthetic. Materials made of biomedical polymers, metals, ceramics, and derived materials are some examples of 3D implantable medical devices that can be separated based on their properties.

3.1.1 Polymer Materials

All sectors of medicine use biomedical polymer materials, which are among the oldest and most extensively used materials in biomedicine. They are divided into two categories based on their features: nondegradable and degradable. Polyethylene, polypropylene, and polyformaldehyde are nondegradable materials [45]. In addition to having excellent physical and mechanical properties, this type of material may also last in a biological setting for an extended amount of time. Degradable biological components include collagen, cellulose, and chitin [46, 47]. Degradation of this material occurs due to biological action, which absorbs and partially removes some of the degraded products in the human body. Biomedical polymers are frequently employed in cardiovascular stent repairs, as well as soft tissue and hard tissue reconstruction.

3.1.2 Metal Materials

In the field of biomedical implants, biomedical metal materials are metals or alloys that are used to treat surgical implants. Stainless steel [48], titanium and its alloys [49, 50], cobalt-molybdenum-chromium alloys [51], and precious metals are the primary materials used. Medical biomedical metals are clinically used metals. This material has excellent mechanical properties, is corrosion-resistant, and is biocompatible despite its high fatigue and mechanical resistance. This component is employed in bone and joint replacements, spinal implants, and cardiovascular implants, among other things.

3.1.3 Ceramic Materials

Biomedical ceramics have also evolved into inorganic biomedical materials. They first started in the early eighteenth century as biomedical materials and were used clinically in China in the 1970s. Biomedical pottery is split into ceramics that are inert, physiologically active, and biodegradable. Alumina, zirconia, and carbon are organic ceramics. These materials are structurally stable and resistant to good abrasion, with great resistance and stability. Bioactive ceramics include bio-glass [52, 53] and tricalcium phosphate [54] that, by chemical body reactions, develop solid chemical connections with tissue. Biodegenerative bioceramics can encourage the development of new bones since they, like β-tricalcium phosphate bioceramics, can be absorbed into the living bones. Artificial hip joints, artificial bones and valves, etc., may be made with biomedical ceramics. Little strength and toughness are the primary difficulties of biomedical ceramics in clinics.

3.1.4 Composite Materials

Biomedical composites are made up of two or more distinct materials. Because they are made up of numerous different materials, composites have unique properties that set them apart from their parts. As a result, they improve the performance of other materials. Depending on how they are used, these new material features could be beneficial to human health. As a result, the composite must not only have the desired physical and chemical qualities but also meet biocompatibility standards. Bone cement, coating compounds, and nano-phosphoric lime are all examples of biomedical materials. To repair or replace human tissues and organs, as well as to improve their functioning and produce artificial organs, biomedical composites are frequently used [55, 56].

3.1.5 Derived Materials

Bio-derived materials sometimes referred to as bio-regenerative materials are generated from biological tissues and are created by a specific processing method. Because they lose their life after a certain treatment, biological tissues are inanimate objects. Tissue topologies and functionalities are similar in bio-derived materials. In the preservation and replacement of human dynamic processes, bio-derived materials play an important role because of their oral makeup, which is similar to genuine issues. Artificial heart valves, masks, bone remodeling, and vasculature regeneration frequently use biological material.

3.2 Medical Devices Printed in 3D

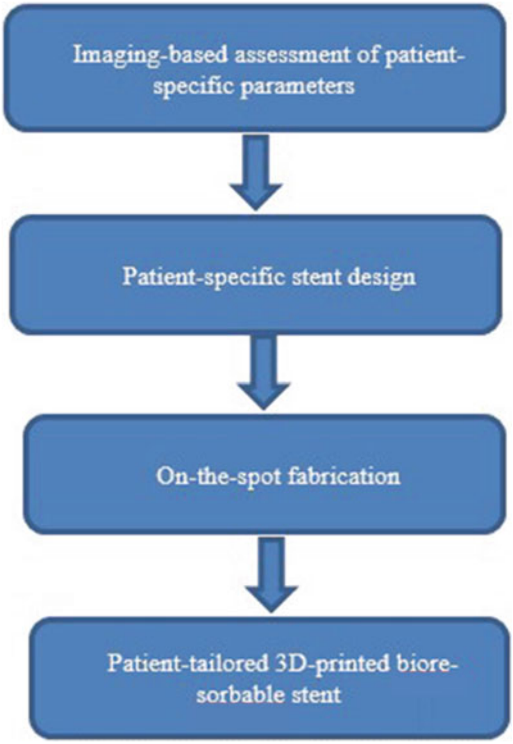
3.2.1 Vascular Stents

CAD (coronary artery disease), a condition that adversely affects the heart, heart valves, and blood vessels, is becoming more widespread and increasingly lethal. Cardiovascular disease takes the lives of millions each year, and it is considered the primary cause of death [57]. Research scholars have been investigating numerous issues involving cardiovascular illness in various places. To treat cardiovascular illnesses, options include medication, surgical procedures, and an interventional vascular stent treatment [58–60]. External surgery might lead to additional harm and can take a long time to heal. Additionally, the effects of treatment may not be immediately clear, and extensive recuperation periods are necessary. Using this new minimally invasive surgery is far better than the previous techniques as the postoperative recovery time is shorter and there is much less stress and suffering. Interventional vascular stent therapy has grown in popularity and become the standard treatment for cardiovascular disease. The 3D printing procedure for vascular stent production is shown in Fig. 7.

Flege et al. [61] used to produce the biodegradable vascular scaffolds, utilized the selective laser melting (SLM) 3D printing method, and dipcoated and spray treated the stent to ensure it was smooth. Figure 8 illustrates the biodegradable stent. Experimental results have shown that the materials and the procedures are great for the biocompatibility of the scaffold and the implant. Figure 8 represented the SLM printed vascular stent.

Kaesemeyer et al. [62] reported the biodegradable vascular stent production using the RSF method as shown in Fig. 9. Lactide, glycolide, ε caprolactone, and lovastatin form the raw materials for the vascular stent. Vascular stents serve not only as scaffolding but also as a delivery device that can inject medication into the

Fig. 7 Preparing blood vessel stent using 3D printing technology



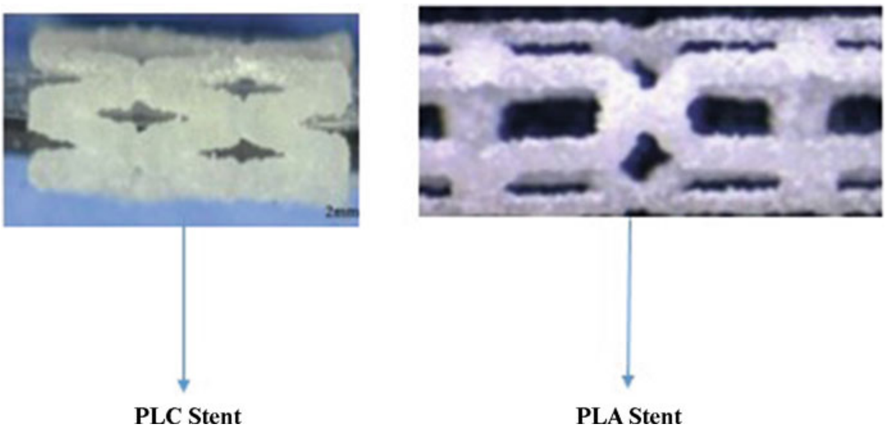


Fig. 8 Vascular stent printed by the SLM Technology

body to help repair damaged endothelium, thereby protecting against blood clotting and providing the user with better quality of life. A biodegradable stent is shown in Fig. 9.

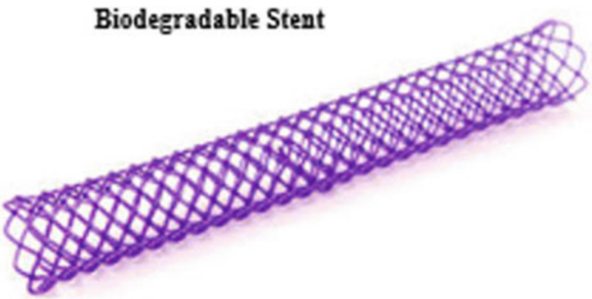


Fig. 9 Biodegradable vascular stent

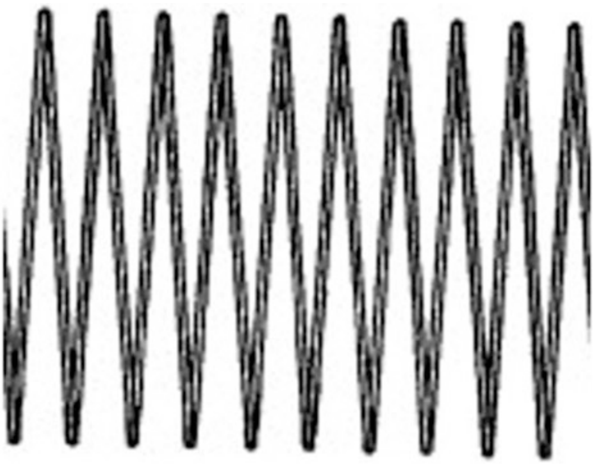


Fig. 10 Spiral vascular stent

Park et al. [63] created a vascular stent that has a spiral bioabsorbable PCL (polycaprolactone) design (Fig. 10). The experimental implant consisted of a covering made of PLGA (poly(lactic-co-glycolic acid)) and PEG (percutaneous endoscopic gastrostomy) combined with sirolimus that was made via ultrasonic atomization. The vascular intimal hyperplasia has been decreased, and the release kinetics of Rolimus are shown in a continuous curve.

In 3D printing technology, Van Lith et al. [64] employed the manufacture of the micro continuous liquid interface to print curable biomaterial based on the use of bioabsorbable citrate and developed the fast, high-resolution, custom-designed bioabsorbable vascular stent. Misra et al. employed 3D printing technology to manufacture the vascular skin with twice drug release potential using a doped composite material of biodegradable polymer PCL and graphene nano flakes.

3.2.2 Prosthetic Valve

Heart valve disease is becoming more common and clinically significant. There are no biological diagnostics or therapy techniques that are clinically successful. Replacement with a prosthetic valve is the sole option. Tissue engineering has the potential to significantly alter heart valve disease by delivering a living valve conduit capable of growth and biological integration.

The job of heart valves is to ensure that blood flows in the appropriate direction through the heart. Individuals' health will be jeopardized by heart valve disease, which will harm their normal quality of life. As people get older, cardiac and penile valve diseases become more common [65]. Cuban disorders are common in today's adults, and they can lead to cardiac valve damage, hyperlipidemia, high blood pressure, and chronic kidney disease. Heart valve disease is a common cardiac ailment in today's medicine. Injury to the heart function and valve loss can occur as a result of cardiac damage and heart failure, preventing normal blood flow. For the treatment of cardiovascular disease, there are three options: medicine, outside surgery, and procedure. A replacement mechanical heart valve, also known as valvuloplasty, is an effective treatment for those who are at high risk of heart valve disease, as well as a radical cure. The precision and stability of the impressed 3D cardiac valve can be changed to improve accuracy and stability in a variety of individuals; the limited body refuse's reaction raises the success rate of cardiac valve replacement [66].

Figure 11 depicts the substitution of biological cardiovascular valves and mechanical valve substitution. Mechanical and biological valves and interventional

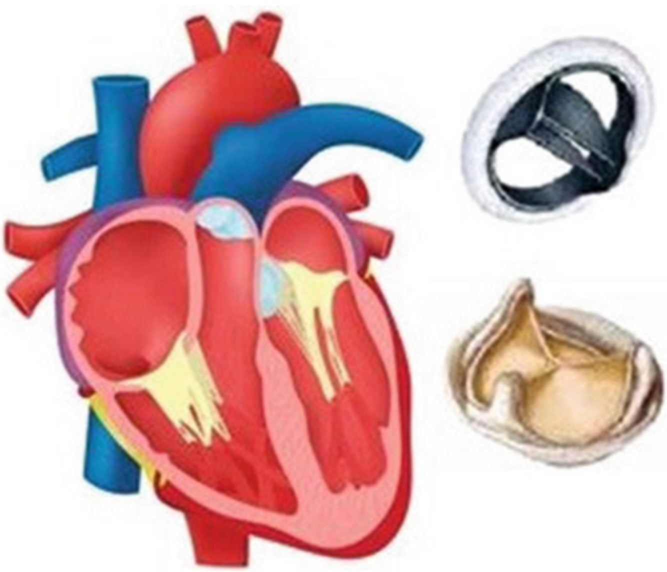


Fig. 11 Mechanical and biological heart valves

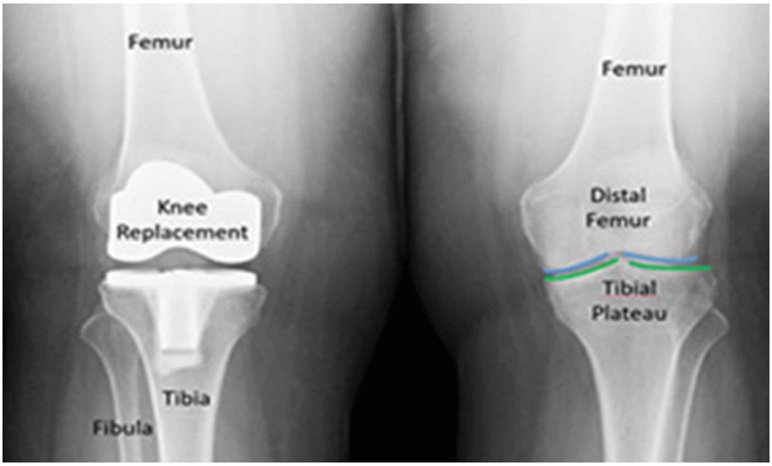


Fig. 12 Human knee replacement x-ray

and tissue-designed valves were developed via the stages of the artificial heart valves [67, 68].

3.2.3 Artificial Joint Prostheses

Degenerative arthritis is a term used to describe osteoarthritis (OA). The condition damages the cartilage in the joints, eventually leading to degenerative joint disease throughout the entire joint, resulting in joint dysfunction and loss of normal joint motion. In the clinical treatment of joint dysfunction, replacement surgery is used as a difficult process in the field of orthopedics, restoring joint shape and function.

Total arthroplasty is the most efficient treatment for moderate to extreme knee arthritis (TKA). On both sides of the knee, surgical implants must be used to replace the articular surfaces of the femur and tibia. The x-ray scans were complete with knee arthroplasty, as shown in Fig. 12.

.

3.2.4 Engineering of Hard and Soft Tissues

Patients with hard tissue abnormalities who need implant insertion often benefit from bone augmentation surgery. The lack of adequate bone leads to implant failure due to a lack of support. This means that before implant placement, bone augmentation may be required in certain circumstances. An autograft made from the patient's bone that has osteoinductive and osteoconductive qualities is the best bone graft to be inserted. Allografts, xenografts, and alloplastic are further options for bone grafts. Grafting is used to bone to repair and rebuild damaged bones which

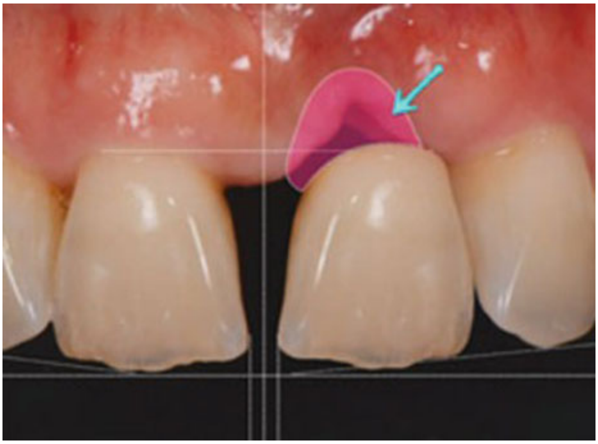


Fig. 13 Management of hard and soft tissues through 3D printing

contains features, such as biocompatibility, similar shape and design to defects and do not provoke an immunological response as well as porosities enable tissues to receive with immune cells and ingest nutrients (Fig. 13) [69]. As hard tissue engineering scaffolds, a variety of biomaterials such as polymers, bio-ceramics, and composites are accessible. Angiogenesis and osteogenesis are both dependent on stem cells, and they should be able to be incorporated into biomaterials to improve regeneration. 3D-printed scaffolds can be fine-tuned with stem cell-infused inks. In this printing technique, stem cells may be precisely positioned. 3D-printed complex scaffolds allow for closer contact with the bone, which speeds healing and improves the aesthetic appearance over conventional ones [70].

Although an autograft taken from the palate or tuberosity is optimal, it causes postoperative discomfort for the patient. Furthermore, 3D printed soft tissue grafts allow for more complex patterns and precision in covering the defect. They are also more well-liked by patients because they produce less postoperative discomfort and speed up the healing process [71].

3.2.5 Peripheral Hip Diseases

Primary and metastatic bone malignancies are most commonly found in the hip [72] which have an unknown cause and are linked to a variety of oncogenes, including circRAB3IP and RNA00511 [73, 74]. Even though they have certain diagnostic tools that are both proven and true and cutting-edge, their therapy and postoperative functional restoration are highly tough [75]. Amputation and tumor resection were once the only options for treating malignant bone tumors surrounding the acetabulum, and prosthesis replacement is now a common treatment for significant bone abnormalities around the acetabulum [76]. Additional treatments for primary

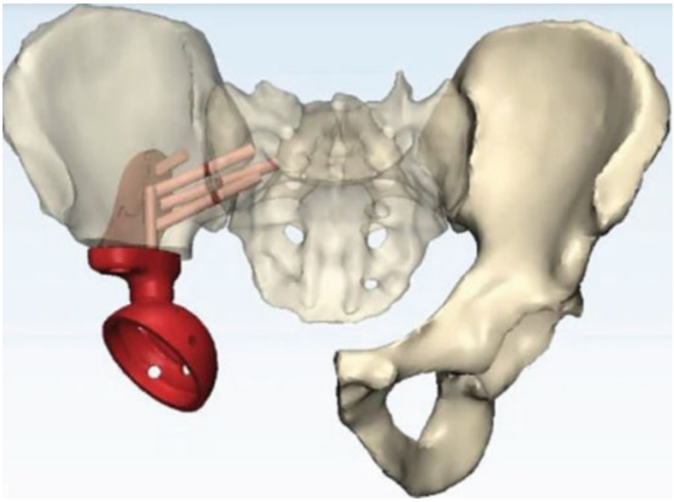


Fig. 14 3D printing (hip diseases)

tumors include allogeneic hemi-socket replacement, proximal femoral eversion implantation, and tumor bone activation prior to implantation [77]. Calycosin-exerting potential anti-OS activities have been discovered using new approaches such as fungal-derived materials [78–80]. Despite this, due to the high likelihood of postoperative complications, there has been a long-running controversy over the optimal treatment for this condition. As a result, the clinical efficacy of 3D printing technology is a hot topic of discussion.

The range of tumor resection can be greatly reduced with 3D printed navigation templates, resulting in less trauma for patients. Using five female cadaveric pelvises from the anatomy laboratory, Sallent et al. [81] compared the accuracy of patient-specific instrumentation (PSI) to the standard manual technique in pelvic tumor resections. PSI osteotomy was used on the left pelvis, whereas the normal manual method was used on the right pelvis. PSI enhanced the precision of pelvic tumor removal as a result (Fig. 14) [81].

4 Conclusion

3D printing technology is now one of the most used methods of implantable clinical devices, which has drawn the attention of relevant researchers from various countries. After becoming the most relevant technology for medical implants, there are a lot of challenges from materials to the processing of medical implants. After accomplishing this, a major challenge of meeting the industry standards is in the next step. The item after this step is not yet complete here; material investigation and its criticality according to the need are important. The material that is used in medical implants should have the best properties to become a biocompatible material and move toward advancement to 3D printing innovation. In the seven techniques of manufacturing medical implants, stereolithography is frequently used in tissue and orthopedic repairs as well as in the printing of skull and hip bones; it is proved to be a more clinically accurate method with a high degree of finish along with an issue of microbial colonization of oral implants. In the sequence, laminated object manufacturing is also best used for orthopedic such as jaws and dental jaws. In selective laser sintering, bone recycling is done with the help of a scaffold. Fused deposition modeling technique is the cheapest among all, but the production rate is proved to be the lowest. In addition to all the techniques, 3D printing is best to use where most complex parts are to be designed and fabricated. 4D printing is one of the free form techniques that is developing nowadays for manufacturing medical implants like building bones on a large scale. 5D printing reduces time and money when compared to 3D printing. The main advantage of this technology is that it allows you to create parts with curved layers that are stronger and more durable. A 5D printed model allows for the creation of a bone substitute for surgery. Because human bones are not flat, artificial bones that are curved and have a curved surface must be made utilizing 5D printing. These bone implants are extremely durable also. 3D printing will play an increasingly essential role in precision medicine and tailored treatment in the future, as it is the medical industry's future. The medical implant industry is booming and will be beneficial to our country's research.

References

1. W. Khan, E. Muntimadugu, M. Jaffe, A.J. Domb, Implantable medical devices, in *Focal Controlled Drug Delivery, Advances in Delivery Science and Technology*, ed. by A. Domb, W. Khan, (Springer, Boston, 2014), pp. 33–59
2. K.K. VanKoevering, D.A. Zopf, S.J. Hollister, Tissue engineering and 3-dimensional modeling for facial reconstruction. Facial Plast. Surg. Clin. North Am. **27**(1), 151–161 (2019)
3. X. Du, B. Yu, P. Pei, H. Ding, Y. Zhu, 3D printing of pearl/CaSO4composite scaffolds for bone regeneration. J. Mater. Chem. B **6**(3), 499–509 (2018)
4. D.B. Jones, R. Sung, C. Weinberg, T. Korelitz, R. Andrews, Three-dimensional modeling may improve surgical education and clinical practice. Surg. Innov. **23**(2), 189–195 (2016)
5. I.M. Mazher, P.F. Angus, K.M. Santosh, I. Gibson, Customised design and development of patient specific 3D printed whole mandible implant, in Proceedings of the 27th Annual International Solid Freeform Fabrication Symposium, Austin, August 2016
6. D.C. Ackland, D. Robinson, M. Redhead, P. Vee Sin Lee, A. Moskaljuk, G. Dimitoulis, A personalized 3D-printed prosthetic joint replacement for the human temporomandibular joint: From implant design to implantation. J. Mech. Behav. Biomed. Mater. **69**, 404–411 (2017)
7. K. Phan, A. Sgro, M.M. Maharaj, P. D'Urso, R.J. Mobbs, Application of a 3D custom printed patient specific spinal implant for C1/2 arthrodesis. J. Spine Surg. **2**(4), 314–318 (2016)
8. L.E. Diment, M.S. Thompson, J.H.M. Bergmann, Clinical efficacy and effectiveness of 3D printing: A systematic review. BMJ Open **7**(12), e016891 (2017)
9. J. Garcia, Z. Yang, R. Mongrain, R.L. Leask, K. Lachapelle, 3D printing materials and their use in medical education: A review of current technology and trends for the future. BMJ Simul. Technol. Enhanc. Learn. **4**(1), 27–40 (2017)

10. R.J. Mondschein, A. Kanitkar, C.B. Williams, S.S. Verbridge, T.E. Long, Polymer structure-property requirements for stereolithographic 3D printing of soft tissue engineering scaffolds. Biomaterials **140**, 170–188 (2017)
11. W. Zhu, X. Ma, M. Gou, D. Mei, K. Zhang, S. Chen, 3D printing of functional biomaterials for tissue engineering. Curr. Opin. Biotechnol. **40**, 103–112 (2016)
12. C. Wang, J. Lai, K. Li, et al., Cryogenic 3D printing of dual delivery scaffolds for improved bone regeneration with enhanced vascularization. Bioactive Mater. **6**(1), 137–145 (2021)
13. ASTM, Standard Technology for Additive Manufacturing-General Principles/Terminology, ASTM ISO/ASTM 52900:2015 (E), West Conshohocken, 2015
14. D. Fan, Y. Li, X. Wang, T. Zhu, Q. Wang, H. Cai, W. Li, Y. Tian, Z. Liu, Progressive 3D printing technology and its application in medical materials. Front. Pharmacol. **11**, 122 (2020)
15. R. Jain, B.S. Supriya, K. Gupta, Recent trends of 3-D printing in dentistry-a review. Ann. Prosthodont. Rest. Dent. **2**, 101–104 (2016)
16. P. Aly, C. Mohsen, Comparison of the accuracy of three-dimensional printed casts, digital, and conventional casts: An in vitro study. Eur. J. Dent. **14**, 189 (2020)
17. A.G. Stevens, C.R. Oliver, M. Kirchmever, et al., Conformal robotic stereolithography. 3D Print Addit. Manuf. **3**(4), 226–235 (2016)
18. J. Bustillos, D. Montero-Zambrano, A. Loganathan, B. Boesl, A. Agarwal, Stereolithography-based 3D printed photosensitive polymer/boron nitride nanoplatelets composites. Polym. Compos. **40**, 379–388 (2019)
19. P. Makvandi, C. Esposito Corcione, F. Paladini, A.L. Gallo, F. Montagna, R. Jamaledin, M. Pollini, A. Maffezzoli, Antimicrobial modified hydroxyapatite composite dental bite by stereolithography. Polym. Adv. Technol. **29**, 364–371 (2018)
20. R. França, J. Winkler, H.H. Hsu, M. Rahimnejad, Z. Abdali, 3D printing—Additive manufacturing of dental biomaterials. Dent. Biomater. **2**, 421 (2018)
21. D. Zheng, Research status report on layered entity manufacturing technology. Times Agric. Mach. **9**, 90–91 (2013)
22. E.O. Olakanmi, R.F. Cochrane, K.W. Dalgarno, A review on selective laser sintering/melting (SLS/SLM) of aluminium alloy powders: Processing, microstructure, and properties. Prog. Mater. Sci. **74**, 401–477 (2015)
23. J.-P. Kruth, B. Vandenbroucke, J. Van Vaerenbergh, I. Naert, Rapid manufacturing of dental prostheses by means of selective laser sintering/melting. Proc. AFPR S. **4**, 176–186 (2005)
24. W. Zhu, C. Yan, Y. Shi, S. Wen, J. Liu, Y. Shi, Investigation into mechanical and microstructural properties of polypropylene manufactured by selective laser sintering in comparison with injection molding counterparts. Mater. Des. **82**, 37–45 (2015)
25. Y. Okazaki, A. Ishino, Microstructures and mechanical properties of laser-sintered commercially pure Ti and Ti-6Al-4V alloy for dental applications. Materials (Basel) **13**, 609 (2020)
26. O. Alageel, M. Abdallah, A. Alsheghri, J. Song, E. Caron, F. Tamimi, Removable partial denture alloys processed by laser-sintering technique. J. Biomed. Mater. Res. Part B Appl. Biomater. **106**, 1174–1185 (2018)
27. M.A. Alhnan, T.C. Okwuosa, M. Sadia, K.-W. Wan, W. Ahmed, B. Arafat, Emergence of 3D printed dosage forms: Opportunities and challenges. Pharm. Res. **33**, 1817–1832 (2016)
28. Y.F. Pan, S. Yizhong, G. Dongdong, X. Chengting, Development status of selective laser sintering technology. Tools Technol. **38**(6), 3–7 (2004)
29. P. Jain, A.M. Kuthe, Feasibility study of manufacturing using rapid prototyping: FDM approach. Procedia Eng. **63**, 4–11 (2013)
30. E. Cuan-Urquizo, E. Barocio, V. Tejada-Ortigoza, R.B. Pipes, C.A. Rodriguez, A. Roman-Flores, Characterization of the mechanical properties of FFF structures and materials: A review on the experimental, computational and theoretical approaches. Materials (Basel) **12**, 895 (2019)
31. T. Yin, W. Li, Characteristics and development direction of rapid prototyping technology, in *Modern Industrial Economy and Information Technology*, vol. 23, (2014)
32. Z. Ping, J. Hua, Z. Zhiting, The principle and process of fused deposition rapid prototyping technology. Mach. Manufacturing Res. **23**(5), 17–18 (2003)

33. C. Xuefei, S. Pengjie, Q. Yongjie, Z. Ping, Research and application of 3D printing bone tissue engineering scaffold. Chinese Tissue Eng. Res. **25**, 4076–4080 (2015)
34. Z. Difeng, Y. Jianming, H. Dazhi, C. Jinsong, T. Yang, The development and research status of 3DP three-dimensional printing technology. Manufacturing Technol. Mach. Tool **3**, 38–43 (2017)
35. J. Suwanprateeb, R. Chumnanklanc, Three-dimensional printing of porous polyethylene structure using water-based binders. J. Biomed. Mater. Res. B Appl. Biomater. **78B**(1), 138–145 (2006)
36. N.W. Choi, M. Cabodi, B. Held, J.P. Gleghorn, L.J. Bonassar, A.D. Stroock, Microfluidic scaffolds for tissue engineering. Nat. Mater. **6**(11), 908–915 (2007)
37. J. Jang, J.Y. Park, G. Gao, D.W. Cho, Biomaterials-based 3D cell printing for next-generation therapeutics and diagnostics. Biomaterials **156**, 88–106 (2018)
38. H. Lipson, M. Kurman, *Fabricated: The New World of 3D Printing* (Wiley, Indianapolis, 2013)
39. H.-N. Mai, K.-B. Lee, D.-H. Lee, Fit of interim crowns fabricated using photopolymer-jetting 3D printing. J. Prosthet. Dent. **118**, 208–215 (2017)
40. A. Dawood, B.M. Marti, V. Sauret-Jackson, A. Darwood, 3D printing in dentistry. Br. Dent. J. **219**, 521–529 (2015)
41. A. Kessler, R. Hickel, M. Reymus, 3D printing in dentistry—State of the art. Oper. Dent. **45**, 30–40 (2020)
42. A. Hazeveld, J.J.R. Huddleston Slater, Y. Ren, Accuracy and reproducibility of dental replica models reconstructed by different rapid prototyping techniques. Am. J. Orthod. Dentofac. Orthop. **145**, 108–115 (2014). https://doi.org/10.1016/j.ajodo.2013.05.011
43. J.W. Stansbury, M.J. Idacavage, 3D printing with polymers: Challenges among expanding options and opportunities. Dent. Mater. **32**, 54–64 (2016). https://doi.org/10.1016/j.dental.2015.09.018
44. A. Mostafaei, E.L. Stevens, J.J. Ference, D.E. Schmidt, M. Chmielus, Binder jet printing of partial denture metal. Mater. Sci. Technol. **2017**, 289–291 (2017). https://doi.org/10.7449/2017/MST
45. W.S. Tan, C.K. Chua, T.H. Chong, A.G. Fane, A. Jia, 3D printing by selective laser sintering of polypropylene feed channel spacers for spiral wound membrane modules for the water industry. Virtual Phys. Prototyp. **11**(3), 151–158 (2016)
46. X. Yang, Z. Lu, H. Wu, W. Li, L. Zheng, J. Zhao, Collagen-alginate as bioink for three-dimensional (3D) cell printing based cartilage tissue engineering. Mater. Sci. Eng. C Mater. Biol. Appl. **83**, 195–201 (2018)
47. R. Shirazi, A. Shirazi-Adl, M. Hurtig, Role of cartilage collagen fibrils networks in knee joint biomechanics under compression. Biomechanics **41**(16), 3340–3348 (2008)
48. R. Biodur, 108 alloy (nickel-free high-nitrogen austenitic stainless steel alloy). Alloy. Dig. **8** (1999)
49. M. Niinomi, Recent research and development in titanium alloys for biomedical applications and healthcare goods. Sci. Technol. Adv. Mater. **4**, 445–454 (2016)
50. M. Niinomi, Fatigue performance and cyto-toxicity of low rigidity titanium alloy, Ti–29Nb–13Ta–4.6Zr. Biomaterials **24**(16), 2673–2683 (2003)
51. R. Yibin, Y. Ke, A new Ni-free drill-based alloy for vascular stent. Rare Metal Mater. Eng. **43**(1), 101–104 (2014)
52. L.L. Hench, R.J. Splinter, W.C. Allen, T.K. Greenlee, Bonding mechanisms at the interface of ceramic prosthetic materials. J. Biomed. Mater. Res. **5**(6), 117–141 (1971)
53. I.D. Xynos, A.J. Edgar, L.D.K. Buttery, L.L. Hench, J.M. Polak, Gene-expression profiling of human osteoblasts following treatment with the ionic products of Bioglass 45S5 dissolution. J. Biomed. Mater. Res. **55**(2), 151–157 (2001)
54. L.L. Hench, Bioceramics: From concept to clinic. J. Am. Ceram. Soc. **74**(7), 1487–1510 (1991)
55. J.K. Sherwood, S.L. Riley, R. Palazzolo, et al., A three-dimensional osteochondral composite scaffold for articular cartilage repair. Biomaterials **23**(24), 4739–4751 (2002)

56. K.J.L. Burg, S. Porter, J.F. Kellam, Biomaterial developments for bone tissue engineering. Biomaterials **21**(23), 2347–2359 (2000)
57. G.A. Roth, C. Johnson, A. Abajobir, et al., Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. J. Am. Coll. Cardiol. **70**(1), 1–25 (2017). BioMed Research International 11
58. D. Qiushuang, Analysis of combined medication regimen in the treatment of cardiovascular disease. Chinese Med. Guide **15**(12), 140–141 (2017)
59. G. Wu, Pay attention to the perioperative management of elderly general surgery patients. Chinese J. Pract. Surg. **29**(2), 109–111 (2009)
60. S. Qian, X. Jianming, The historical development and future prospects of coronary stents. Adv. Cardiovasc. Dis. **37**(1), 23–27 (2016)
61. C. Flege, F. Vogt, S. Höges, et al., Development and characterization of a coronary polylactic acid stent prototype generated by selective laser melting. J. Mater. Sci. Mater. Med. **24**(1), 241–255 (2013)
62. W.H. Kaesemeyer, K.G. Sprankle, J.N. Kremsky, W. Lau, M.N. Helmus, G.S. Ghatnekar, Bioresorbable polystatin fourth-generation stents. Coron. Artery Dis. **24**(6), 516–521 (2013)
63. S.A. Park, S.J. Lee, K.S. Lim, et al., *In vivo* evaluation and characterization of a bio-absorbable drug-coated stent fabricated using a 3D-printing system. Mater. Lett. **141**, 355–358 (2015)
64. R. van Lith, E. Baker, H. Ware, et al., 3D-printing strong high resolution antioxidant bioresorbable vascular stents. Adv. Mater. Technol. **1**(9) (2016)
65. H. Fengyang, S. Jian, S. Lu, G. Gao, Development of artificial heart valves. Med. Equip. **30**(13), 186–193 (2017)
66. W.Z. Marx, S. Duanchao, X. Wang, Z. Long, Analysis of factors affecting the prognosis of patients with severe heart valve disease after artificial heart valve replacement. China Med. Herald **14**(26), 64–67 (2017)
67. S.P. Houstrop, A. Kadner, S. Melnitchouk, et al., Tissue engineering of functional trileaflet heart valves from human marrow stromal cells. Circulation **106**(12, Supplement 1), 143–150 (2002)
68. S.P. Hoerstrup, R. Sodian, S. Daebritz, et al., Functional living trileaflet heart valve grown in vitro. Circulation **102**(Supplement 3), 44–49 (2000)
69. P. Rider, Ž.P. Kačarević, S. Alkildani, S. Retnasingh, R. Schnettler, M. Barbeck, Additive manufacturing for guided bone regeneration: A perspective for alveolar ridge augmentation. (2018). https://doi.org/10.3390/ijms19113308
70. M. Salah, L. Tayebi, K. Moharamzadeh, F.B. Naini, Three-dimensional bioprinting and bone tissue engineering: Technical innovations and potential applications in maxillofacial reconstructive surgery. Maxillofac. Plast. Reconstr. Surg. **42** (2020). https://doi.org/10.1186/s40902-020-00263-6
71. D. Nesic, B.M. Schaefer, Y. Sun, N. Saulacic, 3D printing approach in dentistry: The future for personalized oral soft tissue regeneration. J. Clin. Med. **9**(7), 2238 (2020)
72. M.T. Houdek, C.C. Wyles, J.R. Labott, et al., Durability of hemiarthroplasty for pathologic proximal femur fractures [J]. J. Arthroplast. **32**(12), 3607–3610 (2017)
73. G. Tang, L. Liu, Z. Xiao, et al., CircRAB3IP upregulates twist family BHLH transcription factor (TWIST1) to promote osteosarcoma progression by sponging miR-580-3p [J]. Bioengineered. **12**(1), 3385–3397 (2021)
74. J. Ding, J. Cao, Z. Chen, et al., The role of long intergenic noncoding RNA 00511 in malignant tumors: A meta-analysis, database validation and review [J]. Bioengineered. **11**(1), 812–823 (2020)
75. S.S. Low, Y. Pan, D. Ji, et al., Smartphone-based portable electrochemical biosensing system for detection of circulating microRNA-21 in saliva as a proof-of-concept [J]. Sens. Actuators B Chem. **308**, 127718 (2020)
76. L.R. Menendez, E.R. Ahlmann, C. Kermani, et al., Endoprosthetic reconstruction for neoplasms of the proximal femur [J]. Clin. Orthop. Relat. Res. **450**, 46–51 (2006)

77. W. Sun, P. Zan, X. Ma, et al., Surgical resection and reconstructive techniques using autologous femoral head bone-grafting in treating partial acetabular defects arising from primary pelvic malignant tumors [J]. BMC Cancer **19**(1), 969 (2019)
78. C.W. How, Y.S. Ong, S.S. Low, et al., *How Far Have we Explored Fungi to Fight Cancer?; Proceedings of the Seminars in Cancer Biology, F, [C]* (Academic Press, Elsevier, 2021)
79. W.Y. Chia, H. Kok, K.W. Chew, et al., Can algae contribute to the war with Covid-19 [J]. Bioengineered. **12**(1), 1226–1237 (2021)
80. Q. Pan, K. Wu, J. Tan, et al., Anti-neoplastic characteristics and potential targets of calycosin against bisphenol A-related osteosarcoma: Bioinformatics analysis [J]. Bioengineered. **12**(1), 4278–4288 (2021)
81. A. Sallent, M. Vicente, M.M. Reverté, et al., How 3D patient-specific instruments improve accuracy of pelvic bone tumour resection in a cadaveric study [J]. Bone Joint Res. **6**(10), 577–583 (2017)

Fuzzy-Based Hierarchical Routing Protocol for Wireless Sensor Networks

G. H. Raghunandan, N. Keerthi Kumar, K. Neeraj, Ramya U. Kashyap, S. V. Vishal, and Raveendra Gudodagi

1 Introduction

Wireless sensor networks refer to networks that are geographically dispersed and have dedicated sensors that monitor and record visual changes and environmental conditions. The data which is sensed is then transmitted to the central area, which is usually the base or control station. Even though sensors are used for specialized services, plenty of research has taken place in the area of WSN recently. A sensor network consists of numerous nonlinearly arranged self-governing sensor devices. These devices direct the sensed instruction to the node and make the optimum usage of the obtained information. Initially, due to the increased cost and limitations of wireless sensors, the use was limited. The recent advancements and reduced cost of wireless sensors have increased its field of application. The sensor network consists of various detection channels which are nothing but lightweight, portable, and small-sized sensors. Each sensor node has a transducer that senses a few physical parameters and produces signals based on these parameters. It also consists of a small computer which computes and stores the observed data and a transceiver that receives commands from the control station. The battery source is responsible

G. H. Raghunandan (✉) · K. Neeraj · R. U. Kashyap · S. V. Vishal
Department of Electronics and Telecommunication Engineering, BMS Institute of Technology & Management, Bangalore, India
e-mail: raghunandangh@bmsit.in

N. Keerthi Kumar
Department of Mechanical Engineering, BMS Institute of Technology & Management, Bangalore, India
e-mail: keerthikumarn@bmsit.in

R. Gudodagi
School of Electronics and Communication Engineering, Reva University, Bangalore, India
e-mail: raveendra@revainstitution.org

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_13

for providing the necessary power to operate the sensors properly. The sensory nodes in WSN are used with an onboard processor that controls and monitors the environment. They are connected to a control station that serves as a processing unit in the WSN system. The control station shares the data while connected to the Internet. There are different types of WSNs broadly categorized into the following:

1. *Terrestrial WSNs:* Terrestrial applications generally require the deployment of a large number of nodes (usually hundreds to thousands). These sensor nodes are capable of communicating with base stations in an efficient manner. The nodes are deployed either in a structured (preplanned) or unstructured (ad hoc) manner. The terrestrial applications of WSNs include precision agriculture, disaster management (flood, earthquake, etc.), surveillance and maintenance of machines, area monitoring, habitat monitoring, healthcare monitoring, and so on.
2. *Underground WSNs:* These networks are highly expensive as compared to terrestrial WSNs in terms of arrangement, maintenance, equipment, and planning. Several sensor nodes are buried in the ground to inspect the conditions [10]. These sensor nodes together form an underground WSN. To obtain or extract the required information from the sensor nodes, additional gateway nodes are deployed above the ground. Like any other WSN, the sensor nodes in underground WSNs rely on a limited amount of power supply. In addition, there are high levels of signal loss or attenuation due to underground environment. This makes wireless communication in underground a challenge.
3. *Underwater WSNs:* Approximately three-fourths of the Earth's surface are covered with water, and only about 5% of the ocean is discovered till date. Thus, WSNs have huge scope in underwater applications for habitat monitoring, ecosystem monitoring, and suchlike. Underwater WSNs are a network of dedicated sensors deployed randomly underwater. They consist of several sensor nodes, wireless acoustic links, surface buoy, and broadband radio connection to the Internet and vehicles. Underwater WSN deployment faces a lot of challenges due to the undiscovered environment. Hence, improving energy efficiency involves development of underwater networking techniques and underwater communication as well.
4. *Multimedia WSNs:* These WSNs include nodes that are installed with cameras, microphones, etc. Such WSNs facilitate tracking and monitoring of events. This is done in the form of audio, video, and imagining, which are all forms of multimedia. Multimedia require or take up a large amount of space. Hence, high bandwidth is required to deliver such content effectively. Multimedia WSNs facilitate data retrieval or data compression and correlation. However, multimedia WSNs also face certain challenges that include the requirement of updated data processing and compressing techniques, higher communication bandwidth, and improved energy efficiency.
5. *Mobile WSNs (MWSNs):* In this type of network, the sensory nodes are mobile. They are more flexible than stationary sensor networks as they can cope up in any environment even though they are subject to constantly changing environment and updated technology. The nodes consist of a radio transceiver and a small

battery-powered controller. The mobile WSNs provide several advantages over static WSNs as they provide better coverage, superior channel capacity, higher energy efficiency, etc.

The sensor nodes' deployment in a WSN involves the selection of a routing protocol. The routing protocol helps in selection of an appropriate route from source to destination for transferring the desired data. There are three major classes of routing protocols [9]:

1. *Flat-based routing protocols:* These protocols are those that do not operate under a predefined network structure and/or perimeter. They allow the delivery of packets between routers by any means available without regard to network composition, hierarchy, distribution, and configuration. This protocol is used in networks where individual nodes normally gather and send the routing information to all the adjacent router nodes. All active nodes implementing this protocol play an equivalent role in the entire routing mechanism.
2. *Hierarchical routing protocols:* Cluster-based sequencing protocols divide networks into multiple groups by defining the sequence of sensor nodes. This is a common way of communicating data to improve power consumption. This is done with the help of cluster heads, which aggregate the data and transmit it to the control station. Transmission within a cluster is coordinated by the head of each cluster which is also responsible for transferring the data between clusters and base stations. Hierarchical routing can be divided into chain-based, tree-based, grid-based, and area-based routing, which can be further classified as single-hop and multiple-hop transmission. Such protocols decrease the energy consumption by the sensor nodes and provide high scope for scalability.
3. *Location-based routing protocols:* It is called as geographic- or position-based routing protocols. Such protocols are used in WSNs in which communication between sensor nodes requires the data about the position of the nodes. A source node adds the destination node to its routing table when a packet needs to be sent. Geographically, closest nodes are chosen as the intermediate nodes, and the path for packet routing is opted accordingly. Location-based routing thus does not require the establishment or maintenance of routes.

Apart from routing protocols, there are other aspects that help in the development of an efficient wireless sensor network. Few of these aspects include efficient energy consumption, complexity, time synchronization, data aggregation, delay, scalability, robustness, sensor location, and data transmission models.

The functionality of generally distributed systems and WSNs depends on time, as these systems have to monitor and interact with physical conditions. The sensors use oscillator-driven clocks to measure the time. Due to the phase fluctuations and oscillator flow rates, local time readings of these nodes will begin to vary. This leads to loss of time synchronization. This problem is quite common in distributed systems. Thus, overcoming this problem, by opting for time synchronization protocols, would yield an efficient wireless sensor network.

Data aggregation is another aspect that plays a vital role in the development of an efficient WSN, by ensuring optimal energy consumption [11]. It is considered as one of the most fundamental processing procedures. The sensor nodes in the WSN tend to forward raw data which results in redundant or correlated data, leading to several drawbacks such as network congestion, wastage of network capacity, unnecessary energy consumption, and so on and so forth. Thus, data aggregation, which is a key mechanism to reduce energy consumption in WSNs, is an integral aspect.

It is no hidden fact that WSNs have a wide range of applications. It is not possible to handle a wide range of applications with only single realization of WSNs [8]. However, some common traits are observed, especially in relation to the features and methods required for such systems. Thus, Sect. 2 will discuss about the design challenges faced during the development of a WSN.

Before we come up with an efficient algorithm and protocol to develop a WSN, it is important to learn about the existing WSN systems and the algorithms used in those systems. Hence, Section II will discuss about the existing systems and Sect. 3 highlights about the proposed model. Section IV provides a detailed description about the system model. Section V demonstrates the implementation and result. The conclusion of the paper and brief discussion about the future work are provided in Sect. 4.

2 Existing Systems

Data transmission in a WSN is carried out by the implementation of centrality and gateway-based clustering algorithm (CGBCA). The network sensor nodes are grouped into clusters, and a cluster head is selected for each of these clusters. The cluster head selection is based on the concept of centrality, as opposed to random head selection in typical LEACH protocol. In addition, the proposed system model deploys intermediate gateway nodes that decrease the load on cluster head nodes. This implementation shows considerable improvement over existing models. Authors [1, 2] discussed that sensor networks have received a plenty of considerations recently and received new challenges on correlating with old data routing in wired networks. A brief survey of data routing in sensor networks is done. The most common issue that WSNs face is energy of sensors. When a sensor's energy source is depleted, it loses its ability to function. Even though many routing techniques appear advantageous, they have numerous complications that must be resolved in sensor networks, such as achieving sensor node density, synchronizing time and location, reducing delay time, and improving energy efficiency. Additionally, research challenges include achieving redundancy and effectively utilizing fault tolerance. The energy efficiency of the WSNs can be improved with the help of concepts such as multi-hop network and particle swarm optimization (PSO) technique. The deployment of sensor nodes and random cluster formation may lead to certain clusters having a greater number of nodes than the others. This leads to higher energy consumption and high load on cluster heads. This can be avoided by making use of the PSO technique. The situation where a cluster head node is burdened due to heavy incoming data and data transmission to base station leads to death of the cluster head. This situation is avoided by electing vice cluster head that assists the cluster head. The multi-hop network technique is also incorporated for added energy efficiency of the whole network [3].

The proposed system model makes changes to the existing LEACH algorithm to get a modified LEACH (MOD-LEACH) algorithm. An optimal number of cluster heads have been selected in order to upsurge the network performance. The energy parameter has been given utmost importance during the selection of cluster heads. An arbitrary number amid 0 and 1 is selected, and the sensor nodes having energy greater than that number are considered for the position of cluster head. This helps in the selection of optimal cluster heads for a given network [4].

A similar model as the previous one is implemented with few improvisations. The only difference is that the cluster head is selected depending upon how far it is from the base station. The distance formula has been used for computation [12]. The ordinary nodes, that is, the nodes having energy below the random number selected between 0 and 1, will select a cluster head which has least distance to the base station and sends a message to the cluster head informing that it wants to be a member of its cluster [5].

An approach to enhance the lifetime of the sensors is introduced by adopting the S-LEACH protocol. About ten clusters are formed and cluster heads are nominated using a specific formula. Also, the distance between the nodes and the base station should be optimal for cluster head nomination. Laterally, a secondary cluster head (SCH) is also selected as soon as the present CH is dead. This ensures that the data transmission is not affected even if some nodes are dead [6].

The existing models provide different solutions to improve the performance and energy efficiency of the WSNs. However, these models exhibit certain disadvantages. The difference in energy between the sensor nodes other than the cluster head is not given any importance. Therefore, this leads to the death of the network at a faster rate [4, 5]. The setup time for selecting a secondary cluster head (SCH) is more for every round in S-LEACH approach [6]. Additionally, these existing models consider only a single parameter to enhance the energy efficiency. These systems also provide scope for implementation of mobile base station. The majority of existing protocols are focused on extending the network's life and proficient usage of battery power of sensor node. Additionally, the requirement for rapid sensing and low energy consumption resulted in the development of clustering algorithms which is the most accepted and extensively helpful in many applications. These algorithms assume that nodes make rapid decisions solely dependent on probability or other local information with the preferred quality of the final output of cluster typically treated as a secondary parameter. Another significant characteristic of majority algorithms is the requirement for periodic reelection of CHs (role rotation) across the network's nodes. Clustering algorithms based on more traditional selection criteria such as connectivity, node proximity, and distance have also been developed. They are still in use, resulting in an output that is likely more balanced. However, it is challenging to keep these algorithms and time complexity minimum in clustering algorithms.

3 Proposed Model

In the proposed model which is as shown in Fig. 1, we make use of fuzzy logic (Mamdani) for round-based selection of cluster heads and data transmission in WSNs as well, which is an improvement over the typical LEACH algorithm. The network is divided into clusters which is similar to the LEACH protocol. The main difference in the proposed model is with respect to cluster head selection, which is done based on the criticality or critical degree obtained from the fuzzy logic, and a new head is selected for the clusters after each round based on this critical degree.

Fuzzy logic is a type of multivalue concept where output or truth values can be any real number between 0 and 1 [7]. It is used to build a control system by integrating a set of linguistic control rules known as membership functions, obtained from human operators. The best-known fuzzy system is the Mamdani rule-based system. In a Mamdani system, certain if-then criteria are utilized in order to generate the output of each rule, known as a fuzzy set.

The fuzzy logic system designed accepts three inputs and in turn a single output is generated. The three inputs include energy levels of the node, proximity of the nodes to the control station, and the total number of members in a cluster. Each of the inputs has corresponding membership functions such as high, low, medium, etc. Based on these membership functions, 27 different rules (3 inputs where each input has 3 membership functions, that is, 3ˆ3) are defined. Further, the output of the fuzzy logic system is defined as critical degree or criticality, which is dependent upon these 27 defined rules. The single output generated has nine membership functions that defines the criticality of the nodes (Fig. 2).

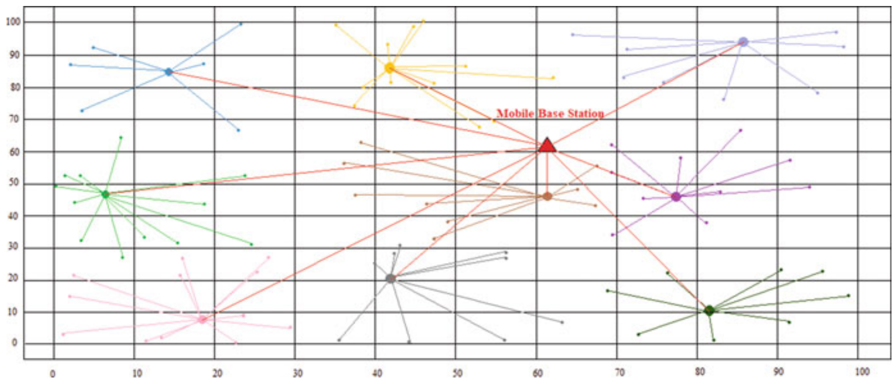


Fig. 1 Proposed system model

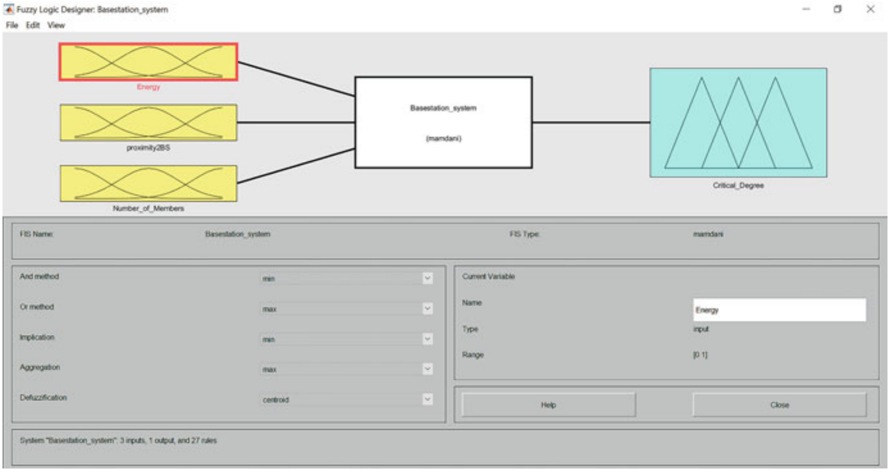


Fig. 2 Fuzzy logic design

4 System Model

The system model involves designing of a wireless sensor network based on certain criteria like total number of dead nodes, residual energy, and energy variation of nodes after each round. By doing so, it is intended to develop a design for wireless sensor networks that are far more efficient in terms of overall network lifetime, as paralleled to the other existing systems. Cluster formation generates a two-tier hierarchical structure, with CH nodes at the top and cluster member nodes at the bottom. The sensor nodes transmit data to the corresponding CH nodes periodically. The CH nodes aggregate the data (thereby reducing the total number of packets relayed) and send it directly to the BS or via internal communication with different CH nodes. On the other hand, CH nodes consume more energy by default because they are constantly transmitting data over greater distances than ordinary (member) nodes. Additionally, the structure formed by the sensor nodes, sink, and BS can be replicated indefinitely, allowing multiple layers of the hierarchical WSN if desired to have a multilevel cluster hierarchy. The multilevel hierarchy is necessary because if the CHs are far from the base station, their energy depletes soon. To overcome this problem, an energy-efficient routing scheme is discussed, which will decrease the energy depletion of CHs for data sending. Cluster formation occurs, which is followed by the selection of a CH. Optimal clustering aims to manipulate the cluster size as regards to the features and data sending or to minimize and equalize the network's energy consumption due to factors such as residual energy, data correlation, and relay traffic. Cluster creation is based on minimizing energy consumption in the proposed algorithm.

The complete system modeling can be divided into four fundamental functions. They are as follows:

A. Cluster Formation

First, the entire area of the wireless sensor network should be divided into equal areas. This results in groups of equal areas known as clusters. Each of them contains a random number of nodes/members.

B. Cluster Head Selection

As mentioned earlier, the hierarchical routing protocols are commonly used protocols for energy consumption optimization. LEACH, being one such protocol, demands the selection of a cluster head for each equal area cluster. The typical approach involves selection of cluster head based on centrality. However, the proposed model selects the cluster heads based on centrality as well as the criticality of the nodes. In addition, the system model consists of a mobile base station that minimally moves toward the cluster head with high critical degree. This novel approach reduces the burden on the cluster heads to some extent and ensures higher lifetime of the entire network.

C. Locating Dead Nodes

The sensor node is declared a dead node as soon as its power level falls below a certain limit. The energy level of the sensor nodes is monitored continuously after each cycle. Any sensor node with a lower energy level than the threshold remains connected to the network but is incapable of sensing or transmitting data. Once such a node loses all its energy, it is located and eliminated from the network.

D. Comparing the Model with LEACH

Finally, the proposed system model is compared with the model that employs typical LEACH algorithm. Parameters such as residual energy, alive nodes, and energy variance are compared in order to determine the better performing model.

5 Implementation and Results

MATLAB version R2021a is used to implement the proposed model. The following assumptions have been made for the proposed network model:

(a) Random deployment of 100 sensor nodes in an area of 100×100 sq. units.
(b) The sensory nodes are stationary while the base or control station is mobile.
(c) It is assumed that all the nodes have approximately same energy level (1 Joule each).
(d) The network has been divided into clusters using centrality approach.
(e) All the clusters have a head which is elected depending on the criticality level determined by the fuzzy logic system.
(f) A new head for each cluster will be selected after each round, based on the amount of residual energy.

The network is simulated for 7000 rounds, and after each round, the energy level of all the sensor nodes is determined. The node with maximum energy remaining

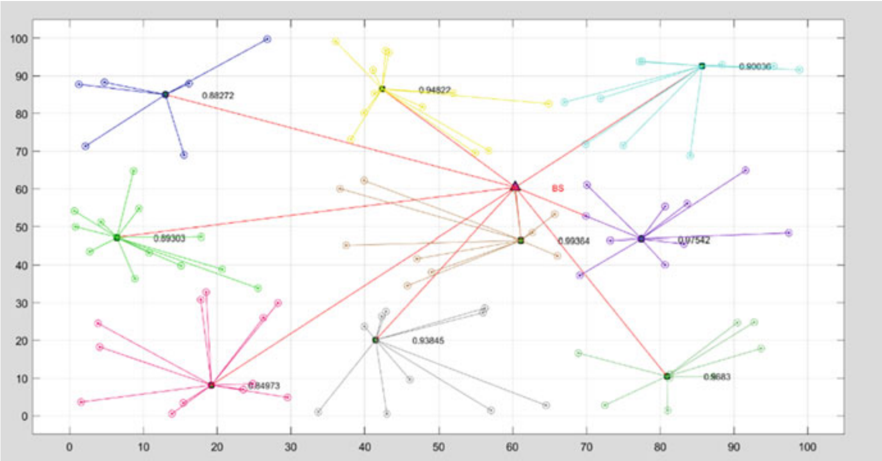


Fig. 3 Initial topology of the network

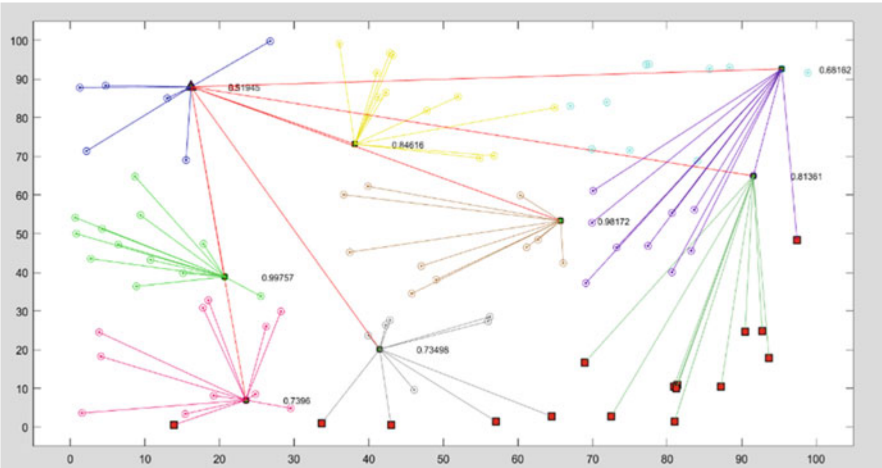


Fig. 4 Network topology after “n” rounds

and which is equidistant from other nodes, along with least critical degree, will be voted as the cluster head for each round, respectively. Figure 3 depicts the initial topology of the WSN which consists of 100 nodes that are randomly deployed. This network is simulated until the energy of most of the sensor nodes becomes 0 J. Figure 4 depicts the network topology after certain number of rounds. It can be observed that the levels of energy of few of the sensory nodes deplete after certain interval of time. As mentioned earlier, they remain connected to the network until their energy is completely drained.

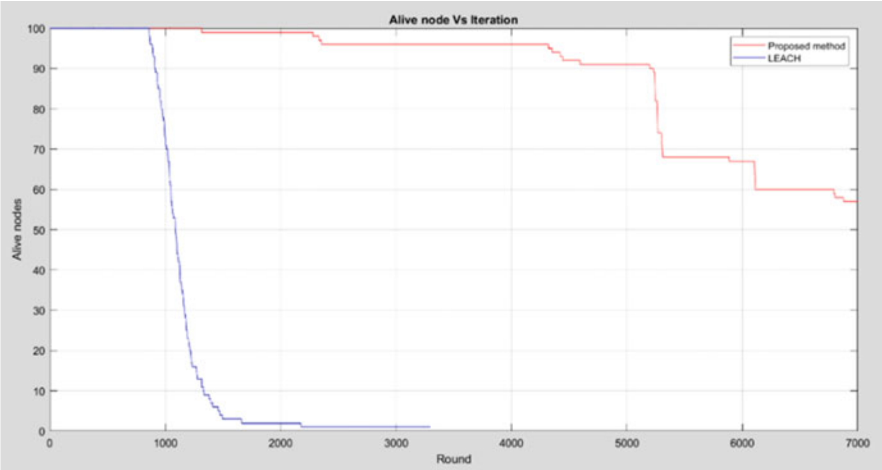


Fig. 5 Comparison of alive nodes in LEACH and proposed model after 7000 rounds

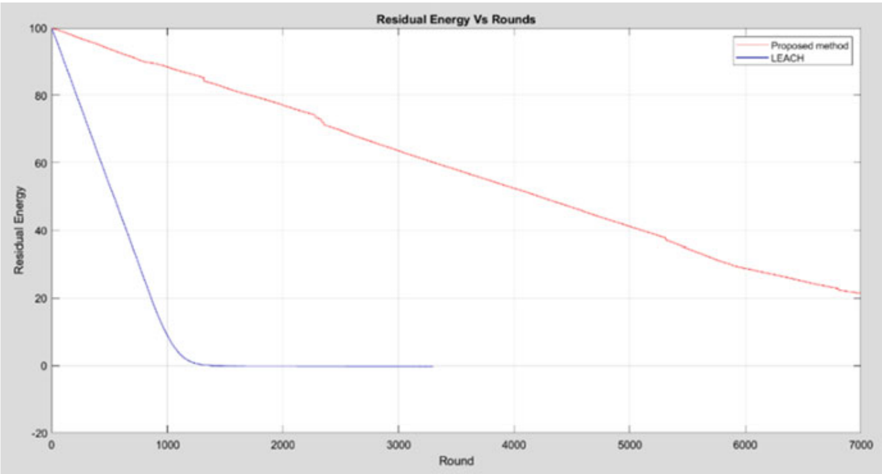


Fig. 6 Comparison of residual energy in LEACH and proposed model after 7000 rounds

Figure 5 contains a plot of total count of alive nodes vs the total number of iterations. After the completion of simulation of 7000 rounds, it is observed that the alive nodes in our proposed model are equal to 57. On the other hand, the typical LEACH protocol halts the simulation at approximately 3500 rounds with total number of alive nodes being barely 2. Thus, at the end of the simulation, the overall number of alive nodes is higher in our proposed model.

Figure 6 depicts a plot of the nodes' residual energy, that is, energy remaining in each node after every round. The residual energy in LEACH is observed to drop drastically, whereas the residual energy of the nodes in our proposed

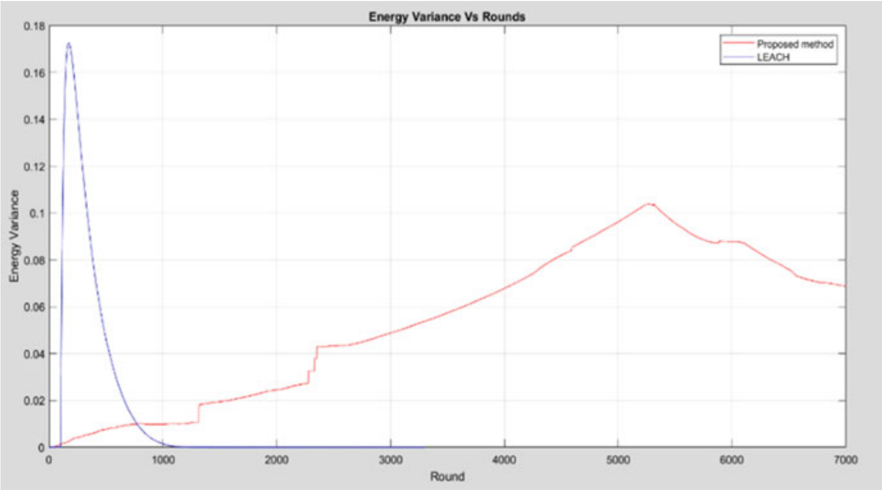


Fig. 7 Comparison of energy variance in LEACH and proposed model after 7000 rounds

model decreases gradually that ensures optimal energy consumption throughout the lifetime of the WSN.

Figure 7 displays a plot of energy variance with respect to the total number of rounds. This parameter helps in analyzing the variance of energy level in the nodes. It determines the amount of deviation of energy levels from the mean value of energy. As observed, variance of energy is smoother or approximately constant in our proposed model, as opposed to the typical LEACH model.

6 Conclusion

Due to a drastic increase in the use of WSNs in a wide range of applications, there is a need for increasing the efficiency of the WSNs. In order to increase the lifetime of WSN, routing protocols play an important role. In this paper, fuzzy logic-based routing protocol with mobile base station is discussed. As discussed in the previous section, it is noted that our proposed protocol works better than other existing LEACH protocols in terms of total count of live nodes, residual energy of the nodes, and energy variance parameters. Thus, the proposed system proves to enhance the energy efficiency and lifetime of the whole network.

Further, the improvisation of energy efficiency of the wireless sensor network provides high scope for research and development due to the limited resources available in the network. Additionally, the performance of the network can be improved by considering various design challenges and services. The deployment of gateway nodes in the network could also possibly contribute to the efficiency of the network.

References

1. G.H. Raghunandan, A.S. Rani, S.Y. Nanditha, G. Swathi, Hierarchical agglomerative clustering-based routing algorithm for overall efficiency of wireless sensor network, 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (2017)
2. G. Xiong, L. Hong, Y. Guangyou, Improving energy efficiency by optimizing relay nodes deployment in wireless sensor networks, 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN) (2017)
3. A. Singh, S. Rathkanthiwar, S. Kakde, LEACH based-energy efficient routing protocol for wireless sensor networks, 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) (2016)
4. Z. Zhao, G. Li, M. Xu, An improved algorithm based on LEACH routing protocol, 2019 IEEE 19th International Conference on Communication Technology (ICCT) (2019)
5. A.O. Abu Salem, N. Shudifat, Enhanced LEACH protocol for increasing a lifetime of WSNs. Pers. Ubiquit. Comput. **23**, 901–907 (2019)
6. S. Nasr, M. Quwaider, LEACH protocol enhancement for increasing WSN lifetime, 2020 11th International Conference on Information and Communication Systems (ICICS) (2020)
7. X. Liu, K. Mei, S. Yu, Clustering algorithm in wireless sensor networks based on differential evolution algorithm, 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (2020)
8. Y. Pinar, A. Zuhair, A. Hamad, A. Resit, K. Shiva, A. Omar, Wireless sensor networks (WSNs), 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT) (2016)
9. M. Rajput, S.K. Sharma, P. Khatri, Energy efficient hybrid multilevel multihop protocol for wide area WSN, 2019 International Conference on Intelligent Computing and Control Systems (ICCS) (2019)
10. L. Mao, Y. Zhang, An energy-efficient LEACH algorithm for wireless sensor networks, 2017 36th Chinese Control Conference (CCC) (2017)
11. N.G. Palan, B.V. Barbadekar, S. Patil, Low energy adaptive clustering hierarchy (LEACH) protocol: A retrospective analysis, 2017 International Conference on Inventive Systems and Control (ICISC) (2017)
12. J. Grover, S. Shikha, M. Sharma, Location based protocols in Wireless Sensor Network — A review, Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT) (2014)

A Review of AI-Based Diagnosis of Multiple Thoracic Diseases in Chest Radiography

Roshan Shetty, Prasad Narasimha Sarappadi, K. M. Sudarshan, and Raveendra Gudodagi

1 Introduction

Diseases seen in the thoracic cavity of the human body are called thoracic diseases. Millions of people suffer from these thoracic diseases worldwide. The chest X-ray is a recommended initial procedure for thoracic diseases in hospitals. Figure 1 shows chest X-ray images of some common thoracic diseases. Interpretation of X-rays and radiologists' decision plays a vital role during the treatment of such diseases. Medical images such as chest X-rays provide valuable information, and AI-based diagnosis of medical images is very much essential due to the shortage of radiologists. So artificial intelligence is alternative for solving such diagnosis problems and improved health care for the public. The AI-based computer-aided diagnosis is dominated by deep learning (DL) algorithms, implemented with the help of the convolution neural network (CNN).

Deep learning falls under neural networks which are not new to the scientific world. Deep learning introduced the deep aspect of neural networks and is termed deep neural network (DNN). The advantage of DNN over neural networks is its ability to discover hidden features from input data. In deep learning, CNNs are specialized in the detection and classification of any images. Today, deep learning with CNN applications is increasing rapidly in computer-aided diagnosis of medical images. It will help the medical community in terms of decreased workload, quick diagnosis, and overall improvement in the health-care system [1].

R. Shetty (✉)
Alva's IET, Moodubidire, India

P. N. Sarappadi · K. M. Sudarshan · R. Gudodagi
REVA University, Bengaluru, India
e-mail: prasadsn@reva.edu.in; sudarshankm@reva.edu.in; raveendra.g@reva.edu.in

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_14

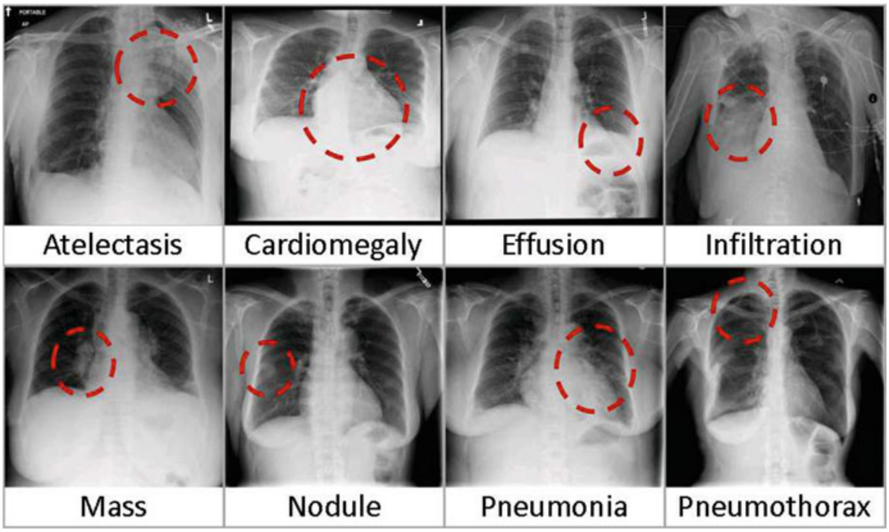


Fig. 1 Eight common thoracic diseases

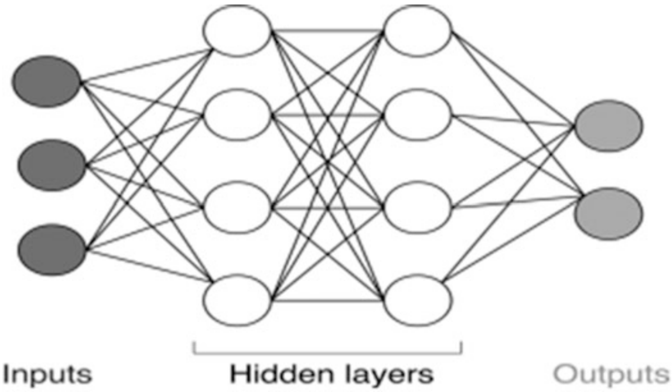


Fig. 2 Multilayer architecture of deep neural network [2]

Neural networks are inspired by the activity of the human brain, particularly the activities of neurons, which led to the development of a new area called artificial neural networks (ANNs). These neural networks and backpropagation training algorithms have a long history in the scientific world [2]. Although neural networks have been used for many years, in recent years, three factors triggered the learning of large neural networks. They are called deep neural networks (DNNs). Those factors are (1) powerful, massive parallel computing hardware called ultrafast graphics processing units (GPUs), (2) availability of large labeled datasets, and (3) advancements in training technologies and architecture. Figure 2 shows the "deep" aspect of deep learning which refers to the multilayer network architecture. Here, information flows from input to output nodes through many hidden layers of nodes.

A type of DNN called as CNN makes use of layers of kernels or filters to extract features from the entire image without anyone's supervision. In 2012, for the first time, human competitive results were reported with CNN such as AlexNet on a widely used computer vision benchmark [3], which had only 8 layers and 62 million parameters. This encouraged many researchers to use CNN on medical images. Important design factors to be considered for building a CNN-based chest X-ray thoracic disease interpretation model are (1) learning forms, (2) network architecture, and (3) training type.

In the next section, we will discuss briefly on these important design factors one by one. Section 3 explains different deep learning models built to tackle multiple thoracic disease detection and classification problem. Section 4 contains a discussion and future trends. Section 5 has concluding remarks.

2 Important Design Factors

2.1 Learning Forms

Supervised learning This is one of the leading learning forms where machines are fed with a lot of labeled input data and output is already known for it. It is predictive in nature with high accuracy. The model considers image labels as ground truths. For instance, in the learning system for classifying thoracic diseases, a particular X-ray image may be labeled as a "hernia." A deep model during supervised training continuously adjusts network parameters or weights in all layers. These adjustments are done by comparing deep model prediction with the ground truth (hernia).

Unsupervised learning The machine is fed with unlabeled input data and output is not mapped with input. It is an independent learning procedure, mainly used to detect hidden patterns without any ground truths.

Semi-supervised A hybrid approach makes use of large unlabeled data and less labeled data. Scarcity of labeled data made semi-supervised learning one of the problems of significant importance in modern-day data analysis. A semi-supervised model is always expected to do more than that in the label, for example, preparing a model to execute many tasks such as detection, classification, and also localization of pathology in the case of medical images.

Weakly supervised learning This is a different form of training method, in which learning happens with inaccurate labels. Supervised and weakly supervised learning are popular, among deep learning-based models on diagnosis of thoracic diseases.

2.2 Network Architecture

CNN: This is the most popular choice with impressive results compared to others in image classification tasks [4]. The CNN model consists of two major components:

1. Convolution filters and pooling layers: Convolution operation is performed by sliding the filter over the entire input image to produce feature maps. The ReLU activation function is used to add some nonlinearity to convolution results. So values in the final feature maps are not actually the sum. The pooling process reduces dimensions of image and number of parameters. This indirectly reduces training time of a model and overfitting.
2. Fully connected layer as classification part: It assigns a probability for the object on the image as what the algorithm predicts. Most of the time, softmax function is the common choice in fully connected layers for classification tasks, but the sigmoid function is popularly used in chest X-ray-based CAD systems. The fully connected layer works with one-dimensional vector numbers. So the three-dimensional output of predecessors needs to be converted to a one-dimensional vector. Figure 3 shows deep CNN architecture for thoracic disease classification.

The performance of CNN is dependent on the following:

1. Network architecture adopted: Different CNNs for image feature learning are AlexNet, VGG, GoogLeNet, ResNet, and DenseNet.
2. Amount of training data: Less training data results in overfitting. So we need to produce more input data from the current dataset. With the help of some common transformations, more input data are created; such a process is called data augmentation. In this way, we artificially boost the size of the training set to reduce overfitting. Dropout is also used to prevent overfitting during training

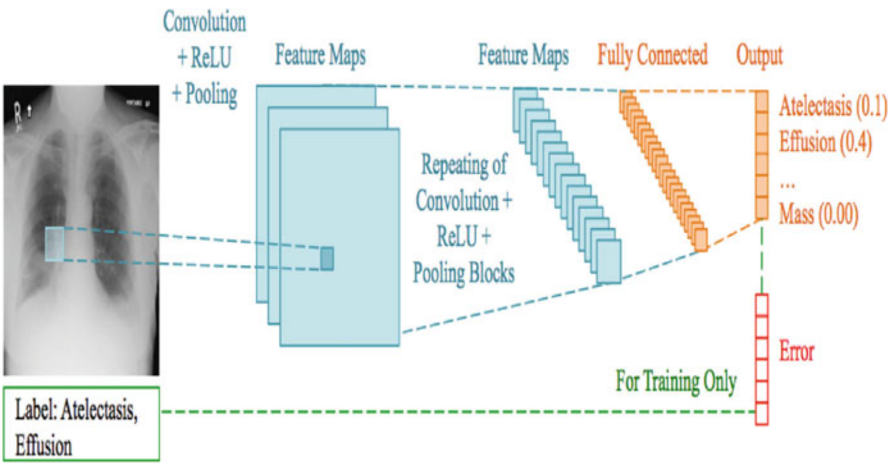


Fig. 3 Deep CNN model for multiple disease classification

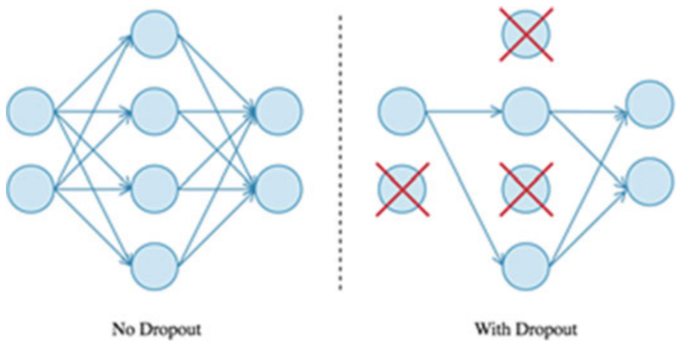


Fig. 4 Dropout of nodes

time. At each iteration, a neuron is temporarily "dropped" or disabled with probability as shown in Fig. 4. Dropout can be applied to the input or hidden layer nodes but not the output nodes. Adding dropout is the most popular regularization technique for a deep neural network's accuracy boost.

2.3 Training Types

Transfer learning is a training strategy of new deep models, where training is initiated with a pretrained model, instead of from scratch. The pretrained model is used, when we have limited availability of training data. The final layers are fine-tuned with the same available training data. The pretrained models are trained with many databases, which contain all types of images. Advantages of transfer learning are performance enhancement in the new deep model, reduction in training time for the new model, and less training data which may be enough for the new task domain.

3 Multiple Disease Classification

The deep neural networks (DNNs) provide good results, only if we feed enough data to them [5]. CAD for thoracic diseases is very challenging due to limited number of chest radiograph database in digital form.

3.1 Evaluation Metrics Used

The evaluation of all existing methods in multiple disease classification frameworks is done using a metric called area under the curve (AUC). It measures the area under the receiver operating characteristic (ROC) curve. The performance of the classifier on individual pathology is evaluated using per-class AUC, and for multiple diseases, the average value of AUC is taken. Hence, a higher average AUC indicates a better classifier. Both AUCs are always in the range of 0–1. Most of the time, AUCs are expressed in percentages.

3.2 Popular Dataset

The first ever publicly available chest X-ray dataset is Open I [6]. It has 7470 X-ray images in digital form. This dataset came in 2016 with automated image annotation, but none of the diseases were detected. Major achievement happened next year, when a new chest X-ray database was made public by NIH (National Institute of Health). It had a total of eight thoracic disease labels. These labels are created using natural language processing (NLP) process [5]. They also proposed a thoracic disease classification framework with four CNN pertained models AlexNet, GoogLeNet, VGG-16, and Resnet-50 for the initialization of a network. Out of these, Resnet-50 shows better performance with an average AUC of 73.8%.

However, they used a binary relevance approach for classification even though most of the images contain multiple pathology as shown in Fig. 5. Later, NIH

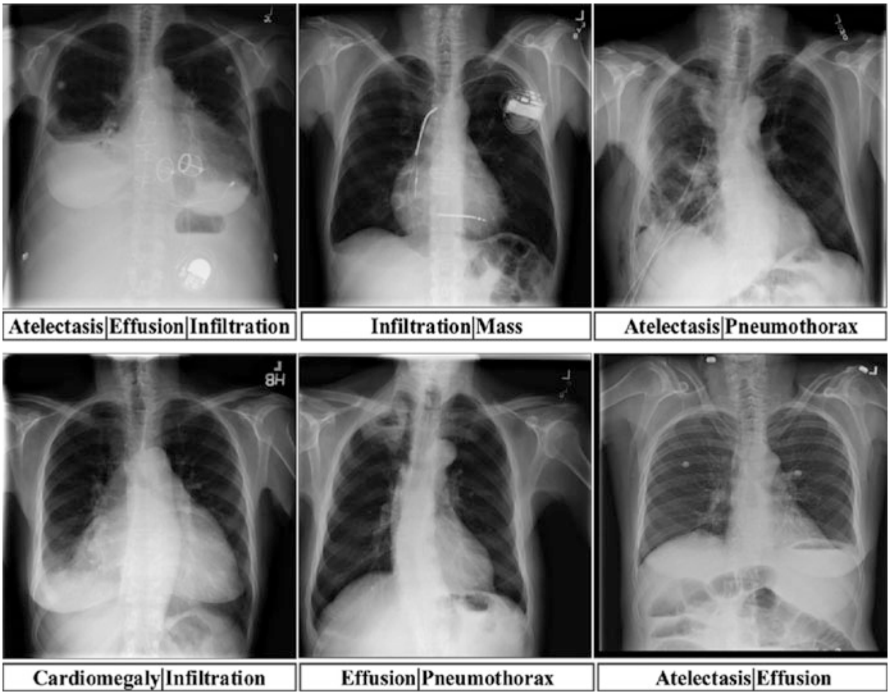


Fig. 5 Multilabel images of chest X-ray14 dataset

expanded this database from 8 to 14 thoracic diseases and named it has chest X-ray14 dataset [5]. More recently, different state-of-the-art models were proposed by researchers to address chest X-ray14 multiple disease classification problems, which are discussed next.

3.3 State-of-the-Art Models

Yalo et al. [7] proposed a DenseNet model with an average AUC score of 80.1%. They tried to discover inter-relation among all 14 pathologies. Their model is trained from scratch without any transfer learning, and their designed LSTM (long short-term memory) model ignored label dependencies also.

Rajpurkar et al. [8] proposed a method called as CheXNet. It is pretrained with DenseNet-121 with batch normalization and achieved an average AUC score of 84.2%. The heatmaps are generated with the help of CAM (class activation map). In CheXNet method, AUC score on pathology pneumonia equaled the prediction of radiologists.

Kumar et al. [9] developed cascaded DenseNet for thoracic disease classification task. The performance classification of thoracic disease, cardiomegaly, was improved with per class AUC score of 91.33% than previous methods. However, the above methods used the entire image for the detection of diseases even though some pathology areas are very small in the images. Usage of entire image in training process resulted in more computational complexity and noise.

Guan et al. [10] developed AG-CNN, a deep learning framework, and obtained good score of 87.1%. Here, they introduced a fusion branch in their method to fuse both global and local information of X-ray image. But here, some misclassification of pathology are observed due to CAM-generated heatmaps.

Wang et al. [11] introduced a model on chest X-ray14 database and called as ChestNet. They introduced Grad-CAM-based attention block to regulate CNN-based feature extractor. The CNN pretrained ResNet-152 is used as a main network with a support of attention block. Their proposed model outperformed previous models with an average AUC result of 89.6%. Figure 6 shows Grad-CAM [12]-generated heatmaps from ChestNet for thoracic disease localization.

Segmentation-based deep fusion network [13] does thoracic pathology detection in a unique way. Here, two CNNs are used to act on entire X-ray and also on a cropped lung region of X-ray. Both features are later fused together for classification work. The average AUC of 81.5% was achieved with very high computation complexity and more computation time.

Chen et al.'s [14] DualCheXNet architecture was introduced. It works with dual asymmetric feature learning from both ResNet and DenseNet architectures. It does embed complementary feature learning using both CNNs. This model achieved a good average AUC score of 85.6%. Figure 7 shows the ROC (receiver operating characteristics) curve plot of their model. The ROC curve of all 14 pathologies of the standard chest X-ray14 dataset is shown.

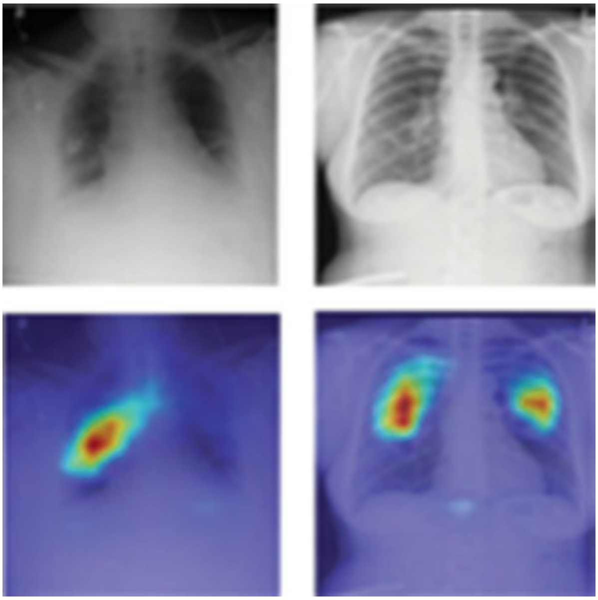


Fig. 6 Grad-CAM generates heatmap for a localization of atelectasis (left) and nodule

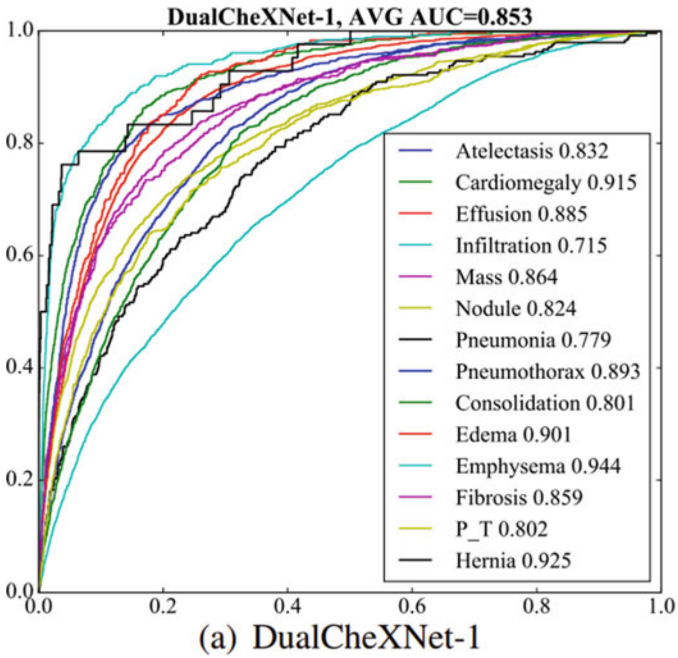


Fig. 7 The ROC curve plot of DualCheXNet

The area under the ROC curve of each pathology is calculated. Later average of all 14 AUC values is used to find the average AUC value of the model. The disadvantage of the above method is its failure in detecting textures of lung portion in an image. Table 1 gives a brief summary of all state-of-the-art deep models using NIH dataset.

4 Discussion and Future Trends

This paper discussed various CNN-based learning algorithms for the classification of multiple diseases belonging to thoracic cavity. It is proved that algorithms' performance reached radiologists' level predictions in most of the pathologies. This could be the main direction toward improved health care for patients of developing countries. However, the limitation of all these deep learning methods is that they are data-hungry. Still, only few chest X-ray dataset are available in digital form. The present dataset has very imbalanced disease class distribution. This imbalance affects model performance on minority class pathology.

Coming to multiple disease classification problems, the future direction is accurate detection of Covid-19 and also degeneration of pathology in the human body, for example, change of Covid-19 to pneumonia, etc. The other difficulty is that thoracic disease patterns in X-ray are very diverse in nature, with many overlaps. In most of the models, localization of pathology is done in a weakly supervised fashion. A lot of work needs to be done on the accurate localization of pathology on chest X-rays.

The selection of appropriate data sampling techniques and loss function also play a critical role in success of diagnosis. There are three sampling techniques. They are under-sampling training data, penalizing the misclassification, and oversampling. Under-sampling the training data results in wasting part of the training dataset, which has already been relatively small in chest X-ray14. Penalizing the misclassification is aimed to impose an additional cost on the model for making classification mistakes or misclassifications of the minority class during training. Weighted cross-entropy loss functions or standard binary cross-entropy loss functions are used to do this penalization in [11]. Here the difficulty is manually fixing weights for different misclassifications. A moderate improvement is noticed in their model [11] to overcome data imbalance issues. A type of oversampling technique called adaptive augmentation with unweighted cross-entropy loss function is used in [16]. A good improvement is observed here to overcome the class imbalance issue of the chest X-ray dataset.

Table 1 The overview of all the models

	Description		
	Methodology	Avg AUC	Advantage
Wang et al. [5]	Pretrained with 4 CNN + binary relevance approach	74.2%	First model
Yao et al. [7]	DenseNet and LSTM	80.2%	Correlation among pathology labels is discovered
Rajpurkar et al. [8]	Pretrained with DenseNet-121 with batch normalization and CAM-generated heatmaps	83.5%	Score of pneumonia (76.8%) exceeded predictions of radiologist
Kumar et al. [9]	DenseNet with cascade learning	78.5%	Score of cardiomegaly (91.33%); more than any model
Guan et al. [10]	DenseNet-based pretrained CNN with attention guidance	87.1%	Improvement in small nodule detection
Wang et al. [11]	Grad-CAM-supported ResNet	88.4%	High average AUC
Liu et al. [13]	Segmentation-based deep fusion network	81.5%	The 95.84% lung region identified
Chen et al. [14]	ResNet and DenseNet used in complementary fashion	86.7%	Unique type of feature learning
Hongyu et al. [15]	Triple attention learning-based DenseNet	82.6%	Three-way attention learning
Wang et al. [16]	Adaptive sampling-based DenseNet	82%	Adaptive augmentation to overcome class imbalance issue successfully

5 Conclusion

This survey discussed various artificial intelligence-based works on chest X-ray radiography, along with a brief background. Transfer learning with pretrained CNN models is used more instead of learning from scratch. At the end, fine-tuning was done in end-to-end manner. Out of several pretrained CNN models, DenseNet showed promising results as a feature extractor than ResNet. In most of the models, sigmoid activation functions with a specific number of neurons are kept at the last layer for the purpose of classification. Evaluation metrics such as AUCs are used for the performance analysis of classifiers. Under localization of abnormal regions, Grad-Cam-generated heatmaps gave better accuracy than CAM-generated heatmaps. In future, these AI-assisted diagnoses will provide improved health care to society and decreased workload to medical community.

References

1. K. Yasaka, O. Abe, Deep learning and artificial intelligence in radiology: Current applications and future directions. PLoS Med. **15**(11), e1002707 (2018) https://doi.org/10.1371
2. G. Chartrand, P.M. Cheng, E. Vorontsov, M. Drozdzal, S. Turcotte, C.J. Pal, S. Kadoury, A. Tang, Deep learning: A primer for radiologists. Radiographics **37**(7), 2113–2131 (2017)
3. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. **1**, 1097–1105 (2012)
4. I. Allaouzi, M.B. Ahmed, A novel approach for multi-label chest X-ray classification of common thorax diseases. IEEE Access **7**, 64279–64288 (2019)
5. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2097–2106
6. Open-i: An open access biomedical search engine
7. L. Yao, J. Prosky, E. Poblenz, B. Covington, K. Lyman, Weakly supervised medical diagnosis and localization from multiple resolutions (2018), arXiv preprint arXiv: 1803.07703
8. P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, A. Y. Ng, CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning (2017), arXiv:1711.05225 [cs, stat]. *arXiv: 1711.05225*
9. P. Kumar, M. Grewal, M.M. Srivastava, Boosted cascaded convnets for multilabel classification of thoracic diseases in chest radiographs (2017), arXiv:1711.08760. [Online]
10. Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, Y. Yang, Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification (2018), *arXiv:1801.09927*
11. H. Wang, H. Jia, L. Lu, Y. Xia, Thorax-net: An attention regularized deep neural network for classification of thoracic diseases on chest radiography. IEEE J. Biomed. Health Inform. **24**(2), 475–485 (2019)
12. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626
13. H. Liu, L. Wang, Y. Nan, F. Jin, Q. Wang, J. Pu, SDFN: Segmentation-based deep fusion network for thoracic disease classification in chest X-ray images. Comput. Med. Imaging Graph. **75**, 66–73 (2019)

14. B. Chen, J. Li, X. Guo, G. Lu, DualCheXNet: Dual asymmetric feature learning for thoracic disease classification in chest X-rays. Biomed. Signal Process. Control **53**, 101554 (2019)
15. H. Wang, S. Wang, Z. Qin, Y. Zhang, R. Li, Xia, Triple attention learning for classification of 14 thoracic diseases using chest radiography. Med. Image Anal. **67**, 101846 (2020)
16. H. Wang, Y.Y. Yang, Y. Pan, P. Han, Z.X. Li, H.G. Huang, S.Z. Zhu, Detecting thoracic diseases via representation learning with adaptive sampling. Neurocomputing **406**, 354–360 (2020)

Introduction to Deep Learning

Sohit Shukla, Neelendra Badal, and Bhawesh Kumar Thakur

1 Introduction

Deep learning is an integral part of concept machine learning, which is based upon artificial neural networks.

Initial research indicates that a linear perceptron cannot be used as a ubiquitous classifier but that a network with function of execution, having non-polynomial nature, and with unbounded width, a hidden layer can be used. Deep learning is the latest aberrant that takes account of a bounded-sized unbounded number of layers, allowing optimization for a functional application while maintaining a theoretical unique approach under mild conditions. Its layers can be heterogeneous as well as indifferent from the biologically informed connection models for good understandability, trainability, and the sake of performance [1].

Artificial intelligence (AI) based on predefined sets of rules given by a subject matter expert is heavily focused on rule-based systems that made predictions. However, since these structures were brittle and relied on "expert advice," they gradually fell out of the course of working profile. These approaches can be placed into a more data-driven approach, machine learning, as the size and volume of data increase (Fig. 1).

A genre of machine learning algorithms is based on the utilization of multiple layers for extracting higher-level features against the raw inputted data, professionally known to be deep learning. You can take examples in image processing, where

S. Shukla · N. Badal
Department of Computer Science & Engineering, KNIT, Sultanpur, U.P, India
e-mail: sohit.2400@knit.ac.in; nbadal@knit.ac.in

B. K. Thakur (✉)
Department of Information Technology, BBDNIIT, Lucknow, U.P, India

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_15

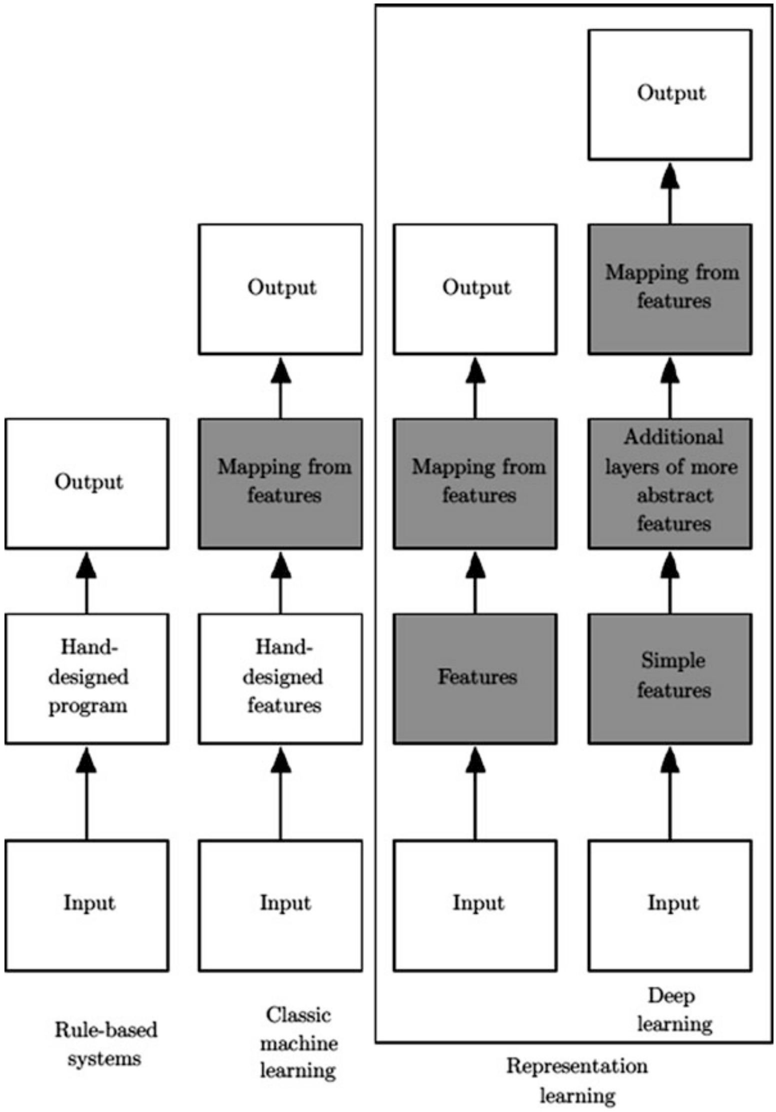


Fig. 1 Flowchart illustrates how various components of AI systems interact within various AI disciplines

higher layers are used to identify the concepts for humans (digits, letters, faces) while lower layers may identify edges.

In the recent model of the machine learning apart from the perspective of neuroscience, "deep learning" is a modern term. It works over the concept of training at the various levels of composition, also that is not influenced neutrally, and can be used by machine learning systems.

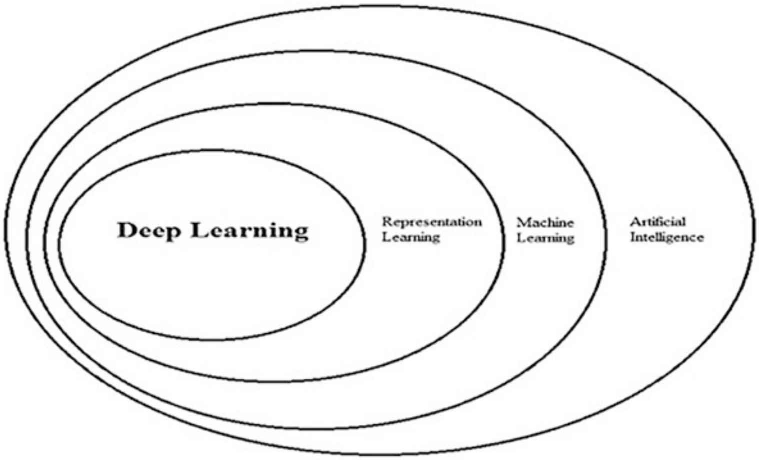


Fig. 2 The diagram shows how AI approaches are used in many but not all learning as well representation learning. An example of AI technology is included in each segment of the Venn diagram

2 Machine Learning Versus Deep Learning

There are several aligned algorithms for the formation of machine learning functions that equip it to trace patterns within data and use this kind of architecture for the solution to a problem. Machines attempt to comprehend these fundamental patterns in a variety of ways. Here, we will try to know how machine learning is connected to deep learning. Also, try to elaborate few of its key applications and challenges [2]. In general, there are some misconceptions attached to deep learning; in the reference to machine learning, deep learning is a competitive technology (Fig. 2).

To align things in order, machine learning is a field in which deep learning belongs. With extreme computational power over the large data sets, deep learning algorithms can learn themselves from hidden patterns across the data for creating a relevant outcome.

In an abstract view, we can say deep learning is a branch of machine learning that is handled by several units with fine computational power in a managed way to make a prediction and is trained on large quantities of data.

3 The Architecture of Deep Learning

Deep learning architecture and algorithms can be expressed in different shapes and sizes. Now, we will create knowledge about the six architectures of deep learning from the past two decades. Remarkably, the oldest approaches are only two, one is long short-term memory (LSTM) and another one is convolutional neural networks (CNNs), used in various applications [3].

In the section below, we will illustrate the supervised and unsupervised learning architecture of deep learning:

1. Convolutional neural networks (CNNs)
2. Recurrent neural networks (RNNs)
3. Long short-term memory/gated recurrent unit (GRU)
4. Self-organizing map (SOM)
5. Autoencoders (AE)
6. Restricted Boltzmann machine (RBM) (Fig. 3)

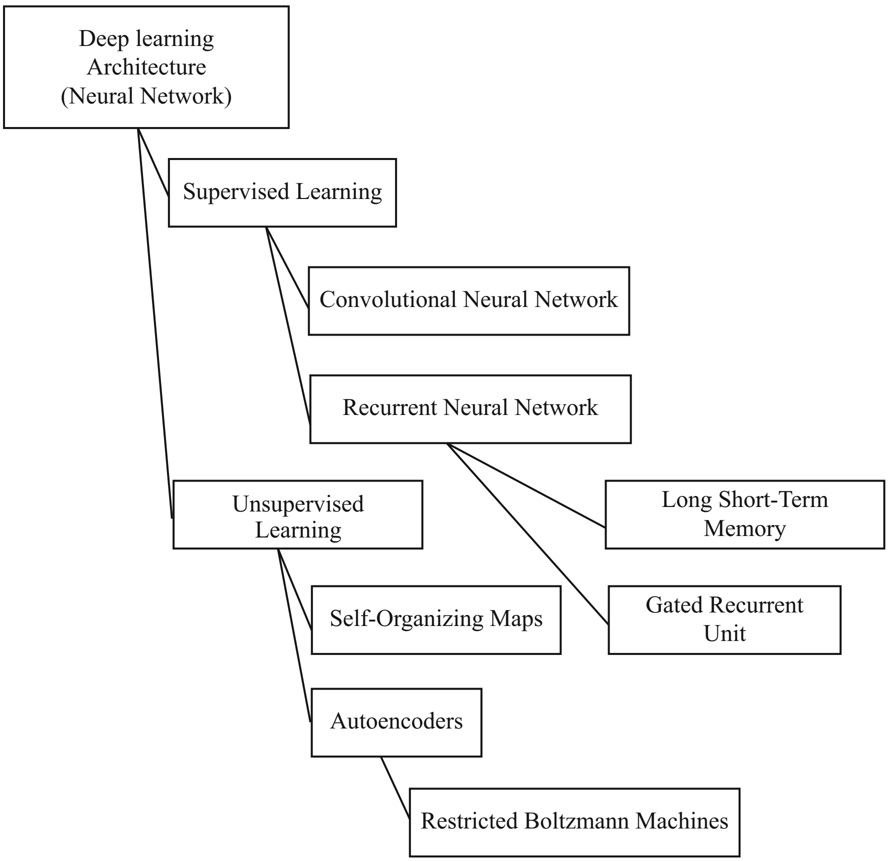


Fig. 3 Deep learning architecture

3.1 Supervised Learning

A learning technique is known to be supervised learning that has a very clearly identified goal, which lies in the domain of the targeted data used for the learning.

3.2 Convolutional Neural Network

To be treated as the biological multilayer neural network, the basic concept of CNN is inspired by the visual cortex of animals [4]. In image processing applications, architecture is especially useful. Yann LeCun developed the first CNN, which focused on character recognition, which is handwritten such as interpretation of the postal code at an instant of time [5]. Early layers of a deep network identify edges as a feature, and the layers above these are responsible to convene these extracted features into the attributes which can be inputted at a higher level.

Classification and feature extraction are required at several layers of LeNet CNN architecture (see the following image). Features were extracted from the input images, which were fed into a convolutional layer and divided into receptive fields. Pooling is the next step, preserving the accurate knowledge from the sample; the dimensions of the feature diminish in order. The data is then fed into a completely connected multilayer perceptron after another convolution and pooling stage. This network's final output layer is a cluster of nodes that define image features. Back-propagation is the most popular method used here for training purposes (Fig. 4).

Convolutions, deep processing layers, pooling, and complete linked classification are exact openings for diverse fresh deep learning neural network applications. Multiple tasks are executed in video recognition that lies within the NLP (natural language processing). Besides this, CNN can be well applied in image processing.

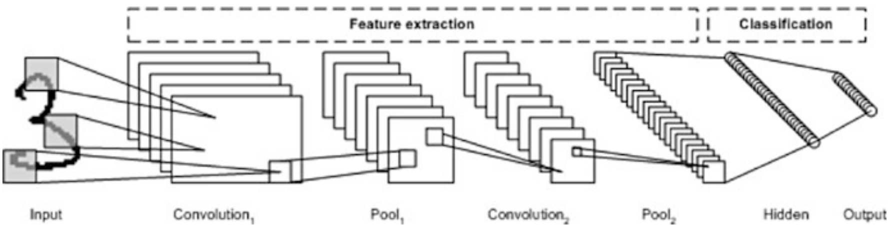


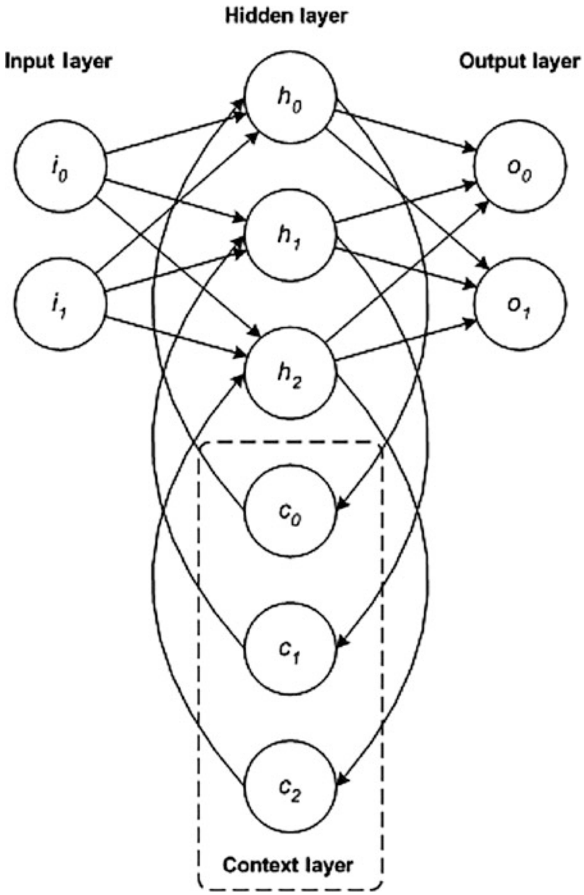
Fig. 4 Feature extraction and classification at various layers of LeNet CNN

3.3 Recurrent Neural Networks

The RNN is considered to be a framework in deep learning as it has basic network architecture that can be used for the design of the other deep learning architecture. The recurrent network might have a connection that feedback to last layers despite full feed-forward (or into the same layer); this differentiates it from the traditional multilayer network. The feedback system enables the RNNs to keep track of recently acquired inputs and loopholes within the model in a run-time environment [6].

RNNs can be built in a variety of ways. The main differentiation is network feedback, which may come from the combination of two layers or one of them (hidden layer and O/P layer) (Fig. 5).

Fig. 5 Recurrent neural networks



3.4 LSTM Networks

Hochreiter and Schmidhuber developed the LSTM in 1997, but it has gained popularity as an architecture framework of RNN that can be used in various applications [7]. LSTMs can be used in items that you use daily, such as smartphones. In IBM Watson®, LSTMs are used to set new standards in conversational speech recognition.

The LSTM used the idea of a memory cell. As a function of its inputs, the cell contains a memory that will maintain its value for a short or long time, allowing the cell to retain important content rather than the value computed at last.

The three gates are responsible to regulate information flow in and out of the LSTM memory cell. The input gate regulates when new data will enter the memory. On the loss of the run timepiece of information, forget gate recalls new data. At the last, the output gate determines whether or not the information present in the cell is included in the cell's output. The weight in the cell is responsible to regulate the gate. On the basis of resulting error in network performance, the training algorithm usually referred to as BPTT (back-propagation through time) optimizes these weights (Fig. 6).

Recent CNN and LSTM applications have resulted in the image and video captioning systems that caption images and videos in natural language. In the processing of the image or video CNN, and LSTM is frequently used; as these are found very suitable to translate the output into natural language.

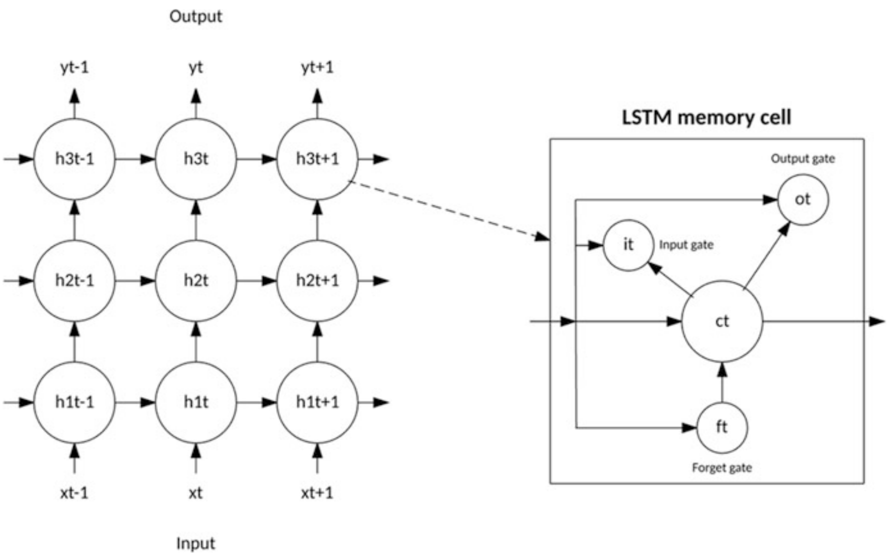


Fig. 6 LSTM memory cell architecture

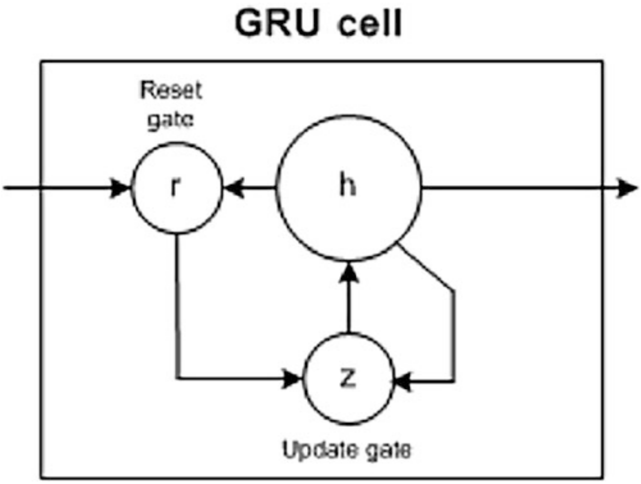


Fig. 7 GRU cell architecture

3.5 GRU Networks

The gated recurrent unit (GRU), a simplified version of the LSTM, was introduced in the year 2014. Two other types of gates are used in place of the output gate, as in the model of LSTM [8]. The gates mentioned are an update gate and a reset gate. Update gate plays a significant role in retaining the quantity of previous cell content that needs to be preserved, whereas the role of the reset gate is to combine the previous cell and the new cell content. The gate value is set to 1 and the update gate value to 0, which makes a gate recurrent unit model a regular RNN (Fig. 7).

GRU is a much simpler model as compared to the LSTM and can be trained faster in an efficient way. The LSTM, on the other hand with huge data, can produce fine results and fine expressiveness.

3.6 Unsupervised Deep Learning

Unsupervised learning is where the target label does not exist in the data of the domain used for the training [9]. In this section, we will discuss the different types of unsupervised deep learning architecture.

3.7 Self-Organized Map

A self-organized map (SOM) come into light by Dr. Teuvo Kohonen in the year 1982 and was popularly known as Kohonen map [10]. By the reduction in the dimension of input, clusters are created in an unsupervised neural network. SOM creates a reasonable difference from the traditional artificial neural network (Fig. 8).

The very first difference is the weight served as the characteristic of the nodes present. The random input is picked just after the normalization. Each of the features of the record inputted is initialized by the random weight and very close to zero. In this way, the input node is represented by these weights, and the variation in the input is represented by the combination of the weights. The Euclidean distance calculation is done between output nodes and the input nodes in one-to-one approach. The least Euclidian distance is used for the precise presentation of the input and marking of the best BMU (building maintenance unit). Similarly, other entities like distance and the radius of points around BMU weights are calculated, which are also assigned to the cluster based on proximity.

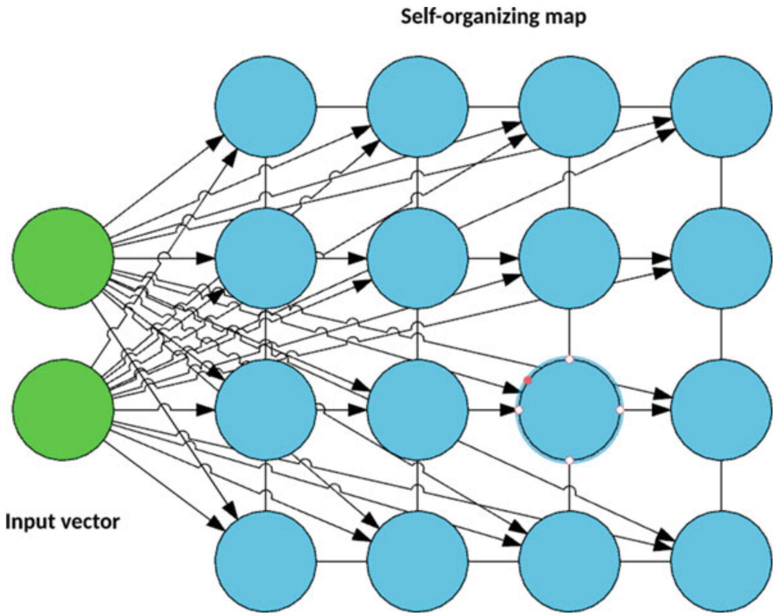


Fig. 8 Self-organizing map architecture

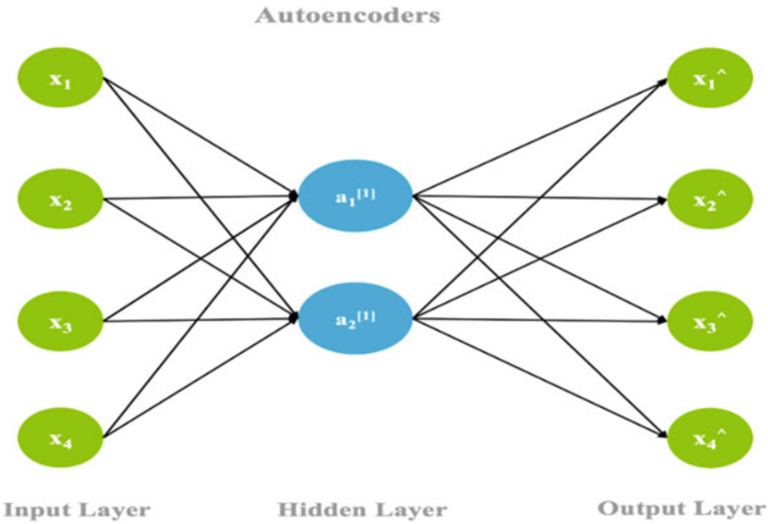


Fig. 9 Autoencoder's architecture

3.8 Autoencoder

There is some uncertainty in the invention of the autoencoder, despite this it was found by LeCun in 1987. The variant itself comprises three layers, input, hidden, and output layers, which is ANN [11].

Firstly, using an appropriate encoding function, the input layer can be encoded to the hidden layer. The counting of nodes in the input layer is more compared to the hidden layer. The precise presentation of actual input is contained in the hidden layer. By the use of the decoder function, output layers reconstruct the input layers [12] (Fig. 9).

During the training phase, by the use of the error function, the gap between the input and the output layers has been calculated, and weights are adjusted to minimize the error. The autoencoders learn continuously using backward propagation. For this reason, autoencoders are classified as self-supervised algorithms.

3.9 Restricted Boltzmann Machine

Paul Smolensky originally invented the RBMs in 1986 known to be Harmonium, yet RBMs became popular much later [13].

An RBM is considered to be an approach that has a two-layer neural network, that is, input layer and hidden layer. The figure given below depicts how every node in a visible layer of RBMs is connected to every node in the hidden layer.

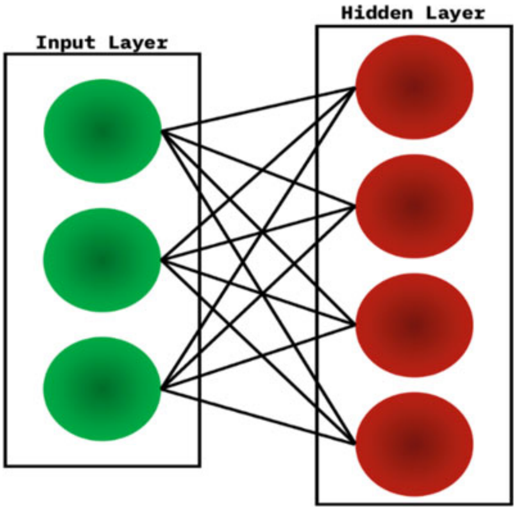


Fig. 10 Restricted Boltzmann machine's architecture

Input and hidden layer nodes are also connected in the traditional Boltzmann machine, whereas the nodes within the layer in restricted Boltzmann machine are not connected due to computational complexity (Fig. 10).

In the training phase, a stochastic approach is in action for the calculation of the probability distribution in the RBMs. Randomized activation of neurons is placed as the training begins. The visible layer helps in reconstructing the input whereas the hidden layer is for a forward pass to create the activation.

When the original inputs are different from the reconstructed input in an RBM, they are known to be generative models [14].

It is also seen that built-in randomness creates similar results for different outputs. For the autoencoder having the deterministic model, the most significant difference is this.

3.10 Deep Belief Network (DBN)

With a unique training algorithm, the DBN is a standard network architecture. The DBN is typically deep, and several hidden layers of multiple layered networks have RBM in each layer pair. So the stacks of the RBMs are known to be DBN [15].

The raw sensory inputs are represented by the input layer in the DBN, and the condensed representation of these inputs is learned from each of the hidden layers. The output layer is differently handled from the other layer to implement the network classification. Unsupervised initial training and guided fine-tuning are two stages of training (Fig. 11).

To recreate the inputs in unsupervised initial training, each of the RBMs is trained. In this sequence, training of RBM is done, and the input layer will be the first

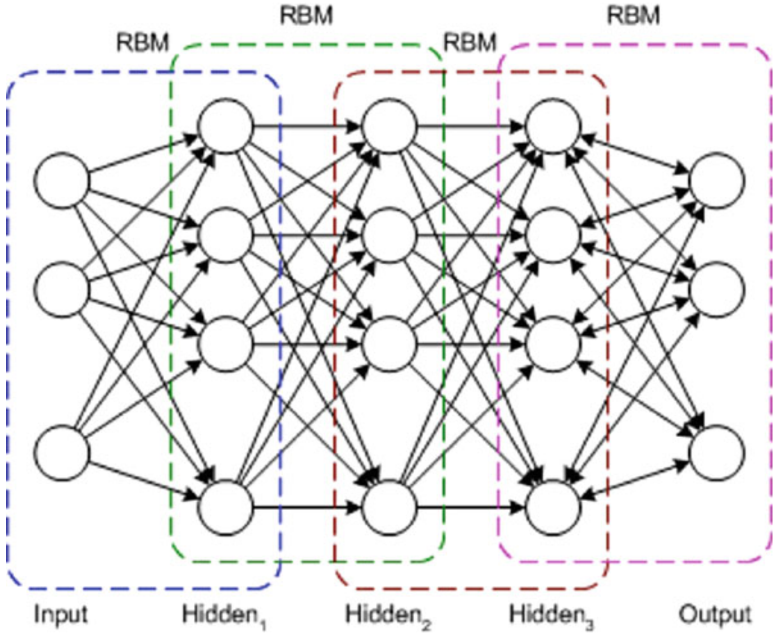


Fig. 11 Deep belief network architecture

hidden layer (or visible). Besides this, RBM training is done by using the first hidden layer's output as the input. Till each of the layers has been pretrained, this process is repeated again and again. As the pretraining was completed, the fine-tuning phase started. To notify them significantly, the output nodes were given labels (in the context of the network what they represent). Using back-propagation or gradient descent learning, full training has to be done to complete the training process.

3.11 Deep Stacking Network (DSN)

The DSN is also known as a deep convex network. It is different from a conventional deep learning system, and it is a deep collection of individual networks each with its hidden layers rather than a single deep network. The difficulty of training is one of the issues with deep learning, and this architecture addresses it. The complexity of the training increased exponentially by each layer of deep learning architecture; the DSN views training as a set of individual training problems rather than a single problem [16].

Several modules are combined to form DSN, each of which of them is a subnetwork in DSN's overall hierarchy. In any of one instance architecture, three modules are utilized in the DSN. These modules are created with the combination

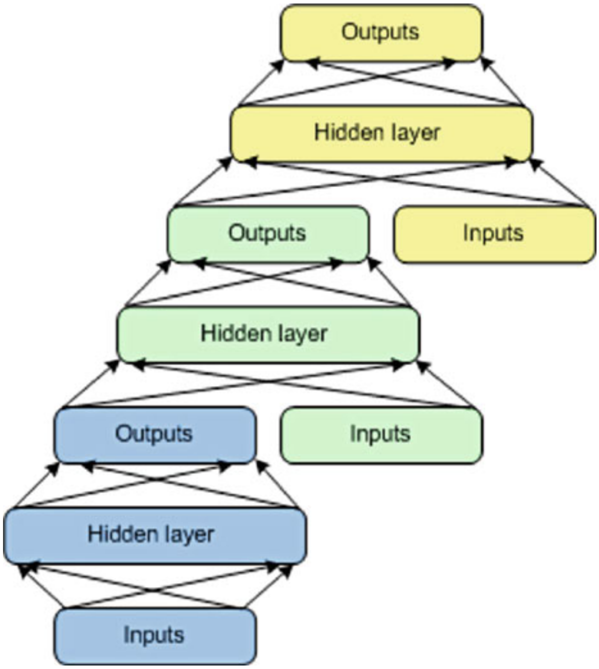


Fig. 12 Deep stacking network architecture

of an input layer, a single hidden layer, and an output layer. These modules are arranged in the form of the stack placed one on top of each other, with the prior layer, output, and the initial input vector serving as inputs (Fig. 12).

Individual modules can be trained in isolation using the DSN; due to the ability of parallel training, it is more effective. Instead of back-propagation across the entire network, for each of the modules, supervised training is implemented as back-propagation. DSNs outperform when used in the traditional DBNs for many problems. With this, they create a common as well as powerful network architecture.

4 Applications

Health care, banking, and image recognition have a wide variety of uses for deep learning. It has been observed that today, it can be applied to the field of inspection of the material science, computer as well as machine vision, speech recognition, natural language processing, bioinformatics, and board game programs and produced results surpassing the human expertise in few of the relevant cases and is comparable to all cases based on the performance.

Medical Image Analysis in the Field of Health Care The various health issues are perfectly applied by deep learning due to easier access to accelerated GPU (graphics processing unit) and inculcating a large amount of data. As input images from MRI and X-ray, the detection of cancer can make easy by the concept of image recognition. Several other domains like drug discovery and genomics are other health-care-based applications.

Autonomous Vehicles The dream of cars to be self-driven seems to be a critical field to automate, but now it has become a reality. From the level of identifying pedestrians over the road, and the various traffic signs, environments to monitor progress can also be performed by deep learning-based training models.

The Recommendation System for e-Commerce Applications In the market-place, deep learning plays a role in product recommendation as one of the profitable applications, as the customers can be anchored by personalized and accurate recommendations. This increases the number of visits to the product and accelerates sales, thus benefiting sellers.

Automatic Speech Recognition as a Personal Assistant The presence of the Alexa or Google Assistant at our office desk and in homes is a prominent example of deep learning-based personal assistants.

Enhancing the Audio and Video Capture With the help of deep learning, the video and the audio streaming gets a new wing to fly with the high quality of the picture.

References

1. W.G. Hatcher, W. Yu, A survey of deep learning: Platforms, applications, and emerging research trends. IEEE Access **6**, 24411–24432 (2018). https://doi.org/10.1109/ACCESS.2018.2830661
2. H.-D. Wehle, Machine learning, deep learning, and AI: What's the difference? (2017)
3. L. Deng, D. Yu, Deep Learning: Methods and Applications. Found. Trends Signal Process. **7**(3–4), 1–199 (2014). https://doi.org/10.1561/2000000039
4. Y. Gong, Q. Zhang, Hashtag recommendation using attention-based convolutional neural network, in Proceedings of 25th International Joint Conference on Artificial Intelligence, New York, 2016, pp. 2782–2788
5. Y. LeCun, Generalization and network design strategies. Technical Report CRG-TR-89-4, University of Toronto (1989)
6. C. Wu, J. Wang, W. Liu, J. Liu, Recurrent neural network-based recommendation for time heterogeneous feedback. Knowl.-Based Syst. **109**, 90–103 (2016)
7. S. Hochreiter, J. Schmidhuber, Long short-term memory. Neural Comput. **9**, 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
8. R. Dey, F. Salem, Gate-Variants of Gated Recurrent Unit (GRU) neural networks (2017)
9. D. Oh, B. Kim, J. Lee, Y.-G. Shin, Unsupervised deep learning network with self-attention mechanism for non-rigid registration of 3D brain MR images. J. Med. Imag. Health Inform.

11, 736–751 (2021). https://doi.org/10.1166/jmihi.2021.3345
10. T. Kohonen, M.R. Schroeder, T.S. Huang, *Self-Organizing Maps*, 3rd edn. (Springer-Verlag New York, Inc, 2001). ISBN: 3540679219
11. S. Deng, L. Huang, G. Xu, X. Wu, Z. Wu, On deep learning for trust-aware recommendations in social networks. IEEE Trans. Neural Netw. Learn. Syst. **28**(5), 1164–1177 (2017)
12. F. Strub, J. Mary, Collaborative filtering with stacked denoising autoencoders and sparse inputs, in Proceedings of the NIPS Workshop on Machine Learning for E-Commerce, Montreal, 2015, pp. 150–158
13. K. Georgiev, P. Nakov, A non-IID framework for collaborative filtering with restricted Boltzmann machines, in Proceedings of the 30th International Conference on Machine Learning, pp III–1148–III– 1156 (2013)
14. F. Yuan, L. Yao, B. Benatallah, Exploring missing interactions: A convolutional generative adversarial network for collaborative filtering, in Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 1773–1782
15. A. Khan, muhammad Islam, Deep belief networks, in *Introduction to Deep Neural Networks* (2016). https://doi.org/10.13140/RG.2.2.17217.15200
16. L. Deng, X. He, J. Gao, Deep stacking networks for information retrieval, in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, (2013), pp. 3153–3157. https://doi.org/10.1109/ICASSP.2013.6638239

Methodological Assessment of Various Algorithm Types for Load Balancing in Cloud Computing

Dhruvi Trivedi, Naina Parmar, and Mrugendrasinh Rahevar

1 Introduction

More than a hundred million computer units are now interconnected. Such devices put in a request and obtain a solution very quickly. In [1], Fig. 1 illustrates how different devices (tablets, PCs, and laptops) may connect to a cloud and access data at any time. The term "cloud" is used to describe the Internet [3]. It offers facilities like hardware, software, applications, and databases on a need-to-know basis. The basic goals of cloud-based solutions are to lower costs, improve reaction times, and improve performance. Thus, cloud computing is also known as a pool of resources. In [2] to put it another way, cloud computing (CC) is a trending technology to deliver facilities to customers at any time. Resources are spread all over the world in a cloud computing system (CCS) to provide speedier service to clients [4].

Clients can use a variety of devices to access information, including PCs, mobile phones, and tablets. Security, effective balancing of the burden, scheduling of resources, QoS management, scaling and energy usage in data centers, continuous service and data lock-in, and observation of efficiency are among challenges that cloud computing has faced. Virtualization, distributed computing, and web services are all used in cloud computing to give a flexible approach to store data and files. The objective of virtualization is to provide as many capabilities as possible for the least amount of money at any moment. Cloud computing services (CCS) are divided into three categories, software as a service (SaaS), platform as a service

D. Trivedi (✉) · M. Rahevar
CSPIT – U & P.U. Patel Department of Computer Engineering, Anand, Gujarat, India
e-mail: dhruvitrivedi.ce@charusat.ac.in; mrugendrarahevar.ce@charusat.ac.in

N. Parmar
Devang Patel Institute of Advance Computer Engineering, Changa, Anand, Gujarat, India
e-mail: nainaparmar.dcs@charusat.ac.in

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_16

Fig. 1 A cloud computing scenario [2]

(PaaS), and infrastructure as a service (IaaS), with four deployment models: private, public, hybrid, and community.

Customers who make a service request and brokers that one operates as a middleman between the client and the cloud provider are two essential components of cloud systems. Following the conclusion of negotiations with various cloud retailers and cloud workers, it retains all cost-effective services. Some big IT firms, including HP, IBM, Apple, Google, Oracle, and others, have embraced mesh computing as a path to utility computing.

Several requests may come in from different geographical routes at the same time in the cloud system. The requests are distributed to several cloud providers at random, ensuring unequal load (request) distribution for each node. As a result, some of the nodes are overloaded while others are underloaded. The system's performance may suffer as a result of this condition. A load balance is essential in solving this problem.

2 Load Balancing

One of the most significant issues and concerns in cloud computing is load balancing. In [4] within a cloud balancing the burden environment, the process of distributing workloads and resources. Load balancing is a technique for allocating resources among numerous computers, networks, or servers to handle application or workload needs. It is commonly used to divide tasks across large amounts of servers and data traffic. To accomplish speed and efficiency in the cloud, advanced architectures are used. Load balancing has various properties including equitable task distribution across all nodes, ease of obtaining user satisfaction, improved

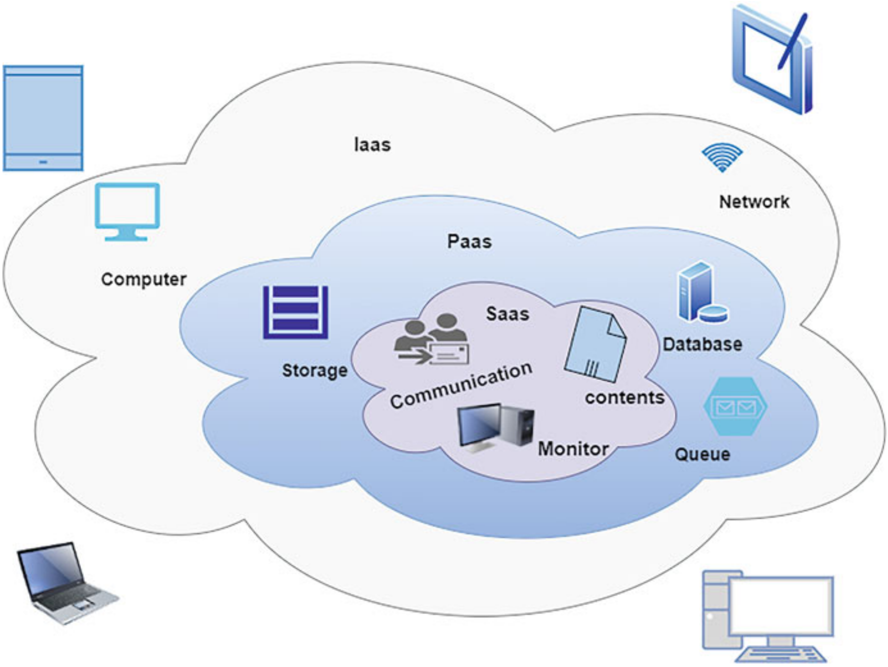


Fig. 2 View of the cloud computing environment [5]

overall performance of the system, a shorter response time, and delivering facility to maximize way use [6]. In cloud computing, load balancing is depicted in Fig. 2. As an example, suppose we create a cloud-based application that hundreds of people are anticipated to access at any given moment. As a result, response times for hundreds of people will become extremely slow and quickly become overburdened leading to poor responses and dissatisfied customers. Work is spread among numerous nodes when we implement load balancing on our application, resulting in improved performance and faster response (Fig. 3).

This section covers the source of scientific paper information, screening process, quality criteria, and impact assessment. Table 1 contains a series of questions for conducting a survey to evaluate the current state of affairs of load balancing challenges.

3 Data Source

As a data source, we searched the Web of Science, Google Scholar, Scopus, books, and magazines for relevant journals, research papers, and conferences. In our search, we used the database listed below. The percentage of publications reviewed from various sources is shown in Fig. 4.

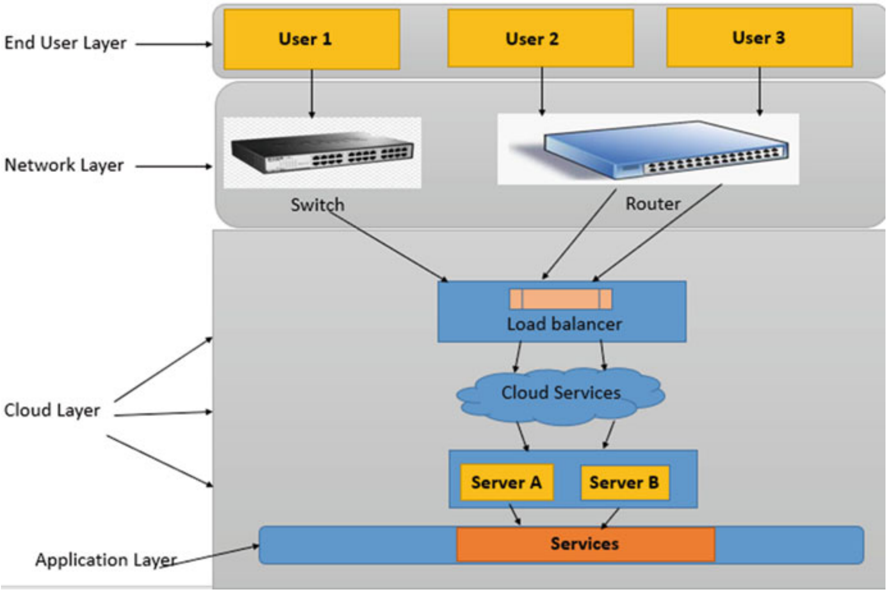


Fig. 3 Virtualization computation offloading

3.1 Search Criteria

3.2 Evaluation of the Level of Quality

We used a quality assessment process on the articles we found using the criterion for individualism and collectivism listed in Table 2. Based on basic keywords, we found 2, 64,000 papers from various publications from multiple roots, including ACM, Science Direct, IEEE, and Springer. Some publications were omitted because their titles were unrelated to our research, after reading the abstract, introduction, and conclusion (Fig. 5).

4 Methods for Having to Balance Workloads

The system's current state and future prospect process beginning are used to classify load balancing approaches [5, 17] (Fig. 6).

Table 1 Cross questions

No	Query for discussion	Motivation
1	How does load balancing play out in mesh computing?	The core goal stays to find numerous studies/articles that have been published over time and their importance with increasing use of the cloud. So far, a variety of load balancing approaches has been presented; all of load balancing approaches have been presented, all of which must be evaluated using various load balancing measures. This survey will also look for concerns and challenges with current load balancing strategies in order to assure QoS-based services. This section discusses load balancing strategies, depending on different categories
2	Why there is a requirement of balancing the burden in cloud?	
3	Some present approaches meet the requisite load balancing measures?	
4	What is the state of cloud load balancing right now?	
5	What criteria should a consumer use to choose a suitable service provider?	
6	What new algorithms must be presented in order to increase system performance?	
7	How may resource overloading and underloading be avoided by effectively managing resources?	
8	What are the quality of service (QoS) standards that the user expects from the cloud?	
9	How can massive amounts of strategies be established that meet the user's QoS requirements?	Various research papers from various load balancing categories must be found in order to uncover critical research issues. With the growing demand for cloud services, it is more important than ever to set load balancing criteria and develop strategies for properly handling high amounts of user requests. The various issues raised here will aid in identification of future study topics
10	How might accessible tools be used to validate existing and developing load balancing techniques?	
11	How does one create an architecture that meets the user's primary QoS requirement?	
12	What is the definition of SLA? What criteria are used to define SLA violations?	

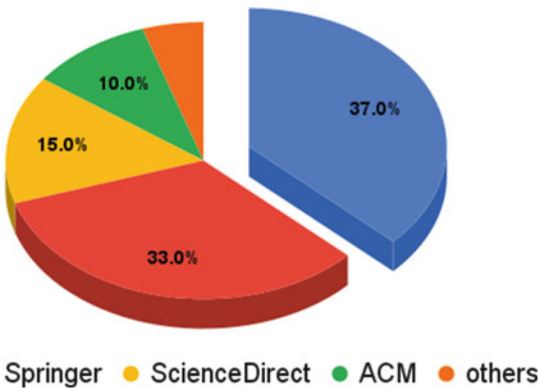


Fig. 4 Origin of a task scheduling research article that was examined

Table 2 Comparison table of load balancing techniques

Paper title	Algorithm	Advantages	Disadvantages
Improved Max-Min Algorithm in Cloud Computing [7]	Static load balancing method	It outperforms the old max-min method	Only RASA and the max-min algorithm are used to make comparisons
Analysis of Variants in Round Robin Algorithms for Load Balancing in Cloud Computing." *International Journal of computer science and Information Technologies* [8]	Static load balancing method	It is compatible with a wide range of processes	It is insufficient for load balancing to be based on processing overhead
An Efficient Data Locality Driven Task Scheduling Algorithm for Cloud Computing [9]	Static load balancing method	It works well in a variety of network states	When the cost fluctuates regularly, it is complex to locate a near-optimal solution
Performance analysis of load balancing algorithms [10]	Static load balancing method	It enhances the system's performance	When all remote processors are overloaded, load balancing becomes a problem
Efficient optimal algorithm of task scheduling in a cloud computing environment [11]	Static load balancing method	It outperforms the round-robin algorithm	It is difficult to achieve the perfect load balancing algorithm
Round-robin with server affinity: a VM load balancing algorithm for cloud-based infrastructure [12]	Static load balancing method	Over the RR approach, take less time to respond and process data in the data center	Important tasks are not given special consideration
A comparative study of artificial bee colony algorithms. Applied mathematics and computation [13]	Dynamic load balancing method	Increase the maximum throughput of the system	Due to the lack of supplemental information when employed in a sequential process, the solution slows down the operation and increases the processing cost
A Bee Colony based Multi-Objective Load Balancing Technique for Cloud Computing Environment [14]	Dynamic load balancing method	It allocates resources in accordance with their priorities	It adds extra load determination parameters
A Load Balancing Model Using Firefly Algorithm in Cloud Computing [15]	Dynamic load balancing method	It gives you more search options.	The results are affected by both distance and radiance effects in this case.
Load Balancing in Cloud Computing Using Dynamic Load Management Algorithm [16]	Dynamic load balancing method	Additionally, the performance is simulated using the cloud analyst simulation based on several characteristics like data processing time and response time, and the results are compared to a previously developed technique known as VM assign	The simulation results show that the current technique evenly distributed the load among servers by uniformly utilizing resources

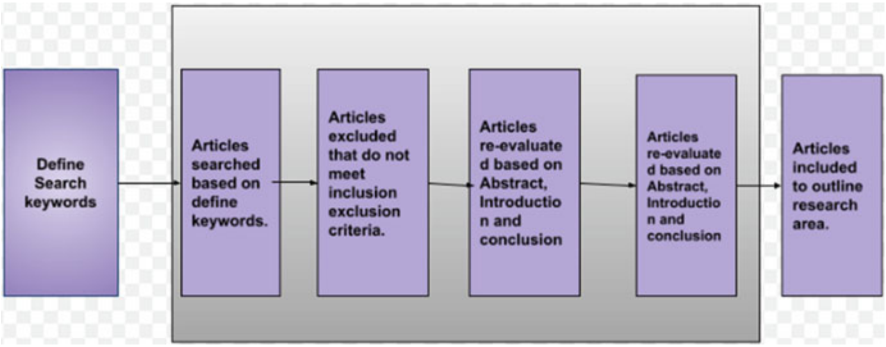


Fig. 5 Overview of the article identification procedure

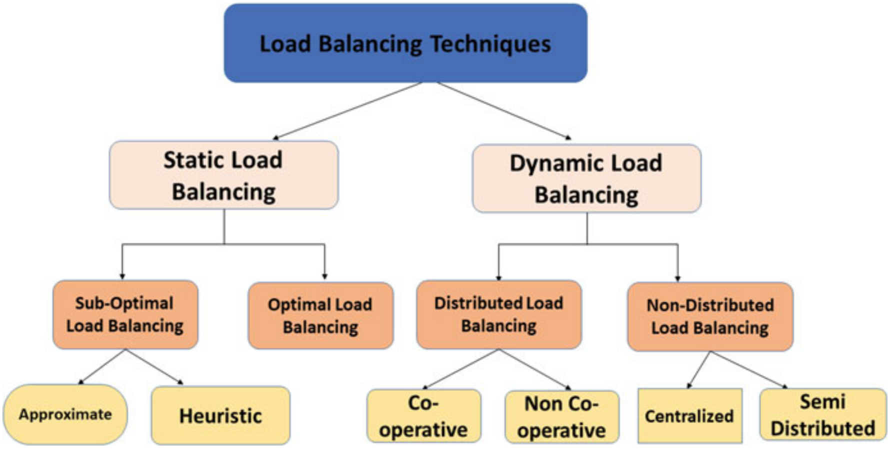


Fig. 6 Techniques for scheduling algorithms

4.1 There Are Two Varieties of It Based on the Current State of the System

1. Static algorithms
2. Dynamic algorithms

4.2 Load Balancing Metrics [18]

The following are some examples of qualitative measurements that might help improve the performance of a cloud load balancer (Fig. 7).

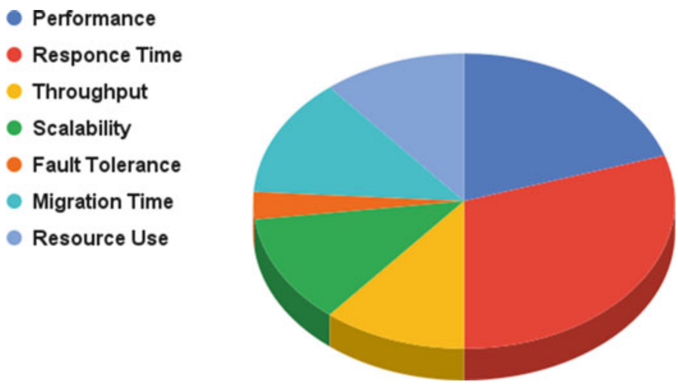


Fig. 7 Massive amount metrics as a percentage of all techniques evaluated

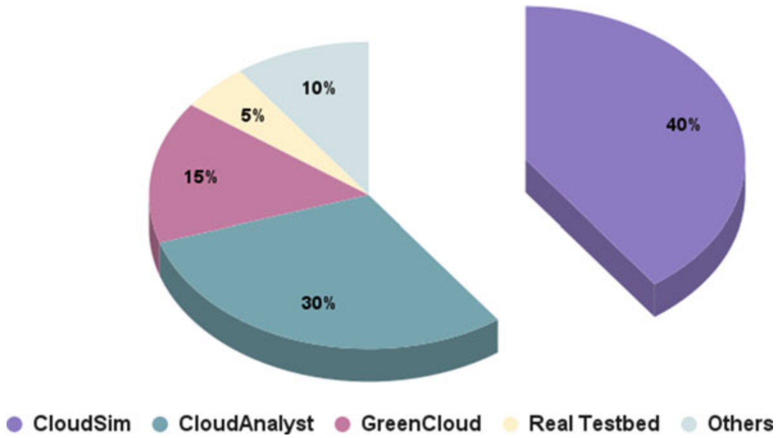


Fig. 8 Evaluation tools/techniques used in survey papers

5 Comparative Study (Fig. 8)

6 Conclusion

Workload balancing among cloud nodes is considered to be one of the most prominent difficulties in today's cloud systems. The above article offers a real-time look at auto-scaling concerns, obstacles, and different load balancing techniques. A comprehensive survey was conducted on a range of load balancing systems employing different parameters, according to the aforementioned research. We have divided these scalability tasks based on our findings within a number of groups, as shown in Fig. 6. We addressed the benefits, drawbacks, concepts, and obstacles of each technique in each area. There are quite a variety of algorithms that take into account the majority of total load parameters while also improving resource utilization and reaction time. The techniques, on the other hand, must be updated in order to boost the system's efficiency in the future. Researchers also looked at the scholarly literature in the field of resource allocation, which is a crucial component of cloud technology. The above study will benefit a massive amount of researchers by identifying research problems and providing a summary of load balancing approaches that are currently accessible.

References

1. P. Urvashi, M. Patel, Load balancing algorithm in cloud computing: A survey
2. S. Aslam, M.A. Shah, Load balancing algorithms in cloud computing: A survey of modern techniques, in *2015 National Software Engineering Conference (NSEC)*, (IEEE, 2015)
3. P. Mell, T. Grance, The NIST definition of cloud computing (2011)
4. E.J. Ghomi, A.M. Rahmani, N.N. Qader, Load-balancing algorithms in cloud computing: A survey. J. Netw. Comput. Appl. **88**, 50–71 (2017)
5. H. Shoja, H. Nahid, R. Azizi, A comparative survey on load balancing algorithms in cloud computing, in *Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, (IEEE, 2014)
6. R.G. Rajan, V. Jeyakrishnan, A survey on load balancing in cloud computing environments. Int. J. Adv. Res. Comput. Commun. Eng. **2**(12), 4726–4728 (2013)
7. O.M. Elzeki, M.Z. Reshad, M.A. Elsod. Improved Max-Min Algorithm in cloud computing
8. P. Samal, P. Mishra, Analysis of variants in round Robin algorithms for load balancing in cloud computing. Int. J. Comput. Sci. Inform. Technol. **4**(3), 416–419 (2013)
9. J. Jin, J. Luo, A. Song, F. Dong, R. Xiong, BAR: An Efficient Data Locality Driven Task Scheduling Algorithm for Cloud Computing. IEEE (2011)
10. S. Sharma, S. Singh, M. Sharma, Performance analysis of load balancing algorithms. World Acad. Sci. Eng. Technol. **38**(3), 269–272 (2008)
11. A Agarwal, S. Jain, Efficient optimal algorithm of task scheduling in cloud computing environment. arXiv preprint arXiv:1404.2076 (2014)
12. K. Mahajan*, A. Makroo*, D. Dahiya, Round robin with server affinity: A VM load balancing algorithm for cloud-based infrastructure, J. Inform. Process. Syst. **9**(3), 379–394 (2013)
13. D. Karaboga, B. Akay, A comparative study of artificial bee colony algorithm. Appl. Math. Comput. **214**(1), 108–132 (2009)
14. A. Soni, G. Vishwakarma, Y.K. Jain, A bee colony based multi-objective load balancing technique for cloud computing environment. Int. J. Comput. Appl. **114**(4) (2015) 0975–8887
15. A. Paulin Florence, V. Shanthi, A load balancing model using firefly algorithm in cloud computing. J. Comput. Sci. **10**(7), 1156–1165 (2014)
16. R. Panwar, B. Mallick, Load balancing in cloud computing using dynamic load management algorithm. 978-1-4673-7910-6/15/$31.00 ©20 15 IEEE
17. V.V. Athani, S. Sandhya, Survey on load balancing techniques in cloud. Int. J. Comput. Appl.. (0975 – 8887) **114**(4) (2015)
18. P. Kumar, R. Kumar, Issues and challenges of load balancing techniques in cloud computing: A survey. ACM Comput. Surv. (CSUR) **51**(6), 1–35 (2019)

Extraction Techniques and Evaluation Measures for Extractive Text Summarisation

Asha Rani Mishra, Mahaveer Singh Naruka, and Shivam Tiwari

1 Introduction

A summary is a "text that is created from one or more texts, provides crucial information in the original text(s), and is no longer than half of the original text(s) and frequently much less" (Radev et al. 2002). Automatic text summarisation is the technique of producing only relevant, correct and precise summaries from a larger size document by retaining only relevant sentences [1]. Automatic text summary minimises the size of an input document to only relevant content that communicates the document's primary concept or context without changing the original information [1]. It has been long time since Luhn proposed the concept of automatic text summarisation (Luhn 1958). Several methods of extraction strategies have been successfully used to retrieve relevant contents as a document summary. In recent years, continuous research is going on this to overcome issues like the lack of coherence, redundancy and lack of flow in summary.

2 Related Work on Extractive Text Summarisation

Text summarisation is the technique of generating only relevant, correct and precise summaries from a larger size document by retaining only relevant sentences [2]. Several methods of extraction strategies have been successfully used in the past to retrieve relevant contents as a document summary. Summarising text can be done based on the number of input documents (single vs multiple), the original text's

A. R. Mishra (✉) · M. S. Naruka · S. Tiwari
GL Bajaj Institute of Technology and Management, Greater Noida, U. P, India
e-mail: asha.mishra@glbitm.ac.in; shivam.tiwari@glbitm.ac

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_17

content type (generic vs query-based), the category of input text (genre-specific vs domain-specific), the purpose (inductive vs informative or critical), the level of linguistic space (shallow vs deep) and the output type (extractive or abstractive) [3]. Table 1 shows broad methods of text summarisation based on some parameters, and Table 2 shows previous commonly used methods of extractive text summarisation.

3 Extractive vs Abstractive Summarisation Method

Extractive and abstractive are widely two approaches for generating accurate automatic summary. Extractive method returns highly informative blocks of text, while abstractive method creates new text which does not exist in that form in the document.

To construct a summary, an extractive-based technique detects key phrases from the input materials. The original sentences are not changed; however, the abstractive method involves employing the linguistic/semantic method to analyse and interpret the input content. Summary produced by abstractive methods is more generalised because generated summary carries more concise information [4]. But sometimes performance of extractive summaries is better as compared to abstractive summaries [5]. This is because abstractive methods exploit the concept of semantic for an information representation; also new information is inferred with the help of natural language generation which is comparatively difficult as compared to sentence extraction for extractive-based methods.

3.1 Steps Followed by Extractive-Based Summarisation Methods

1. Selection of document/text for summarisation
2. Calculation of similarity between sentences present in the document by using similarity measures
3. A numeric value, that is, weight assignment to the sentences based on features [6]
4. Selection of sentences by using suitable ranking method. Higher-ranked sentences are preferred over lower-ranked sentence for summary

The main goal is to extract entities (words, sentences and paragraphs) from the input material for inclusion in the summary. Reranking can be accomplished using similarity measures, which reduce redundancy.

Table 1 Previous study based on different parameters of text summarisation

Study	Type of approach	Merits	Demerits
Nenkova et al. (2011)	Generic	Domain knowledge is not considered while generating summary. It tries to visualise homogeneity in a document.	Generic summaries contain no perspective on the subject and treat the document as a single text; therefore all information is treated equally
Radev et al. (1998), Verma et al. (2007) and Wu et al. (2003)	Domain specific	Domain-specific knowledge plays a key role in sentence selection process	Difficult-to-encode domain-specific knowledge
Nenkova et al. (2011)	Query specific	Information in the generated summary is query specific	Because they are focused on the user's query, they do not provide a precise picture of the document's concepts.
Aliguliyev (2009) and Ko et al. (2008)	Extractive	Extractive summaries contain important sentences selected from the document without any alteration	It carries the risk of causing inconsistency in the text because the selected sentences may not have a semantic relationship with one another, resulting in an incoherent summary
Ganesan et al. (2010) and Khan et al. (2015)	Abstractive	Semantic analysis considers only coherent sentences of the document for combination. It excludes irrelevant sentences from the summary	Longer summaries lack consistency, facts are incorrectly reproduced and created sequences lack novelty, all of which are dependent on profound linguistic skills
S. Gholamrezazadeh et al. (2009)	Indicative	It offers the user a short overview from the actual text by presenting the document's primary topic	So it may not contain all important factual contents
C.T. Shubhangi (2014)	Informative	To the user, communicate the important, clear information from the original text	Difficult-to-find informative sentences for summary
Saggion et al. (2013)	Multiple document	Same topic is discussed in multiple document	Redundancy, sentence ordering, temporal dimension and co-reference are major issues which need to be addressed
Svore et al. (2007)	Single document	A summary is created using a single-source document	Type of text does not have an impact as it does not consider domain-specific knowledge

Table 2 Previous study based on extraction strategies for automatic text summarisation

Study	Type of approach	Concept(s) used	Database used	Merits	Demerits
Nenkova et al. (2005)	Frequency based	Word probability	DUC 2004	Simple	Does not consider uniqueness of the word
Filatova et al. (2004)		TF-IDF		Easy and fast to compute	Redundancy of information is extremely high
Fung et al. (2006)					
Galley (2006)					
Hovy et al. (1998)					
Binwahlan et al. (2009)	Feature based	Sentence length, cue method, etc.		PSO method is used	Does not consider semantics of the sentence
Bossard et al. (2011)					
Suanmali et al. (2009; 2011)			DUC 2002	Fuzzy rules are designed using the features	
Babar et al. (2015)				Uses fuzzy reasoning	
Svore et al. (2007)	Machine learning based	NetSum	CNN articles	Neural network model is used to generate summaries	Effective in learning crucial features that are required to write a summary; however, training corpus varies per language and is not fixed for every document
Hannah et al. (2014)				Used a decision tree	
Kleinberg (1999)	Graph based	HITS (Hyperlink- Induced Topic Search)			Sentence similarity is used to model graph, without "understanding" the relationship between the sentences
Brin et al. (2012)		Google's PageRank (GPR)			
Wan et al. (2006)				Assigned different weights to intra-document links and inter-document links	
Hariharan et al. (2009)				Ranked sentences are non-redundant	
Steinberger et al. (2004)	LSA based	SVD for topic modelling		Clustering of topics for sentence selection	Inefficient representation, due to its distributional nature. A topic may need more than one sentence

3.2 Steps Followed by Abstractive-Based Summarisation Methods

The abstractive-based method creates new text which does not exist in that form in the document:

1. Selection of document for summarisation
2. A numeric value, that is, weight assignment to the sentences based on context
3. Selection of sentences by using suitable semantic method
4. Highly correlated sentences are preferred over lower correlated sentences for summary

Sentences which contain specific context help understand language in abstractive-based methods. For better generalisation, these methods need large volume of data. Sequence-to-sequence models with some natural language generation techniques help produce good abstractive summaries [7].

4 Commonly Used Extraction Strategies for Extractive Text Summarisation

4.1 Frequency-Based Approach

Finding chance of occurrence of word (word probability) and number of occurrences of a term TF-IDF (term frequency-inverse document frequency) are two popular methods for text extraction based on frequency approach.

4.1.1 Word Probability

Counting the word occurrence in the document directly varies to the length of the document, so word probability measures significance of a word. Total count of word occurrence in a document (fc) divided by total count of words (n) in the document is defined as the word probability (P(word)):

$$P\,(\mathrm{word}) = \frac{\mathrm{fc}\,(\mathrm{word})}{n}$$

where word € document(d)

SumBasic [8] uses word probability to create summaries. It selects best scored sentence based on sentence weight.

4.1.2 Term Frequency–Inverse Document Frequency (TF-IDF)

The TF-IDF technique, in which uniqueness is given more importance or relevance to create the feature vector, solves the problem of determining how rare or common a word is [9]. TF-IDF score helps in finding relevant word for summary generation because thematic words are considered as informative sentences and are likely to include in the summary. TF-IDF is TF * IDF [10] where TF is equal to the count of appearance of each word in each document and IDF is the total documents under consideration by total documents containing that specific word. TF-IDF reduces the impact of common occurring words which are more frequent in nature by relating to its corresponding count of occurrences in the document set.

4.2 Feature-Based Approach

One of the approaches for generating summary can be those identifying features which will help to find relevant sentences for summary. According to Edmundson (1969), the following features are commonly used to find relevant sentences for generating summary: location of a sentence in the input document, presence of title word in the sentence and more prompting (cue) words, length of sentence, term weight, proper noun, similarity with the title, similarity with keyword and presence of special characters or words. In this approach, suitable weight is assigned to the features considered. Linear combination of the weights (w_i) and feature's value (f_i) is done for the sentence scoring as below

$$S = \Sigma\,(w_i) \times (f_i)\,, \text{for } i = 1, 2, 3 \ldots\ldots n$$

Popular example of feature-based extractive summarisation is text teaser. A weight is attached to each extracted feature from each sentence. Strength of text teaser is the capability to extract features on both sentence and word levels (Jagadeesh et al. 2005) whereas PyTeaser (Gunawan et al. 2017) uses labelled sentence-level features like length, location and keyword frequency in the sentence.

4.3 Machine Learning Approach

This approach can be suitable for large set of training data consisting of documents along with the summary extract as its label for supervised learning whose objective is to classify a test sentence as a summary (1) or non-summary sentence (0). In training phase, labelled sentences are provided for learning for developing model. In prediction phase, trained model predicts whether a sentence in the new document (test instance) should be included in the summary or not. These is a feature-dependent approach, that is, annotations or labelling should be done to the data to be trained. Popular supervised machine learning algorithms used for text summarisation are Naive Bayes classifier [11–15], decision tree [16, 17], maximum entropy [18], neural networks [19, 20] and support vector machine. On the other hand, unsupervised approaches generate summaries without needing of training data. Hidden Markov model [21] and clustering and deep learning techniques (RBM, autoencoder, convolutional network, RNN) are instances of unsupervised learning technique.

Another technique for summarisation belonging to supervised approaches is conditional random field (CRF), a popular probabilistic model that focuses on machine learning and is used for structured prediction. CRF in [22] is used as a sequence labelling problem to detect the correct features that include the interactions between sentences.

4.4 Graph-Based Approach

Another technique is to represent sentences as vertices in a graph representation of a document. Edges/arcs define interrelatedness between two sentences. A numeric score is assigned to the edges based on similarity or how much there is relatedness between each sentence. Cosine similarity, Jaccard coefficient and Euclidean distance are the popular ones. After graph modelling, ranking algorithms like HITS (Hyperlink-Induced Topic Search) (Kleinberg 1999) and PageRank (Brin and Page 2012) can be used. LexRank (Erkan et al. 2004) and TextRank (Mihalcea et al. 2004) use unsupervised learning to compute score of importance for summary generation. Semantic similarity is often ignored in bag-of-words (BOW) method. Word2vec and word embedding can capture semantic similarity on the word level. Important issues to be solved are polysemy and the correct interpretation of meaning of phrases. Graph-based ranking algorithms perform good or better for good feature matrix.

4.5 Algebraic Method-Based Approach

4.5.1 LSA

LSA (latent semantic analysis) is a vector-based method for finding representation of text based on the semantics of observed words. The main idea in this method is mapping data from higher-dimensional to lower-dimensional space. Mapping should not result in any loss of significant information. For news domain, Gong and Liu [23] used LSA which extracts highly ranked sentences. This was applicable for both single and multi-document summarisations. This approach uses the concept of SVD (singular value decomposition) to find similarities in pairs of text. The first step is to create word-sentence matrix A having dimension $n * m$, where total n words and m sentences are present in a text. Each entry (a_{ij}) of the matrix A is the weight of the word i in sentence j. TF-IDF technique is used for calculation of weights, and if a sentence does not have a word, the weight of that word in the sentence is zero. Applying singular value decomposition (SVD) on the matrix (A) gives result as $A = U\ \Sigma VT$. Matrix U (n*m) represents a matrix showing relation between term and topic with weighted words. Matrix Σ is a diagonal matrix (m * m) where each row "i" corresponds to the weight of a topic "i". Matrix VT is the topic sentence matrix. The matrix $D = \Sigma VT$ reflects importance of a sentence for representation of a topic. The variable "d_{ij}" shows the weight of the topic/word "i" in sentence "j".

4.5.2 NMF (Non-negative Matrix Factorisation)

Dimension reduction techniques using the concept of NMF can be used in text summarisation tasks. Weighting factors both local and global can be used for the value of each entry in word sentence matrix. NMF method has advantages over SVD in terms of non-negativity and sparseness. Following the decomposition step, an NMF-based sentence selection approach is utilised to generate summaries of given documents by taking into account the sentences with the highest weight [24].

4.6 Hybrid Approach

An artificial neural network was used to select summary sentences for a news story based on features such as the paragraph that follows the title, the location of the paragraph and sentence in the source document and the positioning of a paragraph's phrase. During the training phase, phrase length, the amount of meaningful words in the sentence and the number of title words in the sentence were all measured [25]. Fuzzy logic can be applied to text summarising to aid in the extraction of sentences for summary. The values for the retrieved features were submitted to a fuzzy-based inference system in fuzzy logic for summarising to identify the sentences to be included in the summary. Further sentences can be categorised into unimportant, average and important, which were used in selecting sentences for the summary [26]. Lexical chain methods which are used to find semantic relationship between the words can be used to extract sentences for summary [27]. Document clustering can be used for text summarisation where grouping of similar sentences from each cluster can be done based on some score. High scored sentences are included in the summary [28].

Frequency-based and feature-based approaches are simple in nature but lack in semantic analysis. Machine learning approaches need huge training data for proper prediction on unseen data and also appropriate feature representation. Graph-based approaches are good at generating query-specific and domain-specific summaries, but selecting the right relatedness function is a challenge. Latent semantic analysis method captures semantic relation to generate summary but cannot handle problems like polysemy, etc.

5 Evaluation Metrics for Automatic Text Summarisation

Many measures have been introduced so far as to analyse the performance of summary generated in text summarisation which can be categorised into intrinsic and extrinsic methods [27]. Preference to human-generated summary is the main strength of intrinsic methods due to which it mainly focuses on the text quality and content. This method produces coherent and informative summary. Text quality evaluation ensures that the summary produced should be grammatically correct, should be non-redundant, should have referential clarity and should possess coherent structure. The content evaluation approach involves sentence co-selection, which is quantified by measures such as precision, recall and F-score [29] as well as relative utility. Expression for precision, recall and F-score is given below

Precision = total count of sentences present in both summaries to be evaluated and expert written summary/total count of sentences in the summary to be evaluated
Recall = total count of sentences occurring common to summary to be evaluated and expert written summary/total count of sentences in the expert written summary

$$\mathrm{F-Score} = (2 \times \mathrm{Precision} \times \mathrm{Recall}) \,/\, (\mathrm{Precision} + \mathrm{Recall})$$

Co-selection method which involves precision and recall measures may face difficulty while selecting n topmost important sentences since there is difference in the opinion of human experts. To solve this, relative utility can be used where experts assign score based on utility to all sentences in a document which helps in the selection of top sentences.

Various content-based similarity measures which are used in content evaluation for evaluation of summary in intrinsic method are mentioned below:

Cosine Measure [30]

$$\cos{(X, Y)} = \frac{\Sigma X_i Y_i}{\sqrt{\Sigma (X_i)^2} \times \sqrt{\Sigma (Y_i)^2}}$$

Where X and Y represent candidate and reference summary, respectively, which generally uses vector space model (VSM) for finding cosine measure.

Unit Overlap Measure [31]

$$\text{Unit Overlap } (X, Y) = \frac{\| X \cap Y \|}{\| X \| + \| Y \| - \| X \cap Y \|}$$

Where X and Y represent candidate and reference summary, respectively. It finds the similarity on the basis of the number of common words with respect to the number of non-overlapping words.

***N-Gram Co-occurrence Statistics Method: ROUGE* [32]**

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is one of the popular tools for checking the accuracy of generated summary. Document Understanding Conference (DUC) uses ROUGE for performance evaluation. ROUGE compares n-grams between the summary which is to be evaluated and expert-written summaries. ROUGE n-grams, ROUGE-skip bigrams, ROUGE based on LCS, ROUGE-weighted LCS and ROUGE-skip bigrams and unigrams are popular methods used by ROUGE.

ROUGE (n-grams) score of generated summary is:

$$\text{ROUGE}\,(n - \text{grams}) = \frac{\Sigma_{c \in \text{RSS}}\, \Sigma_{\text{gram}_n} \text{count}_{\text{match}}\left(\text{gram}_n\right)}{\Sigma_{c \in \text{RSS}}\, \Sigma_{\text{gram}_n} \text{count}\left(\text{gram}_n\right)}$$

Reference summary set (RSS) = number of creative notes or text to create reference summaries

$\text{count}_{\text{match}}$ (gram_n) = highest number of matched n-grams between generated summary and reference summary

count (gram_n) = total n-grams in the summary whose score needs to be determined

Another measure used for evaluation of summarisation is BLEU (bilingual evaluation understudy) metric. The popular evaluation is by using machine translation. BLEU uses unigrams, bigrams, trigrams, etc. of variable length and weighted average. Both summary evaluations with ROUGE and BLEU metrics should contain high score.

Pyramids

The pyramid method identifies certain content units from the summary (SCUs (semantic content units)) which can be further used for comparing the information content. Key steps are as follows:

1. Identification of similar sentences with deeper assessment to find the related subparts
2. Provide more weightage to the SCUs which appear more in reference summary so that they will appear at the top of the pyramid
3. Comparison of the candidate pyramid formed with an existing pyramid for evaluating the information content common in both candidate and reference summary

The extrinsic method is a task-oriented approach that evaluates performance based on the usefulness of the generated summaries for the topic at hand. Mainly it is used for document classification, IR and QA [27].

6 Conclusion

We attempted to highlight prior work in the field of text summarisation using extractive approaches in this study. The advantages and disadvantages of commonly used extraction methods are highlighted. The generated summary must have the suitable amount of information, adequate flow and coherency across words, minimal redundancy and the right size. Finding appropriate evaluation metrics to analyse performance is equally difficult. With recent advances in artificial intelligence such as machine learning, deep learning and natural language processing (NLP), efficient summaries can be generated for bigger volumes of data with accurate feature representation and better semantic understanding.

References

1. E. Lloret, M. Palomar, *Text Summarisation in Progress: A Literature Review* (Springer, New York, 2012), pp. 1–41
2. V. Gupta, G.S. Lehal, A survey of text summarization extractive techniques. J. Emerg. Technol. Web Intell. **2**(3), 258–268 (2010)
3. A. Khan, N. Salim, A review on abstractive summarization methods. J. Theor. Appl. Inf. Technol. **59**(1), 64–72 (2014)
4. N. Andhale, L.A. Bewoor, An overview of text summarization techniques, in *International Conference on Computing Control and Automation (ICCUBEA)*, (IEEE Xplore, 2016)
5. G. Erkan, D.R. Radev, LexRank: Graph-based lexical centrality as salience in text summarization. J. Artif. Intell. Res. (JAIR) **22**(1), 457–479 (2004)
6. R. Ferreira, L. De Souza Cabral, R.D. Lins, G. Pereira, E. Silva, F. Freitas, G.D.C. Cavalcanti, R. Lima, S.J. Simske, L. Favaro, Assessing sentence scoring techniques for extractive text summarization. Expert Syst. Appl. **40**(14), 5755–5764 (2013)
7. J. Gehring, M. Auli, D. Grangier, D. Yarats, Y. Dauphin, Convolutional sequence to sequence learning (2017), arXiv preprint arXiv:1705.03122
8. L. Vanderwende, H. Suzuki, C. Brockett, A. Nenkova, Beyond SumBasic: Task-focused summarization with sentence simplification and lexical expansion. Inf. Process. Manag. **43**, 1606–1618 (2007)
9. G. Forman, An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. **3**(Mar), 1289–1305 (2003)
10. J. Ramos, Using TF-IDF to determine word relevance in document queries, in Proceedings of the First Instructional Conference on Machine Learning, 2003, pp. 1–4
11. B. Trstenja, S. Mika, D. Donko, KNN with TF-IDF based framework for text categorization, in *24th DAAAM International Symposium on Intelligent Manufacturing and Automation*, (Elsevier, Amsterdam, 2013)
12. J.L. Neto, A.A. Freitas, C.A.A. Kaestner, Automatic text summarization using a machine learning approach, in *Advances in Artificial Intelligence*, ed. by G. Bittencourt, G. L. Ramalho, (Springer), pp. 205–215
13. D.D. Lewis, Naive (Bayes) at forty: The independence assumption in information retrieval, in *Proceedings of Tenth European Conference on Machine Learning*, (Springer, New York, 1998), pp. 4–15
14. J.D.M. Rennie, L. Shih, J. Teevan, D.R. Karger, Tackling the poor assumptions of nave Bayes classifiers, in Proceedings of International Conference on Machine Learning, 2003, pp. 616–623

15. T. Mouratis, S. Kotsiantis, Increasing the accuracy of discriminative of multinomial Bayesian classifier in text classification, in *Proceedings of Fourth International Conference on Computer Sciences and Convergence Information Technology*, (IEEE, 2009), pp. 1246–1251
16. C.Y. Lin, Training a selection function for extraction, in *Proceedings of the Eighth International Conference on Information and Knowledge Management*, (ACM, New York, 1999), pp. 55–62
17. K. Knight, D. Marcu, Statistics-based summarization – step one: Sentence compression, in Proceeding of the Seventeenth National Conference of the American Association for Artificial Intelligence, 2000, pp. 703–710
18. M. Osborne, Using maximum entropy for sentence extraction, in *Proceedings of the AssociaCL-02 Workshop on Automatic Summarization*, (ACL, 2002), pp. 1–8
19. Z. Cao, F. Wei, L. Dong, S. Li, M. Zhou, Ranking with recursive neural networks and its application to multi-document summarization, in *AAAI*, (2015), pp. 2153–2159
20. Z. Cao, F. Wei, S. Li, W. Li, M. Zhou, H. Wang, Learning summary prior representation for extractive summarization. ACL **2**, 829–833 (2015)
21. C. Zhou, S. Li, Research of information extraction algorithm based on hidden Markov model, in *Proceedings of Second International Conference on Information Science and Engineering*, (IEEE, 2010), pp. 1–4
22. D. Shen, J.-T. Sun, H. Li, Q. Yang, Z. Chen, Document summarization using conditional random fields, in *IJCAI*, vol. 7, (2007), pp. 2862–2867
23. Y. Gong, X. Liu, Generic text summarization using relevance measure and latent semantic analysis, in *Proceedings of the 24th Annual International ACM SIGIR Conference 2001on Research and Development in Information Retrieval*, (ACM, 2001), pp. 19–25
24. J.H. Lee, S. Park, C.M. Ahn, D. Kim, Automatic generic document summarization based on non-negative matrix factorization. J. Inf. Process. Manag. **45**(1), 20–34 (2009)
25. K. Kaikhah, Text summarization using neural networks, in *Proceeding of Second Conference on Intelligent System*, vol. 1, (IEEE, 2004), pp. 40–44
26. M.E. Hannah, T. Geetha, S. Mukhe, *Automatic Extractive Text Summarization Based on Fuzzy Logic: A Sentence Oriented Approach* (Springer, 2011), p. 530538
27. A. Nenkova, K. McKeown, Automatic summarization. Found. Trends Inf. Retr. **5**(2–3), 103–233 (2011)
28. D.R. Anjali, L.M.R.J. Lobo, Text summarization using clustering technique. Int. J. Eng. Trends Technol. **4**(8), 3348–3351 (2013)
29. J. Steinberger, K. Jezek, Evaluation measures for text summarization. Comput. Inform. **28**(2), 1001–1028 (2009)
30. G. Manku, A. Jain, A.D. Sharma, Detecting near-duplicates for web crawling, in *Proceedings of the 16th International Conference on World Wide Web*, (ACM Press, New York, 2007), pp. 141–150
31. H. Saggion, D. Radev, S. Teufel, W. Lam, S. Strassel, Developing infrastructure for the evaluation of single and multi-document summarization systems in a cross-lingual environment, in Proceedings of LREC, Las Palmas, 2011, pp. 747–754
32. K. Zhang, H. Xu, J. Tang, J. Li, Keyword extraction using support vector machine, in *Advances in Web-Age Information Management*, (Springer, New York, 2006), pp. 85–96

Encrypted, Compressed, and Embedded Text in Audio WAV File Using LSB Audio Stenography

Anju Gera and Vaibhav Vyas

1 Introduction

With the rapid development of the Internet and communication strategies, the need to protect information from unauthorized access has arisen, and various techniques for information hiding such as cryptography, hashing, authentication, and steganography have been developed and are being developed today. The goal of steganography is to securely communicate in a completely undetectable manner while avoiding drawing attention to the transmission of secret data. The information concealing applications' most critical boundaries are protection, reliability, invisibility, complexity, and data hiding capability; these parameters are mostly linked to each other [14].

This leads to an extensive study of encryption-encryption methods, digital signature, digital holography, authentication, and watermarking methods. Many encryption-encryption algorithms are designed not only for data but also for image, audio, and video files. The most powerful encryption algorithm for securing electronic communications is the advanced encryption standard (AES), which is a symmetric key algorithm, which means that the encryption and encryption keys are identical [1, 6]. Moreover, there is an RSA (Rivest-Shamir-Adleman) algorithm that is a public key encryption, which means that there is a public key used in the encryption and decryption process related to private key encoding [2, 7].

With the development of computer and its use in various spheres of life and work, the subject of information security has acquired a special meaning. A message is covered up within a block audio sign called embed and sound preparation using a stego key that is the same on the side of the transmitter and the receiver. The yield

A. Gera (✉) · V. Vyas
Banasthali University, Jaipur, India
e-mail: vvaibhav@banasthali.in

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_18

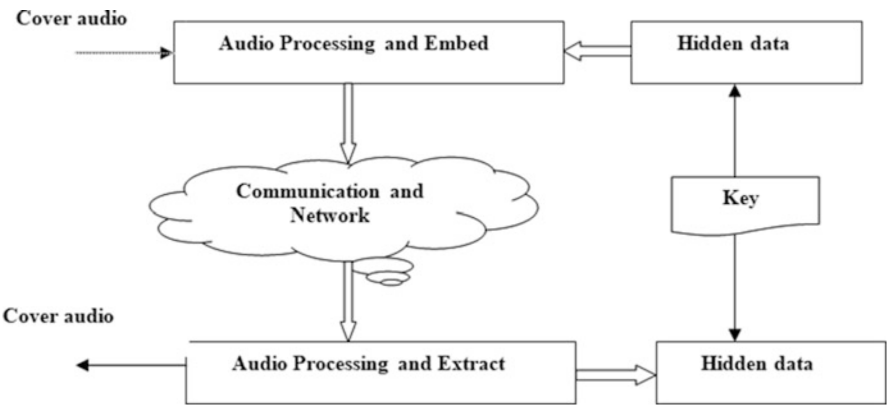


Fig. 1 Steganography workflow

of this output is a stego sound sign [1]. At their receiver side, the inserted text is recovered from the cover sound sign utilizing stego key in the blocks called extract and sound workflow as shown in Fig. 1.

It has been tested in various embedded methods such as AES [1, 2], audio steganography [13, 18, 20, 24, 25, 30], information hiding [3–10], LSB [11–15, 19, 26, 27, 31, 36], and DCT transform and MP3 audio [16, 21, 28, 29, 32, 35] and data compression and Lempel-Ziv [17, 22, 23, 33, 34]. Cryptography and steganography combination improves the productivity and offers a higher degree of security. This research works combined cryptography and steganography to hide a text file in a digital audio signal. Lempel-Ziv-Welch (LZW) algorithm is used for text compression to reduce audio file distortion and least significant bit technique to increase stego file stability.

2 Literature Review

In this section, history of the material is used to construct the audio steganography and integrate encryption methods such as AES and text extraction using Lempel-Ziv encoding.

Cryptography is the modeling and analysis of secrets [1]. The algorithm is widely known. Private key is the secret key. A simple cryptosystem involves encrypting plain text in other text and decrypting the cipher text in plain text. Cipher text is a secret notification message. Encryption is an action that converts plain text (or text) into cipher text (or cryptographic software) [1]. There are two types of digits, symmetric and asymmetric ciphers, in cryptosystems [3, 16].

2.1 Symmetric Cipher

In symmetric cryptographic system or the private cryptographic system, the same key is shared by sender and receiver. In encryption system, the sender identifies the secret message and key. The recipient uses the same key to retrieve the private message [4, 5, 7].

2.1.1 Stream Cipher

It combines a binary message using the exclusive combined action or (XOR) stream with a pseudorandom key stream [1]. The secret key is the initial state of the key stream. The key together with the initial vector is placed in a pseudorandom byte generator, which produces any length of the main key stream. The plain text is also converted to a bit stream. Main-stream combination XOR operation with plain text stream performance of the cipher text stream [30, 37]. Transmitting the encryption requires less time than block encryption to encrypt a document.

2.1.2 Block Cipher

A block cipher is an encryption technique that applies a deterministic formula to decrypt a block of text alongside a symmetric key, as opposed to encrypting the one bit at once as in stream ciphers [6]. For instance, in a typical block cipher, this research work utilizes advanced encryption standard (AES) over another encryption. The overall encryption method is scientifically represented in equation:

$$E\,(K, M) = \{C\}\,K \tag{1}$$

where E denotes the encryption function, M denotes the message, K denotes the key, and C is the cipher text.

2.2 Asymmetric Cipher

An asymmetric cryptographic system or a secret system of public keys consists of a pair of keys [3]. For example, sender needs the receiver's public key to encrypt the concealed message. After the receiver receives the cipher text, the receiver may register the private key with cipher text to retrieve the plain text. These two keys are exceptional and determine from one another computationally infeasible.

2.3 Steganography

The term steganography means keeping knowledge concealed. Digital stereo imaging is more common on the Internet because of its frequency. The most common image stereography algorithm is the least significant bit algorithm. This algorithm uses extra bits in the image. The extra bit function in the image provides much more precise details if needed. This way, a person can exchange a secret message with extra bits without being easily detected.

2.3.1 Audio Steganography

Audio steganography is a way of hiding audio sounds using the weaknesses of the human auditory system. Audio steganography can be much more complex than digital image steganography, which users do not see. Changing the volume can change the sound quality. Successful audio recordings can produce results like the original sound and the ability to detect changes in the person's ear. LSB is a method that hides information in audio technologies.

2.4 Audio Steganography Techniques

The hidden message is sent via a digital signal in audio steganography which causes a change in the collection of appearance files. Furthermore, audio signals are characteristically repetitive and volatile in nature and have better features making them suitable for use as a mask for encrypted communication to conceal secret messages [10, 15].

2.4.1 Low Bit Encoding

For encryption, the least significant bit (LSB) is replaced for the audio file through confidential file information. Even though this technique is basic and can be utilized to send bigger messages, this strategy does not shield the concealed message from minor changes that may happen because of format conversion [8, 9].

2.4.2 Phase Coding

Phase coding [14] is based on the assumption that the phase components are not responsive to a person's development on earth. In a digital signal's phase range, the equipment is encoded with a phase signal.

2.4.3 Echo Hiding

Like SS encryption, echo cancellation provides higher data rates and greater stability than noise generation methods [25]. You need three echo parameters to successfully hide data. The offset is also changed to mask the binary message.

2.5 Data Compression Techniques

Data compression is a technique for reducing the reliability of transmitting data, reducing data storage requirements, and reducing communication costs. The purpose of text compression is to reduce the redundancy of text data, so that data can be stored or transmitted efficiently. Data compression methods are lossless and lossy.

2.5.1 Lossless Compression Methods: Run-Length Encoding (RLE)

This is particularly useful for data that contain many of these renditions, such as simple graphics such as icons, line art, and animation [12]. Since there are two different intensities (black and white), it is probable that the adjacent pixel will be similar.

2.5.2 Huffman Coding

Huffman coding is an algorithm for encoding entropy used to compress Bezel data. It is Huffman who created it. Huffman coding is also used as a "template" for a few other items today [11].

2.5.3 Lempel-Ziv Encoding

Static and dynamic can be a type of dictionary-based encoding. In static dictionary coding, the dictionary remains when it comes to decoding processes [26, 29]. In dynamic dictionary, the dictionary is quickly updated. Algorithms are easy to implement and can be implemented as other devices. LZW (Lempel-Ziv-Welch) compression is built into a wide range of file formats, including GIF, TIFF, and PNG.

At this stage, two synchronized events occur: the construction of the indexed dictionary and the compression of the character strings. The algorithm extracts the smallest subdirectory from the string left without compression which cannot be contained in the dictionary. It then saves a copy of that subdirectory as a new entry in the dictionary and assigns it index rank. Compression happens as the index contained in the dictionary replaces yet another subfolder with the last word. The

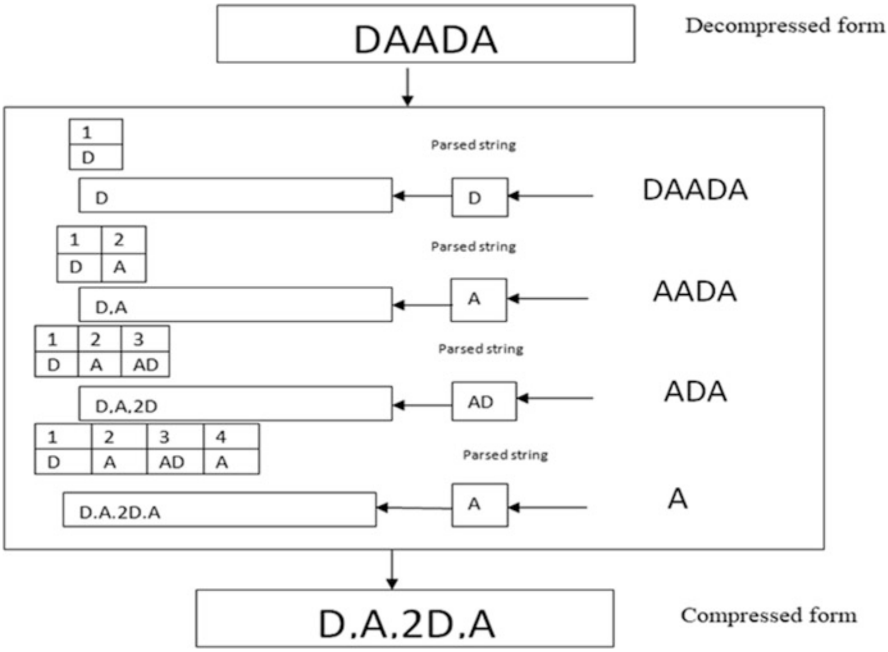


Fig. 2 Lempel-Ziv encoding

method then inserts the substring index and end character into the compressed string (Fig. 2).

2.6 Lossy Compression Methods

The slightest difference our eyes and ears do not show. In such instances, we can use a method for the loss of data compression. In such instances, we can use a method for the loss of data compression. Several methods have been developed using video compression tools such as JPEG (Joint Photographic Experts Group) used for image compression, MPEG (Digital Image Motion) compression for video, and MP3 (MPEG audio) compression for audio.

3 Proposed System

We present a new method for encrypting text from advanced encryption system (AES), data compression (Lempel-Ziv), and least significant bit (LSB) using audio steganography. In fact, this is done for text and audio files. This proposal presents

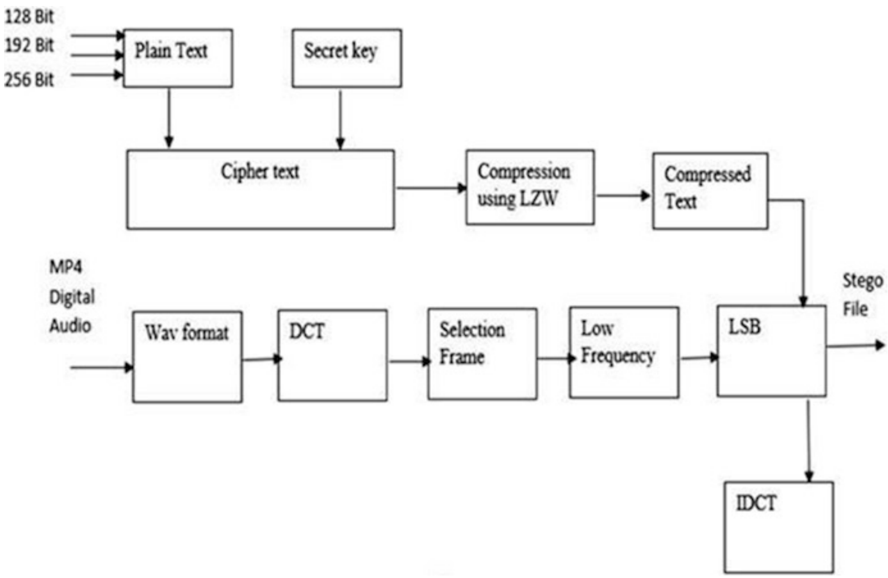


Fig. 3 LSB (least significant bit) is used to conceal the encrypted and compressed secret message (text file) within the digital audio signal

stego files (embedded audio file) that retain their original size after evaluation of the proposed method, and the amount of information that the system can develop and may be removed is too large. It is widely and effectively used when there is an abbreviated system that might be able to shield and share large amounts of confidential data or information without a doubt. A detailed analysis was carried out for this system to show sustainability in Fig. 3.

Lempel-Ziv is denoted by.

$$Z_{\mathrm{LEMP}} = {}^{a}\mathrm{Huff}\left(\log_2 a + \log_2 p + 2\right) \tag{2}$$

where ^{a}HUFF denotes the compressed cipher text using Huffman model, P represents the fixed codes, and Z_{LEMP} represents the LZW compression. By using this system, anyone can make sure that this system is very secure. The algorithm of the proposed method contains two parts as mentioned below. We used this as seen in Fig. 4 to evaluate it on an audio file with an embedded message and to retrieve confidential information.

```
from bitstring import Bits

def lsb_embed (carrier, data, n_bits=1):
  # Assumes that both carrier and data have dtype of int16

  # Convert all integer values of secret message to binary strings:
  secret_bits = []
  For value in np.nditer (data):
    secret_bits.append (np.binary_repr (value, 16))

  # Join all binary strings together
  secret_bits = ''.join (secret_bits)

  # Ensure that the length of binary string is the same as the size of carrier
  secret_bits = secret_bits.ljust (carrier.size * n_bits, '0') [:carrier.size * n_bits]

  # Modify the least significant bits of carrier to contain hidden data
  audio_with_hidden_data = np.zeros (carrier.shape, dtype=carrier.dtype)
  For i in range (len (carrier)):
    # Convert ith value of carrier to binary string:
    binary_string = np.binary_repr (carrier[i], 16)
    # Set the last bit of the binary string to be a bit from the secret message:
    altered_binary = binary_string [:-n_bits] + secret_bits [i*n_bits: i*n_bits+n_bits]
    audio_with_hidden_data[i] = Bits (bin=altered_binary).int # Binary string to int

  return audio_with_hidden_data
```

Pseudo code for embedded data

```
def lsb_retrieve (signal, n_bits=1):

  # Collect the least significant bits of the 'stego' signal
  secret_bits = []
  for value in np.nditer (signal):
    ls_bit = np.binary_repr (value, 16)[-n_bits:]
    secret_bits.append (ls_bit)

  # Join bits together to form a binary string
  secret_bits = ''.join (secret_bits)

  # Ensure that the length of binary string is divisible by 16
  secret_bits = secret_bits [:-( len (secret_bits) % 16)]

  # Convert chunks of 16 consecutive bits to 16 bit integers to retreive the secret
data
  retrieved_audio = np.zeros (len (secret_bits) // 16, dtype=np.int16)
  for i in range (retrieved_audio.size):
    retrieved_audio[i] = Bits (bin=secret_bits [i*16 :( i+1)*16]).int

  return retrieved_audio
```

Pseudo code for retrieves the embedded hidden data

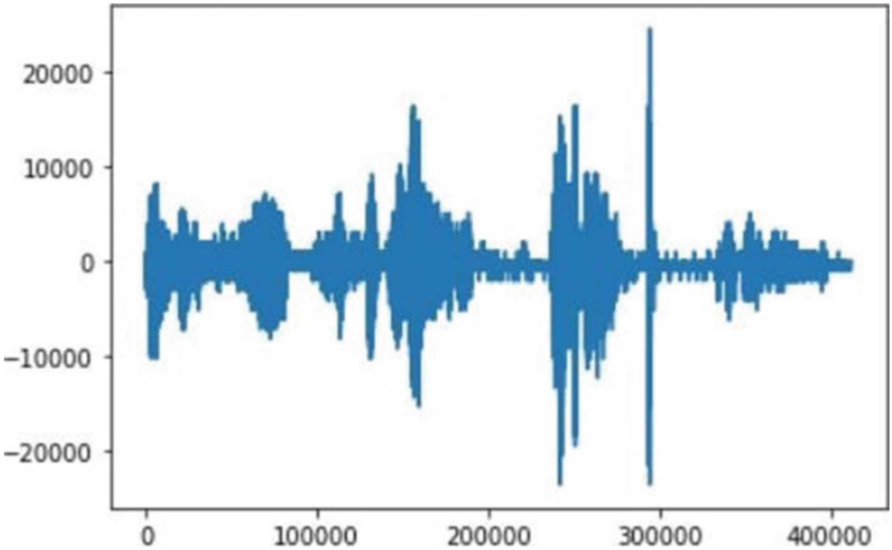


Fig. 4 Cover audio

4 Experimental Work and Results

The framework is designed to run using Python programming language, MP3 and MP4 advanced sound record locations were used in this created system as the distributed media, and various scopes of mystery text were covered up for evaluation within them. Table 1 below shows a summary of different methods of audio steganography in terms of power, vulnerability, and hiding rate. Because of their high embedding capacity, wavelet coefficients and the least significant bit method are ideal for data hiding, whereas other techniques offer low embedding capacity. Lossy data is recovered in the event of a wavelet coefficient. But we cannot detect the initial message so that it can be altered. In LSB coding, the standard of information protection and consistency is preserved throughout the transmission of any sensitive information. So we used LSB method.

Suppose "A.wav" is the sender, and "B.wav" is the hidden message as shown in Table 2 (secret message may be any other form of data as well, including text). Table 3 shows the SNR and MSE results.

The study's goal is to create quantitative metrics that instantly improve audio quality. Table 2 and Figs. 4, 5, and 6 measure and show audio, stego audio, and cover audio output. Using LSB, secret bits are embedded, contributing to increased sound robustness.

Table 1 Comparison of audio steganography methods

Hiding domain	Methods	Strengths	Weakness	Hiding rate
Temporal domain	Least significant bit	Simple and easy way of hiding information with high bit rate	Easy to extract and destroy	16Kbps
Frequency domain	1. Phase coding	Data Retrieval needs the original Signal	Low capacity	333bps
	2. Spread spectrum	Provide better robustness	Vulnerable to time scale modification	20bps
Wavelet domain	Wavelet coefficients	Provide high embedding rate	Lossy data retrieval	200kbps

Table 2 Audio recoding

	Size (KB)	Speed	Duration
A.wav	804.68	1.23 Mb/s	0.6
B.wav	49.2	193Kb/s	0.3

Table 3 Stego audio, cover data, and hidden data

Cover data	Hidden data	Stego	
		MSE	SNR
Dialogue	Welcome	0.43	73.6
Female	Welcome	0.42	73.6

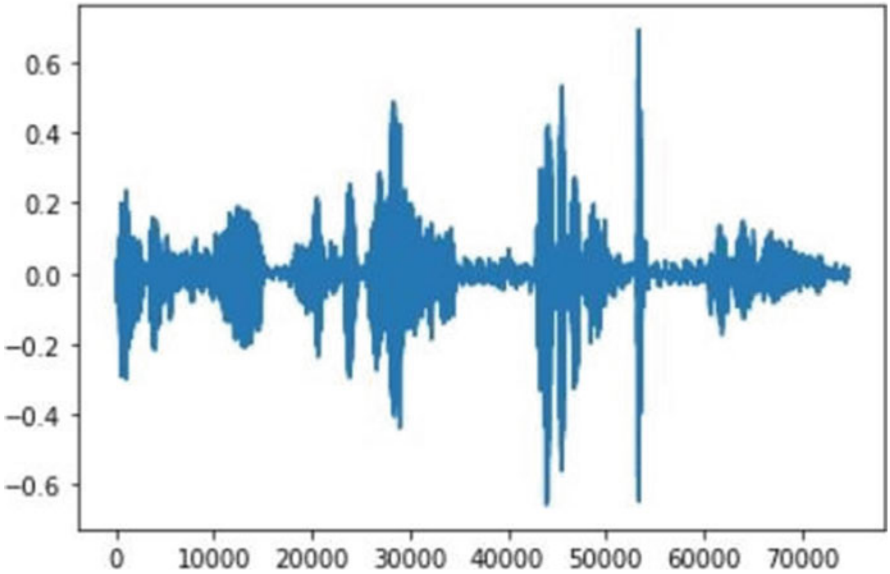


Fig. 5 Stego audio

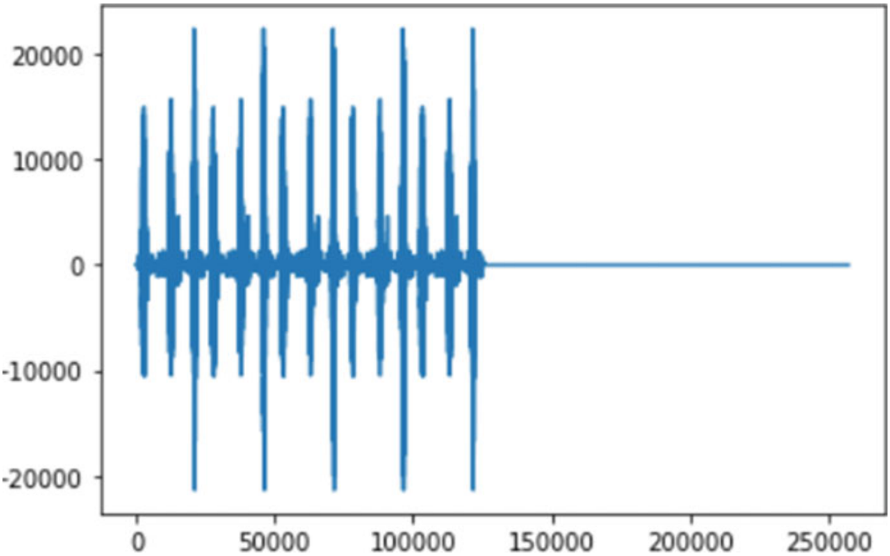


Fig. 6 Retrieved hidden audio

5 Discussion

This study shows how audio steganography can be used to hide encrypted text. In comparison to previous studies, the proposed method increases hiding ability by 25% while keeping stego tolerable. The SNR for audio is 73.6 dB. The new model is 25% more powerful than the previous model. Table 4 describes the description of various schemes.

J. Uthayakumar proposed a highly reliable and low-complexity image compression approach. To further compress the image, the NCS method executes the bit reduction process, which is ultimately encoded using a codec. With high fidelity, the proposed NCS method improves compression performance and lowers sensor node energy consumption. A. Habibian, T. V. Rozendaal, J. Tomczak, and T. Cohen developed a technique for generating a two-dimensional (2D) audio signal as an input to a VVC encoder and investigated its application to 2D audio compression using the video coding scheme. The experimental outcome demonstrates the feasibility of further investigation into 2D audio encoding. Shukla offers a novel audio compression method that includes a lossless text compression technique. By eliminating duplicate data, the goal of audio compression is to minimize the amount of data necessary to represent digital audio. The suggested algorithm's performance is evaluated using a variety of audio specimens of varying sizes and audio signal characteristics. M. Farzaneh proposed a suitable audio graph structure.

(Muin, F.) [10]. There are two goals in this work. The first goal is to examine and evaluate the IEEE 1857.2 standard's performance. The predictor and the preprocessing block are the focus. The predictor's fundamental mechanism is

Table 4 Literature review

Paper name	Author	Year of publication	Results
robust image hiding in audio based on integer wavelet transform and Chaotic maps hopping	El-Khamy *et al.*	2017	Using two phases of integer wavelet transformation, change of wavelet coefficients, XOR, and chaotic map methods, this methodology boosts the concealing capacity, security, and robustness of audio steganography
DWT/DCT-based Invisible Digital Watermarking Scheme for Video Stream	Panyavaraporn and Horkaew	2018	For improved robustness and extraction even after HEVC compression, this paper introduces the DWT-DCT hybrid watermarking
Data Security with the combination of Cryptography and Audio Steganography	Kapoor, Kapil	2019	This paper proposes a combination of the echo hiding process of cryptography and steganography for the hidden conversion of text data through an audio file. Metrics such as PSNR, MSE, and SNR have been used for approach assessment
Evolutionary Detection Accuracy of Secret Data in Audio Steganography for Securing 5G-Enabled Internet of Things	Mohammed J. Alhaddad, Monagi H. Alkinani, Mohammed Salem Atoum, and Alaa Abdulsalm Alarood	2020	This paper provides an experimental analysis assessing the performance of the hidden data embedded in MP3 detection. Training information has been used to incorporate and detect text messages into MP3 formats
LSB Image Steganography with Data Compression Technique Using Goldbach G0 Code Algorithm	Jan Carlo T. Arroyo, Allemar JhoneP. Delima	2020	This study used a compression technique in securing text data using the Goldbach code algorithm. Peak Signal to Noise

linear predictive coding (LPC). The linear predictive encoder's error residue is normalized in the preprocessing step. The second goal is to show the outcomes of experimenting with various wave sound file types as inputs. O. Salau, I. Oluwafemi, K. F. Faleye, and S. Jain introduced a discrete cosine transform (DCT) with a temporal auditory masking (TAM) method for audio signal reduction. The suggested technique achieves a compression ratio (CR) of 4:1 when compared to the original signal, according to the results. B. Patil and K. D. Kulat implemented to find the sampling frequency first preprocessed the audio file and then encoded data in the sample file. After receiving files, RLE and Huffman techniques are applied.

6 Conclusion and Future Work

In this way, we have proposed another audio steganography technique by concealing content by bit substitution within a music text. The statistics also indicate the original audio signal distribution and the resampling of the audio signal after the message is inserted in it. After the hidden message is inserted, the changed speech file appears as the initial carrier file. Consequently, information protection standard is maintained during the transmission of any sensitive information. Future work is to concentrate on how much maximum data can hide in the audio signal and make it stable.

References

1. A.A. Olawale, G. Arome, Embedding text in audio steganography system using advanced encryption standard, text compression and spread spectrum techniques in Mp3 and Mp4 file formats. Int. J. Comput. Appl. **177**(41), 46–51 (2020). https://doi.org/10.5120/ijca2020919914
2. A. Kanhe, G. Aghila, *Robust Audio Steganography Based on Advanced Encryption Standards in Temporal Domain* (IEEE, 2015)
3. F. Djebbar, B. Ayad, H. Hamam, K. Abed-Meraim, A view on latest audio steganography techniques, 2011 International Conference on Innovations in Information Technology, 2011, pp. 409–414. https://doi.org/10.1109/INNOVATIONS.2011.5893859
4. P.P. Balgurgi, S.K. Jagtap, Audio steganography used for secure data transmission, in Proceedings of International Conference on Advances in Computing, 2013, pp. 699–706
5. F. Djebbar, B. Ayad, K.A. Meraim, H. Hamam, Comparative study of digital audio steganography techniques. Eurasip J. Audio, Speech, Music Process. **2012**(1), 1–16 (2012). https://doi.org/10.1186/1687-4722-2012-25
6. R. Jayaram, Anupama, Information hiding using audio steganography – A survey. Int. J. Multimed. Appl. **3**(3), 86–96 (2011). https://doi.org/10.5121/ijma.2011.3308
7. H.A. Abdullah, A.A. Abdulameer, I.F. Hussein, Audio steganography and security by using cryptography. i-manager's J. Inf. Technol. **4**(4), 17–24 (2015). https://doi.org/10.26634/jit.4.4.3644
8. I. Bilal, R. Kumar, M.S. Roj, P.K. Mishra, Recent advancement in audio steganography, in *Proc. 2014 3rd Int. Conf. Parallel, Distrib. Grid Comput. PDGC 2014*, (2015), pp. 402–405. https://doi.org/10.1109/PDGC.2014.7030779

9. A. Neyaz, Audio steganography via cloud services: Integrity analysis of hidden file. Int. J. Cyber-Secur. Digit. Forensics **7**(1), 80–86 (2018). https://doi.org/10.17781/p002368
10. S.V. Jadhav, A.M. Rawate, A new audio steganography with enhanced security based on location selection scheme. Int. J. Sci. Eng. Res. (2016)
11. M. Bazyar, R. Sudhirman, A new method to increase the capacity of audio steganography based on the LSB algorithm. J. Teknologi Sci. Eng. **74**(6), 49–53 (2015)
12. B. Datta, P. Pal, S.K. Bandyopadhyay, *Robust Multi- Layer Audio Steganography* (IEEE, 2015)
13. H.B. Kekre, A. Athawale, S. Rao, U. Athawale, Information hiding in audio signals. Int. J. Comput. Appl. **7**(9), 0975–8887 (2010)
14. J. Bahl, D.R. Ramakishore, Audio steganography using parity method at higher LSB layer as a variant of LSB technique. Int. J. Inno. Res. Comput. Commun. Eng. **3**(7) (2015)
15. K.P. Adhiya, S.A. Patil, Hiding text in audio using LSB based steganograph. International Institute for Science, Technology and Education **2** (2012)
16. A.B. Watson, *Image Compression Using the Discrete Cosine Transform* (NASA Ames Research Center, California, 1994)
17. B. Forouzan, F. Mosharraf, *Data compression* (Foundation of computer Science)
18. M.S. Atoum, O.A.A. Rababah, A.I. Al-Attili, New technique for hiding data in audio files. J. Comput. Sci. **11**(4), 173–177 (2011)
19. P.E. Kresnha, A. Mukaromah, A robust method of encryption and steganography using ElGamal and spread spectrum technique based on MP3 audio file, in Proceeding Conference on Application and of Electromagnetic Technology, 3(9), 11–15 (2014)
20. K. Sakthisudhan, P. Prabu, C.M. Marimuthu, Dual steganography approach for secure data communication, in *ELSEVIER International Conference on Modelling, Optimization and Computing*, (2012)
21. M. Qiao, A.H. Sung, Q. Liu, MP3 audio steganalysis. Inf. Sci. **231**, 123–134 (2013)
22. M.R. Hasan, M.I. Ibrahimy, S.M.A. Motakabber, M.M. Ferdaus, M.N.H. Khan, Comparative data compression techniques and multicompression results, in *IOP Conference*, (2013)
23. C. Lamorahan, B. Pinontoan, N. Nainggolan, Data compression using Shannon-Fano algorithm. JdC **2**(2), 10–17 (2013)
24. P. Yellamma, N. Challa, Performance analysis of different data compression techniques on text file. Int. J. Eng. Res.Tech. (IJERT) **1**(8) (2012)
25. D. Wheeler, D. Johnson, B. Yuan, P. Lutz, *Audio steganography using high frequency noise introduction* (2012)
26. D. Ghanwat, R.S. Rajan, Spread Spectrum based audio steganography in transformation domain. Glob. J. Adv. Eng. Tech. **2**(4) (2013)
27. S.K. Bandyopadhyay, B. Datta, Higher LSB layer based audio steganography technique. IJECT **2**(4), 129–135 (2011)
28. E. Bertino, Data security–challenges and research opportunities, in *Workshop on Secure Data Management*, (Springer, Cham, 2013), pp. 9–13
29. X.U. Shuzheng, Z. Peng, P. Wang, H. Yang, Performance analysis of data hiding in MPEG-4 AAC audio. Tsingua Sci. Tech. **14**(1), 55–61 (2009)
30. P. Jayaram, H.R. Ranganatha, H.S. Anupama, Information hiding using audio steganography––A survey. Int. J. Multimed. Applica. IJMA **3**(3) (2011)
31. R.F. Olanrewaju, H.A. Othman Khalifa, R. Suliman, Increasing the hiding capacity of low-biten coding audio steganography using an ovelembedding technique (2013)
32. Z. Su, G. Zhang, F. Yue, L. Chang, J. Jiang, X. Yao, SNR-constrained heuristics for optimizing the scaling parameter of robust audio watermarking. IEEE Trans. Multimedia **20**(10), 2631–2644 (2018). https://doi.org/10.1109/TMM.2018.2812599
33. N. Pm, R. Manicka Chezian, Behavioral study of data structures on Lempel Ziv Welch (LZW) data compression algorithm and its computational complexity, in *Proc. – 2014 International Conference on Intelligent Computing Applications ICICA*, (2014), pp. 268–274. https://doi.org/10.1109/ICICA.2014.64
34. N. Zhang, T. Tao, R.V. Satya, A. Mukherjee, Modified LZW algorithm for efficient compressed text retrieval, in *Proceedings of the International Conference on Information Technology: Coding and Computing-2004*, (IEEE, New Jersey, 2004)

35. R. Tanwar, K. Singh, M. Zamani, A. Verma, P. Kumar, An optimized approach for secure data transmission using spread Spectrum audio steganography, chaos theory, and social impact theory optimizer. Int. J. Comput. Netw. Commun. (2019). https://doi.org/10.1155/2019/5124364
36. A. Thakur, Enhanced steganographic method using LSB, Parity and spread spectrum technique for Audio Signals. October 2014. https://doi.org/10.13140/RG.2.2.16950.91203 (2016)
37. S. William, Cryptography and network security: For VTU, Pearson Education India (2006)

Improvement of Flat Plate Collector Performance Using Nano-additives

N. Keerthi Kumar, G. H. Raghunandan, and Raveendra Gudodagi

1 Introduction

Till date, major source for satisfying huge demand of energy worldwide is fossil fuel due to its high-energy conversion efficiency than alternative energy source. On the other hand, the exhaustive use of fossil fuel resources led to negative impact on environment. This made many researchers to find better alternative for fossil fuel such as biodiesel, solar, and wind. Biodiesel produced from plant seeds is considered as better alternative for diesel fuel in engines, but processing of seed oil to biodiesel incurred much cost and also energy obtained is not in line with fossil fuel source. As solar energy is cleaner and freely available, it has been considered as a best alternative for conventional fossil fuels in household applications like water heating system and drying system especially in countries where footprints of fossil fuel are not available and have hot climates. Stationary water heating system with absorber plates have better absorption capacity of solar energy than other heating systems. Percentage of energy extraction from heating system mainly depends on types of fluid used, materials used for tubes and absorber plates, geometry of the system, and nature of hybrid technique involved. Stationary water heating systems, flat plat collector is cheapest due to their design and fewer parts' involvement. In spite of the position of the sun, efficiency of these collectors depends on various factors like tilt angle of absorber, material of absorber and tube, mass flow rate, type of fluid used, and coating material used on absorber plate and tubes. Enhancement in efficiency of existing solar collectors can be done using advanced fluids having

N. Keerthi Kumar (✉) · G. H. Raghunandan
BMS Institute of Technology and Management, Bangalore, Karnataka, India
e-mail: keerthikumarn@bmsit.in; raghunandangh@bmsit.in

R. Gudodagi
Reva University, Bangalore, Karnataka, India
e-mail: raveendra@revainstitution.org

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_19

higher thermal conductivity than conventional working fluids like water and EG. Mixing nanoparticles into the water or EG is termed as nanofluid, and this enhances thermal conductivity. The use of nanoparticles directly into the conventional fluids enhances the thermal performances; on the other hand, stability of suspended particles also plays a major role in increasing the efficiency of system. Yousefi et al. used Triton X-100 surfactant to mix Al_2O_3 nanoparticles into conventional base fluids. Tests were carried with different mixing ratios and different mass flow rates using flat plate collectors to evaluate the efficiency of collector system. An increasing efficiency of 28.3% was recorded when surfactant was used, whereas without surfactants, efficiency enhancement was 15.6%. Effect of collector was studied using various flow rates and 0.1% of ZnO nano particles in weight fraction. Increased efficiency was recorded at lower flow rates by using nano fluids compared to water [2].

Along with type of nanofluid, percentage of nanofluid, and mass flow rate, Brownian motion and Lewis number are also important parameters for enhancing the efficiency of solar flat plate collector (Bhatti et al.). Enhancement in Prandtl number is mainly due to decrease in temperature profile [3].

From the detailed literature, it is reviled that most of the work was concentrated on enhancement of collector efficiency with reduced temperature parameter of $(T_{c1} - T_a)/G_T$ [4–8].

Till date, no study has been carried to examine the enactment of solar flat plate collector using Al_2O_3 and ZnO with EG altogether. Hence, the present study is aimed to prepare the experimental setup in the first phase, and in the second phase, test fluids were prepared in-house. In the subsequent phase, experiments were carried out to study the effect of flat plate collector operated using nanofluids with EG at sunny days. Results obtained from this study demonstrated only for the application of indirect heating, use of nano particles provides faster heating with higher efficiency.

Excluding introduction, this paper is organized under four different sections: methodology, results, discussion, and conclusions.

2 Methodology

The performance of commercially available flat plate collectors in terms of its efficiency was evaluated using six different working fluids which also include distilled water. Initially, experimental setup was installed, and the solar flat plate collectors are tilted at an angle of 30 ° with horizontal plane. Two nanoparticles, namely, Al_2O_3 and ZnO, and EG were collected from the Vasa Scientific, Bangalore.

Solar water heater is designed and constructed. In the first phase, base fluid (distilled water) was filled, and the air present in the flow line is removed by vacuum breakers. Base fluid is allowed to flow throughout the test setup to find any leakages at the point of connections and also to ensure proper working of various equipments like flow meter, pump, flow control valve, thermocouples, and data logger. Thermocouples are placed in various places of test setup to record the temperatures of inlet and exit fluid, absorber plate temperature, and atmospheric temperature. Temperatures from various locations were measured using T-type thermocouple and are measured in Kelvin [9]. Solar radiation values during test period were collected from pyranometer at Berambadi site, Gundlupet, Karnataka (11.76°N; 76.58°E), at an altitude of 870 m. Initially, tests were carried out to find the efficiency of solar flat plate collector between 9 ante meridiem and 3 post meridiem for base fluid, and results were in line with results of [10]. Temperatures from the test setup are collected at a time interval of 30 min. Every experiment was carried five times each time a day, and the best combination of results was taken for results. In the second phase, other test fluids such as water, combination of Al_2O_3 and base fluid, and ZnO and base fluid including and without including EG were at various proportions, and different mass flow rates are used to evaluate the performance of flat plate collector. Based on the obtained results, different graphs are plotted.

2.1 Experimental Setup

Technical specification and photo of the test setup are shown in Fig. 1 and Table 1, respectively. Absorber plate, riser tubes, and header tubes are made up of copper. Storage tank is made up of stainless steel with capacity of 60 l, and it is covered using foam to provide thermal insulation. For the effective circulation of working

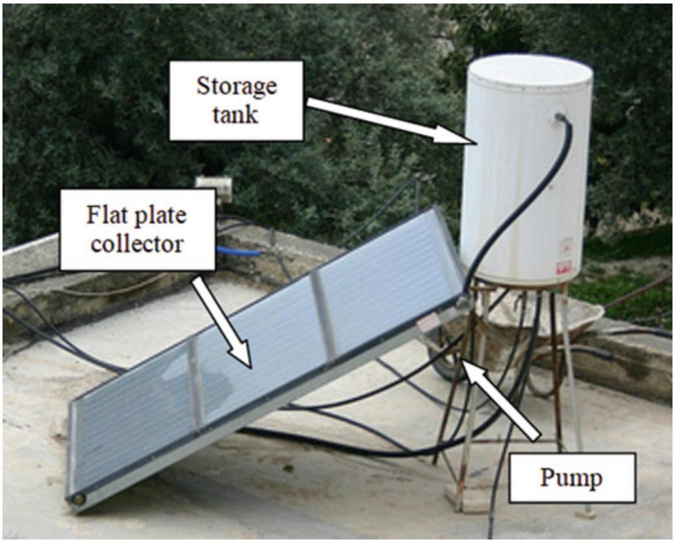


Fig. 1 Experimental setup

Table 1 Specification of experimental setup

Particulars	Dimensions
Solar collector	
L, m	1.72
W, m	0.93
T, m	0.08
Absorber plate	
L, m	1.7
W, m	0.9
T, m	0.05
Glass cover thickness, m	0.05
Insulation thickness, m	0.04
Diameter of riser tubes, m	0.019
Thickness of fin, m	0.03

Table 2 Properties of nanoparticles [11, 12]

Particulars	Al_2O_3	ZnO
Density, g/cm^3	5.61	3.94
Average size of nanoparticles, nm	30–50	35
Purity	99.99	99.99
Surface area, m^2/g	120–140	31
Molecular weight, g/mol	101.96	81.37
Morphology	Spherical	Polyhedral roughly round
Thermal conductivity, W/m°C	40	29

fluid, pump with flow rate capacity of 1.35 kg/s is needed. Rotometer is used to measure the actual flow rate within experimental system [4]. Experimental setup was installed at Berambadi site, Gundlupet, Karnataka (11.76°N; 76.58°E), at an altitude of 870 m. Flat plate collector was tilted at an angle of 30°.

2.2 *Test Fluid Preparation*

To study the efficiency of collector, experiments were performed using different test fluids. Before conducting the experiments, different test fluids were prepared using EG and nanoparticles. Nanoparticles of Al_2O_3 and ZnO were selected and mixed at 0.25% volume fraction. Experiments were carried by considering EG along with nanoparticles. Al_2O_3, ZnO, and EG were collected from Vasa Scientific. Properties of nanoparticles are shown in Table 2.

Six different test fluids were prepared including base fluid (distilled water). EG is mixed with distilled water at a ratio of 25:75. Nanoparticles are mixed at 0.25% based on volume fraction. To obtain proper dispersion of nanoparticles, nanofluids were mixed using ultra sonicator at a frequency of 50 kHz for 30 min.

The volume concentration and heat capacity of dispersed fluid are represented by Eqs. 1 and 2, respectively [13, 14]:

$$\varnothing_{\mathrm{v}} = \frac{1}{\left[{}^{100}\big/_{\varnothing_{\mathrm{m}}}\right]\left[{}^{\rho_{\mathrm{n}}}\big/_{\rho_{\mathrm{f}}}\right] + 1} \times 100\,(\%) \tag{1}$$

where $\varnothing_{\mathrm{v}}$, $\varnothing_{\mathrm{m}}$, ρ_{n}, and ρ_{f} are volume and mass concentration, nanoparticle density, and fluid density:

$$c_{\mathrm{p,nf}} = \varnothing_{\mathrm{v}} C_{\mathrm{p,n}} + [1 - \varnothing_{\mathrm{v}}]\, c_{\mathrm{p,f}} \tag{2}$$

2.3 Energy Analysis

Thermal performance in terms of efficiency of flat plate collector is evaluated as per the ASHRAE 93-2010 standards.

To evaluate the performance of collectors, the effect of solar rays striking the collector surface (total absorbed radiation) based on hitting angle and area of collectors was calculated excluding refracted solar radiation and considering collector losses.

Heat removal factor (Fr) was calculated from ratio of output and inlet temperature of fluid. To find the total system efficiency, Fr value with total absorbed energy is used.

Total energy absorbed (Q) is obtained by inlet (T_{c1}) and outlet (T_{c2}) temperature of fluids along with flow rate and heat capacity of nanofluid using the formula below [15]:

$$Q = \dot{m} \times C_{\mathrm{pn,f}} \times [T_{\mathrm{c2}} - T_{\mathrm{c1}}] \tag{3}$$

Collector efficiency is a function of total energy absorbed, total incident radiation, and the area of the collector [15]. Total collector efficiency is calculated using the formula below:

$$\eta = \frac{Q}{A \times G_{\mathrm{T}}} \tag{4}$$

Useful energy is obtained by Eq. 5. It is the difference between energy absorbed and energy lost.

$$Q = A F_{\mathrm{R}} \left[G_{\mathrm{T}}\,(\tau\alpha) - U_{\mathrm{L}}\,(T_{\mathrm{c1}} - T_{\mathrm{a}})\right] \tag{5}$$

The total thermal efficiency is calculated by Eq. 6 [15]:

$$\eta = F_R(\tau\alpha) - \left[F_R U_L \frac{T_{c1} - T_a}{G_T}\right] \quad (6)$$

3 Results

3.1 Effect of Nanofluid Without Ethylene Glycol on Efficiency and Reduced Temperature Parameter

Figures 2, 3 and 4 represent the effect of nanofluid without ethylene glycol on efficiency and reduced temperature parameter. Graphs were plotted for water, Al_2O_3, and ZnO at flow rate of 0.05, 0.07, and 0.09 kg/s. The present study was carried to determine the influence of nanoparticles, and the results were compared with base fluid. Efficiency of the test setup was calculated similarly for all working fluids. Based on results, efficiency of base fluid mixed with Al_2O_3 nanoparticles was highest than other two fluids. This may be attributed to highest thermal conductivity of Al_2O_3 than ZnO and base fluid. Highest efficiency for base fluid is 64.36% at mass flow rate of 0.09 kg/s. At the same flow rate, the use of Al_2O_3 enhanced the efficiency by 11.15%, whereas the use of ZnO nanoparticles enhanced the efficiency by 5.81% (Fig. 4).

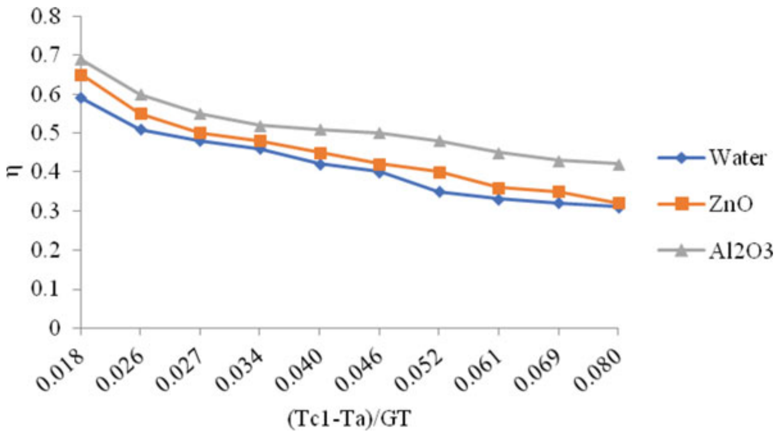


Fig. 2 Collector efficiency with reduced temperature parameter without EG with flow rate of 0.05 kg/s

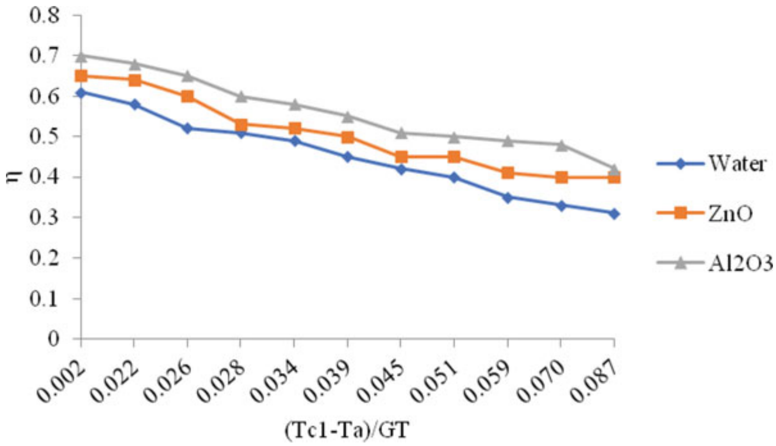


Fig. 3 Collector efficiency with reduced temperature parameter without EG with flow rate of 0.07 kg/s

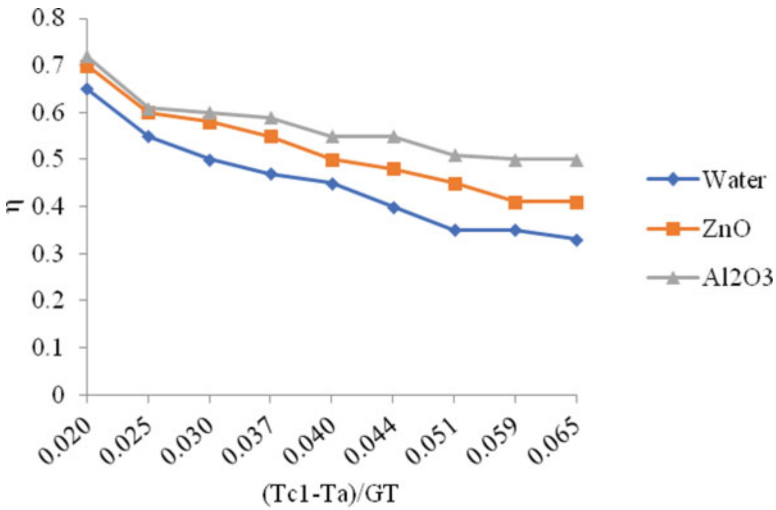


Fig. 4 Collector efficiency with reduced temperature parameter without EG with flow rate of 0.09 kg/s

3.2 Effect of Nanofluid with Ethylene Glycol on Efficiency and Reduced Temperature Parameter

Figures 5, 6, and 7 show the efficiency of test setup for selected working fluids with EG. All the test fluids were mixed with EG, and the test fluids were called as water/EG, Al_2O_3-water/EG, and ZnO-water/EG, respectively. Based on the results, Al_2O_3-water/EG shows highest efficiency than ZnO-water/EG. Results revealed that

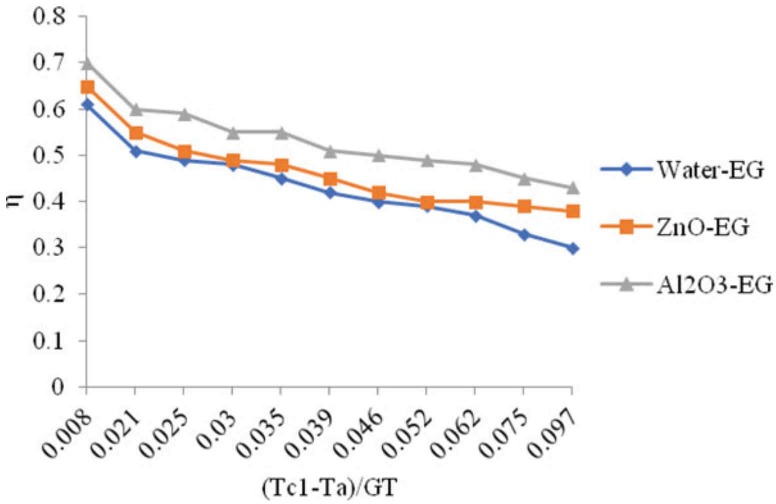


Fig. 5 Collector efficiency with reduced temperature parameter with EG with flow rate of 0.05 kg/s

the influence of EG is much greater on efficiency for all the flow rates and all the test fuels. The use of EG for base fluid increases efficiency by 8.55%, 10.29%, and 13.35%, respectively, for mass flow rate of 0.05, 0.07, and 0.09 kg/s. Further increase in efficiency is obtained by using nanoparticles. This may be attributed as for base fluid mixed with EG and nanoparticles, energy parameter and FRUL were lower whereas absorbed energy parameter $F_R(\tau\alpha)$ was highest than the base fluid with EG and nanoparticles. Hence, the efficiency of solar collector using EG and nanofluids was higher than working with base fluid [4].

By using EG with water at mass flow rate of 0.05 kg/s, 0.07 kg/s, and 0.09 kg/s, the efficiency is enhanced by 1.32%, 1.62%, and 1.78%. The efficiency of Al_2O_3 base fluid/EG at the same mass flow rate is enhanced by 1.84%, 2.34%, and 3.97%, respectively, when compared with Al_2O_3 base fluid. Further, the efficiency of ZnO with EG was enhanced by 1.54%, 2.03%, and 2.86%, respectively, when compared with ZnO water without EG.

3.3 Influence of the Mass Flow Rate on Efficiency for Different Test Fluids

Figure 8 represents effect of mass flow rate on efficiency for different test fluids. Efficiency of collector was increased for higher mass flow rates for all the test fluids. Similar trend in the graph was observed for all test samples in three various mass flow rates. It is evident that mass flow rate is directly proportional with Reynolds number, and hence, it may be attributed that as the Reynolds number increases, the

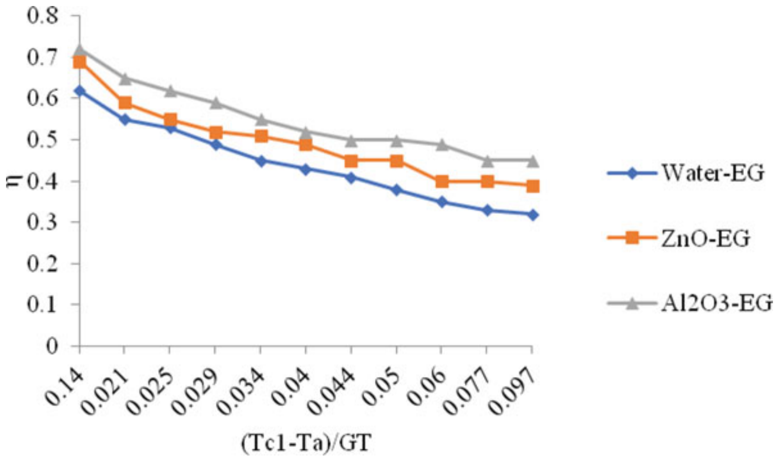


Fig. 6 Collector efficiency with reduced temperature parameter with EG with flow rate of 0.07 kg/s

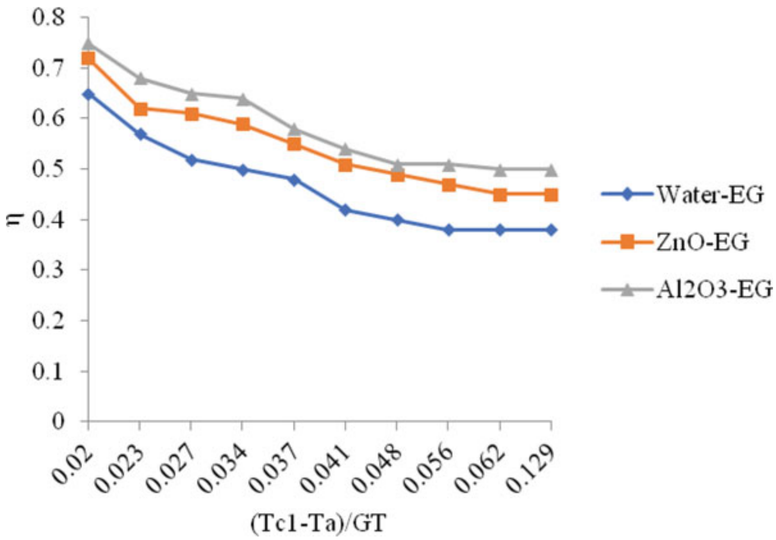


Fig. 7 Collector efficiency with reduced temperature parameter with EG with flow rate of 0.09 kg/s

efficiency of the test setup also increases [1]. Highest efficiency obtained for Al_2O_3-water/EG at mass flow rate of 0.09 kg/s is 77% when compared with base water; enhancement in efficiency for Al_2O_3-water/EG is 10.41%. As nanofluids mixed with base fluid have higher thermal conductivity, heat absorbing ability is higher [16].

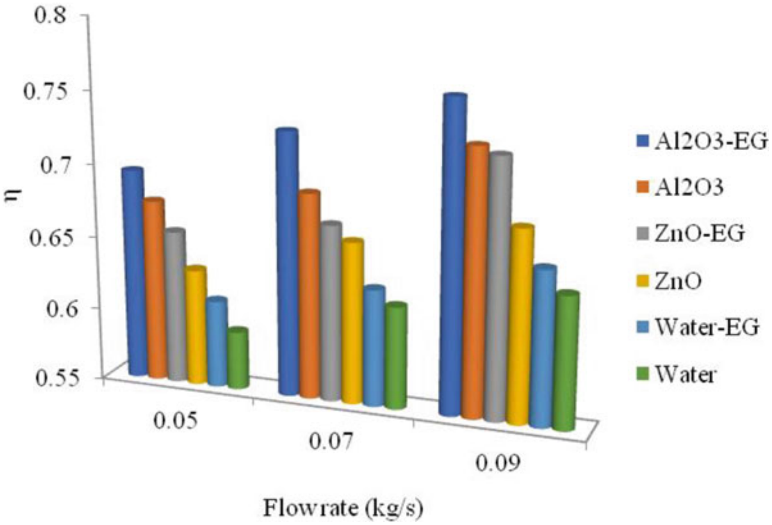


Fig. 8 Variations of efficiency for test fluids with mass flow rate

4 Conclusions

Present experiment was carried to evaluate the effect of nanofluid and EG on performance of collectors. Experiments were performed at three different mass flow rates, namely, 0.05 kg/s, 0.07 kg/s, and 0.09 kg/s. Test fluids were prepared on volume basis and the different test fluids with and without EG in the ratio of 75:25. Efficiency of test setup was calculated based on ASHRAE 93-2010 standard.

Key outcomes of the present study are summarized as follows:

- As mass flow rate increased, collector efficiency also increased
- Use of nanofluid increased the acquired energy parameter test setup
- Use of EG further increased the efficiency of all the test fluids
- Addition of Al_2O_3 and EG for water resulted in increased efficiency compared to other fluids for all the flow rates. In addition 5.3% increased efficiency is obtained by mixing Al_2O_3 and EG for water compared to mixing Al_2O_3 alone for water at a flow rate of 0.09 kg/s
- Efficiency if collector is enhanced by 4% by using ZnO in water for flow rate of 0.09 kg/s compared to water at same flow rate

References

1. T. Yousefi, F. Veysi, E. Shojaeizadeh, S. Zinadini, An experimental investigation on the effect of Al_2O_3-H_2O Nanofluid on the efficiency of flat-plate solar collectors. Renew. Energy **39**, 293–298 (2012)

2. S.A. Babu, M. Raja, An experimental investigation on the effect of ZnO/water nanofluid on the efficiency of flat-plate solar collector. Adv. Nat. Appl. Sci. **7**, 40–48 (2016)
3. M.M. Bhatti, T. Abba, M.M. Rashidi, Entropy generation as a practical tool of optimization for non-newtonian nanofluid flow through a permeable stretching using SLM. J. Comput. Des. Eng. **4**, 21–28 (2017)
4. S.T.J. Suthahar, S. Jaisankar, S. Saravanan, Experimental investigation on solar flat plate collector using alumina nanofluid with tube inserts. Mater. Technol. **37**(3), 179–189 (2022)
5. O. Montoya-Marquez, J.J. Flores-Prieto, The effect of inclination on the efficiency in a medium-temperature flat plat solar collector. Energies **10**, 71 (2017)
6. A. Noghrehabadi, E. Hajidavalloo, M. Moravej, Experimental investigation of efficiency of square flat-plate solar collector using SiO2/water nanofluid. Case Study. Therm. Eng. **8**, 78–386 (2016)
7. M. Sheikholeslami et al., Recent progress on flat plate solar collectors and photovoltaic systems in the presence of nanofluid: A review. J. Clean. Prod. **293**, 126119 (2021)
8. Q. He, S. Zeng, S. Wang, Experimental investigation on the efficiency of flat-plate solar collectors with nanofluids. Appl. Therm. Eng. **88**, 165–171 (2015)
9. PASCO, PASPORT Temperature Sensor—PS-2125: PASCO. Available online: www.pasco.com/prodCatalog/PS/PS-2125_pasport-temperature-sensor/index.cfm. Accessed 11 Mar 2018
10. J.J. Michael, S. Iniyan, Performance of copper oxide/water nanofluid in a flat plate solar water heater under natural and forced circulations. Energ. Conver. Manage. **95**, 160–169 (2015)
11. M. Millipore, ZnO—SDS report. online: www.merckmillipore.com/search/1314-13-2. Accessed 22 May 2018
12. M. Millipore, Al_2O_3—SDS report. online:www.merckmillipore.com/search/1344-28-21. Accessed 22 May 2018
13. A.M. Ajeena, P. Víg, I. Farkas, A comprehensive analysis of nanofluids and their practical applications for flat plate solar collectors: Fundamentals, thermophysical properties, stability, and difficulties. Energy Rep. **8**, 4461–4490 (2022)
14. K.M. Pandey, R. Chaurasiya, A review on analysis and development of solar flat plate collector. Renew. Sust. Energ. Rev. **67**, 641–650 (2017). ISSN 1364-0321
15. American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc. ASHRAE 93-2010 (RA 2014), Methods of testing to determine the thermal performance of solar collectors; ASHRAE: Atlanta (2014)
16. L.S. Syam et al., Solar energy absorbed thermosyphon flat plate collector analysis using Cu/H2O nanofluid–An experimental study. Energy Clim. Change **2**, 100028 (2021)

Green Computing: A Sustainable and Eco-friendly Approach for Conservation of Energy (A Contribution to Save Environment)

Kanchan Awasthi and Sameer Awasthi

1 Introduction

Green computing is one of the developing registering innovations in the field of computer engineering to give green information innovation. It is for the most part used to safeguard climate, enhance energy utilization and keep green climate. Expanding energy productivity and lessening the utilization of unsafe materials are the fundamental objectives of green figuring. Green computing eventually centres around routes in lessening by and large natural effects. The significant power utilization parts are processors and the principal memory in the servers. Green processing is the idea which is attempting to bind this strategy by designing new techniques that would work productively while consuming less energy and making less populace.

Green computing is the naturally dependable utilization of PCs and related assets; such practices incorporate the execution of energy-proficient focal handling units, servers and peripherals as well as decreased asset utilization and appropriate removal of electronic waste (e-squander). It is also defined as the process of planning, fabricating/designing, involving and discarding figuring devices in such a way that their natural effect is reduced. Principal objectives of green registering are to diminish the utilization of harmful and peril materials and further develop the energy effectiveness, reusing processing plant squander. Green computing is the prerequisite to save the energy with the costs.

In green computing, PCs have adverse consequences to our current circumstance as well as clients. The misuse of equipment materials contains synthetic compounds

K. Awasthi
Maharishi University of Information Technology, Lucknow, India

S. Awasthi (✉)
Babu Banarasi Das Northern India Institute of Technology, Lucknow, India

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_20

which are poisonous. A PC left for such countless hours will deliver high electronic expense, and it produces carbon dioxide in the environment. Yet, in present days, an ever increasing number of PCs are creating superfluous ozone-harming substances each year which can cause an Earth-wide temperature boost, so a product designer should utilize the PC in an eco-accommodating way. So it is important to find out about green computing. Green figuring is planning, assembling, utilizing and discarding PCs, server and related subsystems productively with insignificant effect on the climate [4].

Over the most recent couple of years, there has been a ton of publicity about green computing and about giving greener figuring answers for nearly everything. Green computing is another methodology which targets planning PC frameworks that accomplishes better handling and execution with least measure of force utilization. Many survey sand studies have already shown that the power costs put together form the lion's share of total costs of management of a data centre. Till now green computing measures had been restricted to mainly achieve lower power consumption within the organizational framework. Tiwari, S. had explained the importance of green computing beyond reducing power consumption and towards delivering better customer value and achieving the ultimate goal of sustainable IT growth [10].

2 History of Green Computing

One of the first manifestations of the green innovation movement was Energy Star programme, which was founded in 1992. Energy Star was a voluntary label awarded to computing products that reduced their energy use while enhancing their efficiency. The Energy Star programme applies to computer monitors, television sets and thermal management devices such as refrigeration systems, air conditioning units and other similar items. One of the first green computing results was the sleep mode function of computer monitors, which puts a consumer's electronic devices into standby mode when no user activity is detected after a specific duration. As the concept of green computing evolved, it expanded to include thin client solutions, energy standard costing, virtualization procedures, e-waste and other topics [3].

Tactical instrumentalists are currently one of the most well-known green computing ensembles. Green computing ideas are used and used by this group primarily to save money rather than to protect the environment. Businesses are under pressure to maximize resources in order to compete effectively in the market, and this green computing concept arose naturally as a result. Rather than political pressure, this movement was sparked by economic sentiments.

New and emerging technologies have social and environmental consequences, which strategic leaders consider. This particular approach considers additional variables such as marketing and branding in addition to lowering costs. Strategic leaders, unlike tactical instrumentalists, acknowledge the necessity to revamp some existing policies or the organization's fundamental structure. This can be observed in recent efforts to make IT staff directly responsible for energy management, minimization and efficiency.

3 Current Trends in Green Computing

Green computing alludes to the investigation of planning, designing, legitimate utilization and removal of registering gadgets so that it diminishes their effect on the climate.

(a) *Energy Consumption*

As indicated by ecological insurance organization, around 30–40% of PCs are continued during the ends of the week and, surprisingly, after available time, and around 90% of these PCs stay inactive. Assuming we foster any application in green figuring climate, it will utilize ideal actual assets. The Climate Savers Computing Initiative (CSCI) is a work to decrease the electric power utilization of PCs in dynamic and latent states. Another strategy is energy–proficient coding which means diminishing the product utilization of the equipment. More effective calculation will promptly lessen the quantity of assets expected to finish a specific registering capacity.

(b) *Cloud Computing*

Organizations are quickly moving from conventional framework to cloud-based framework as a result of its quicker increase/downsize limit, pay-per-use and admittance to cloud-based administrations without purchasing and overseeing on-premises foundation. The compensation per-use office of cloud foundation gives energy and asset efficiencies and elevates clients to consume just those assets which are fundamental (Fig. 1).

(c) *E-waste Recycling*

Many developed nations are more grounded in innovation; consequently an immense measure of PC frameworks and related items are disposed of each and every day. These items are sold out to other emerging nations. E-waste is a reasonable piece of the waste stream, and reusing e-waste is simple to embrace. Reusing registering hardware like lead also, mercury empowers to supplant hardware that in any case would have been made. The reusing pattern in green computing will in general continue to register hardware materials like lead and mercury out of landfills, and reusing old PC parts to fix or redesign other PC frameworks will save energy, diminishing the age of e-waste influencing the climate [5] (Fig. 2).

(d) *Virtualization*

The significant pattern in green figuring these days is virtualization. It is the method involved with running various PC frameworks on one bunch of actual equipment. Energy productivity can be accomplished with less actual gear connected, which decreases power and consumes less power. Many organizations and open-source projects presently offer programming bundles

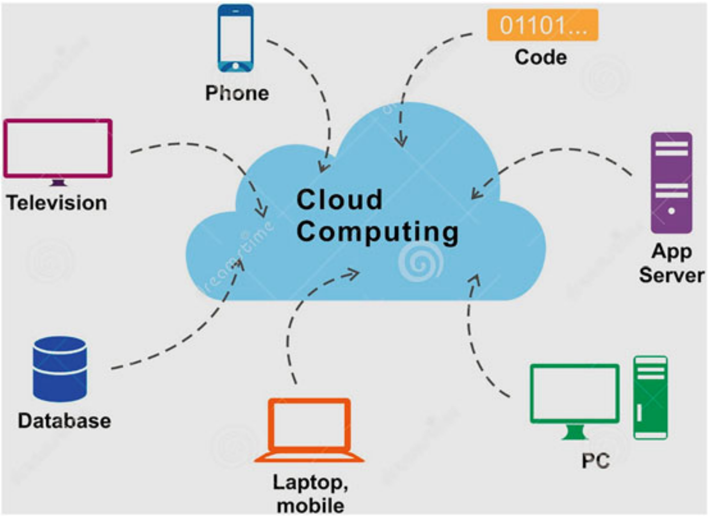


Fig. 1 Cloud computing

to empower a progress to virtual computing. Virtualization can be grouped into three classes, specifically storage, server and desktop virtualization (Fig. 3).

(e) *Developing a Green Machine*

Power management feature in PC sets aside energy and cash. To make the PC climate agreeable, Sleep and Hibernate settings can be utilized. These capacities can be actuated either physically or by powering the board settings of working framework.

(f) *IT product and Eco-labelling*

An extra pattern towards green processing is the "eco-label" which has been upheld by a few associations all over the planet. These eco-marks are given by associations to IT items in view of a few variables planning to save the climate. It likewise incorporates the capacity of reusing the framework, commotion energy utilization and so forth (Fig. 4).

(g) *Modularity*

It expands the adaptability of the organizations which want to foster server farms that are presently anxious to adopt better secluded strategy, adding applications and individual parts as additional necessity. It diminishes the capital cost; however, it likewise builds the adaptability which further assists with forestalling shortcomings. Conversely, secluded item structures may likewise forfeit item execution or lead to over-planned items. These after-effects can increment environmental effect [9].

Green computing benefits include the following:

Fig. 2 E-waste recycling

- Reduced energy use from green figuring methods translates to lower CO_2 emissions
- Saving assets indicates that goods will be created, used and discarded with less energy
- Set money aside through conserving energy and assets
- Green computing even includes a shift in government policy to encourage reuse and reducing energy consumption by individuals and enterprises
- Reduce the risks present in the workplace, such as substances known to cause sickness, nerve damage and people’s safe responses

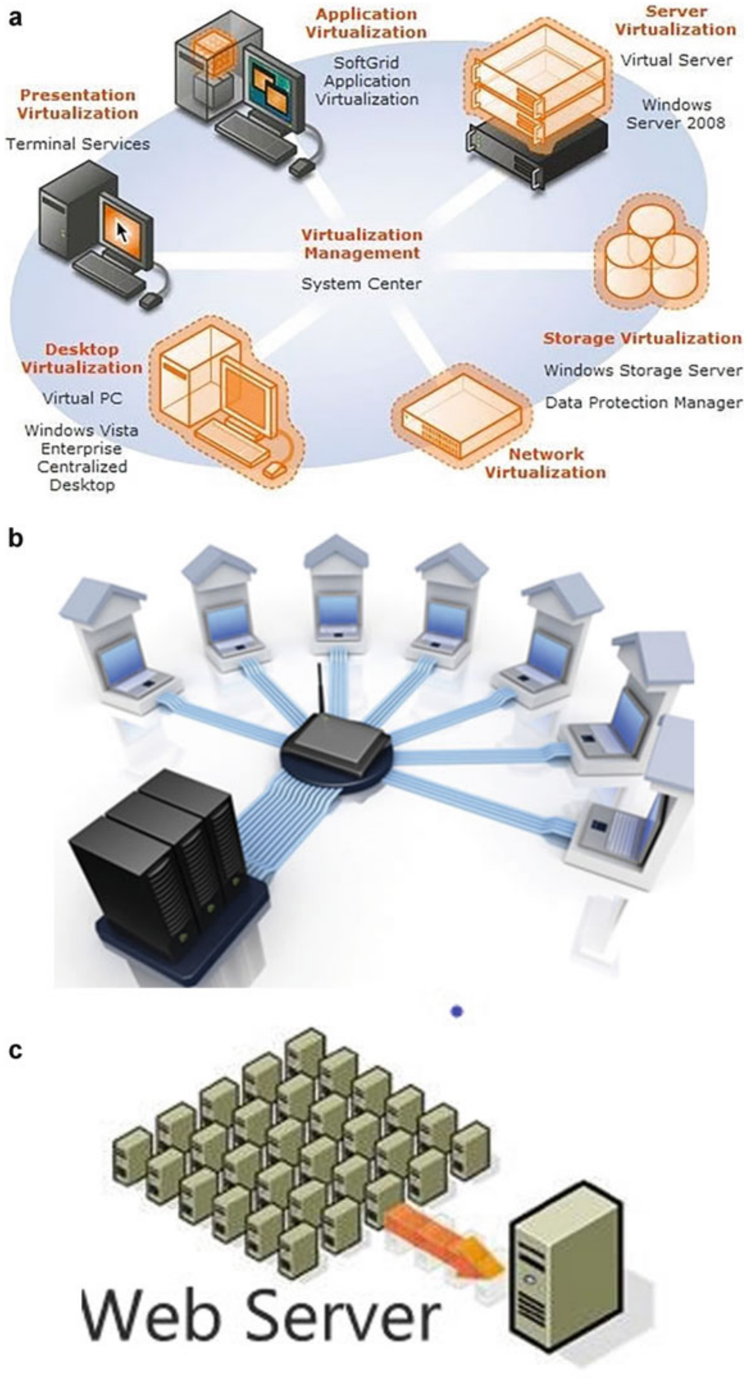


Fig. 3 (**a**–**c**) Virtualization

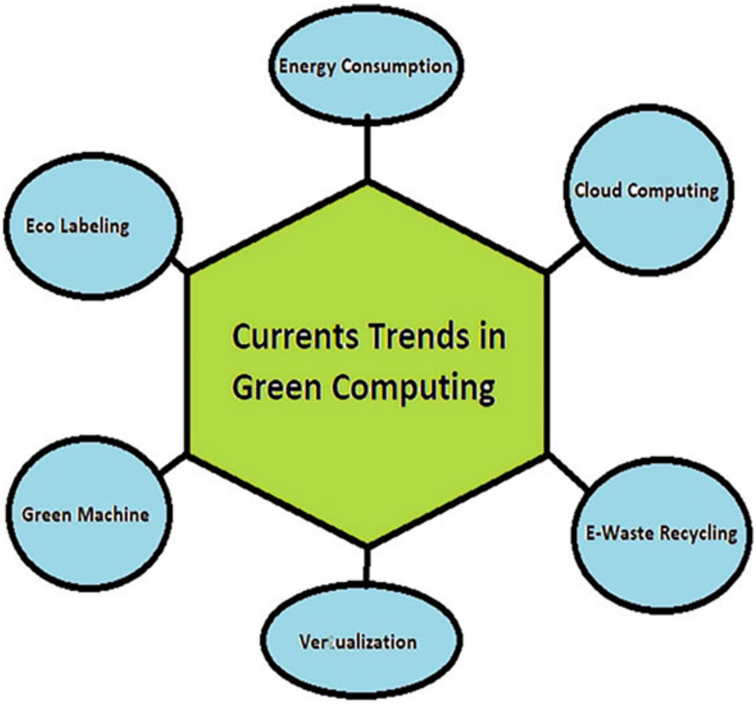


Fig. 4 IT product and eco-labeling

4 Effect of Green Computing in Shaping Education

Green computing has impressive importance in our daily life. Saving energy utilization, decreasing fossil fuel by-products, reducing working expenses and guaranteeing a practical climate have turned into the first concern of the present current world. The green computing area is an exceptionally wide assortment of exploration exercises that have been completed in IT and instructive foundations throughout the course of recent years. In instructive establishments and in IT organizations, electronic gadgets like PCs, climate control systems and lights consume a decent measure of power too. Saving energy assists with reducing expenses and assists with consolidating green computing [8] (Fig. 5).

5 Initiatives Worldwide

Governments have initiated many energy management programmes. In 1998, the China National Development and Reform Commission (NDRC) established the Green computing effect China Energy Conservation Program (CECP), a not-for-

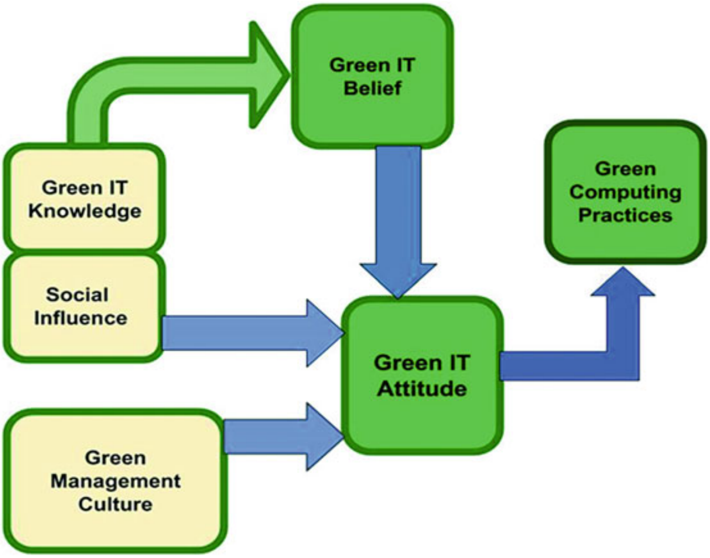


Fig. 5 Green computing effect

profit association responsible for the organization, the executives and execution of the certification for energy rationing, water-saving and use of items harmless to the ecosystem. The European Union adopted the Restriction of Hazardous Substances Directive (RoHS) in February 2003. Six hazardous materials have been restricted for the manufacture of various types of electronic and electrical equipment.

6 Need for Green Computing

In the long run, conserving energy and time saves money, and the benefits are noticed on both macro and micro levels. Green innovations can be applied to an entire company or a single individual's workstation. This innovation is gainful as it:

- Lessens energy utilization of registering assets during top activity
- Saves energy during inactive activity
- Uses eco-accommodating wellsprings of energy
- Diminishes destructive impacts of computing resources
- Diminishes computing wastes

7 How Does Green Computing Conserve Power

Green computing limits the energy utilization of the association, for example, limits the power bill. Utilization of non-poisonous material in the supplies makes the labourer protected from medical issue and word-related perils. It saves the asset of the country in general.

(a) *Practices of Green Computing*

There is a lot of discussion about environmental responsibility these days: global warming, recycling, energy conservation, hybrid cars, buying environmentally friendly products, reducing waste and so on. Simultaneously, a major field of debate on the subject of "green computing" has emerged:

1. Buy laptops, monitors, desktops, printers and other household items with the "Energy Star" label. When not in use, the "Energy Star" gadgets should be designed to "power down" to a low power state. This saves energy and allows them to operate cooler, allowing them to last longer [6].
2. When not in use, computers should always be in "sleep" mode. The EPA (Environmental Protection Agency) has recorded that this reduces their electricity use by 65–75% [7].
3. Switching off PCs and other equipment when not in use is far and away superior. Despite the debate over whether it is better to leave your computer on or turn it off, the truth is that it is better for the environment to turn it off.
4. E-cycle is a company that recycles computer equipment. Reduce, reuse and recycle as much as possible. Reduce, reuse and recycle has long been recommended as a best practice for businesses and/or individual consumers seeking to reduce waste [2].

(b) *Approaches to Green Computing*

For the advancement of green registering, four methodologies are used:

1. *Green Use*

   Lessening the power use of PCs and their fringe subsystems and involving them in an eco-accommodating way. Additionally taking on virtualization lessening the need of energy.
2. *Green Disposal*

   Reusing and reusing existing hardware, appropriately arranging the squandered IT/registering materials, electronic gear and so forth.
3. *Green Design*

   Planning energy proficient as well as viable frameworks which negligibly affect the green climate.
4. *Green Manufacturing*

   Fabricating biodegradable materials, additionally producing long usable, recyclable items and lessening wastage during manufacturing process.

The below figure represents the approaches of green computing (Fig. 6):

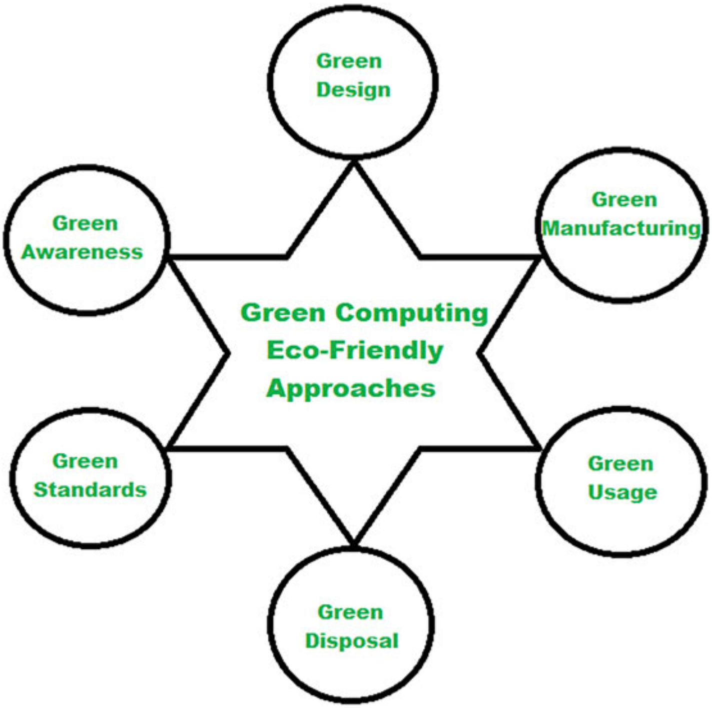


Fig. 6 Green computing approaches

8 Sort of Green Computing

- Solar power system
- Wind turbine programme
- Geothermal power

- *Solar Power System*

  The sun-based power framework is a programme in which we employ daylight to generate sun-oriented force for personal and commercial purposes. California, Spain and Canada are in the best position to succeed in this endeavour. When it comes to green charging, sunlight-based chargers are the most impressive model because they effectively convert electrical power into electrical energy (Fig. 7).
- *Wind Turbine Programme*

  The wind turbine system is the next and most notable sort of green registering, as it lets anybody to create power. There has been no detrimental impact on the climate as a result of the implementation of the breeze turbine programme. It aids in the reduction of CO_2 levels. It is not feasible for everyone to set up a wind turbine programme since it is pricey.

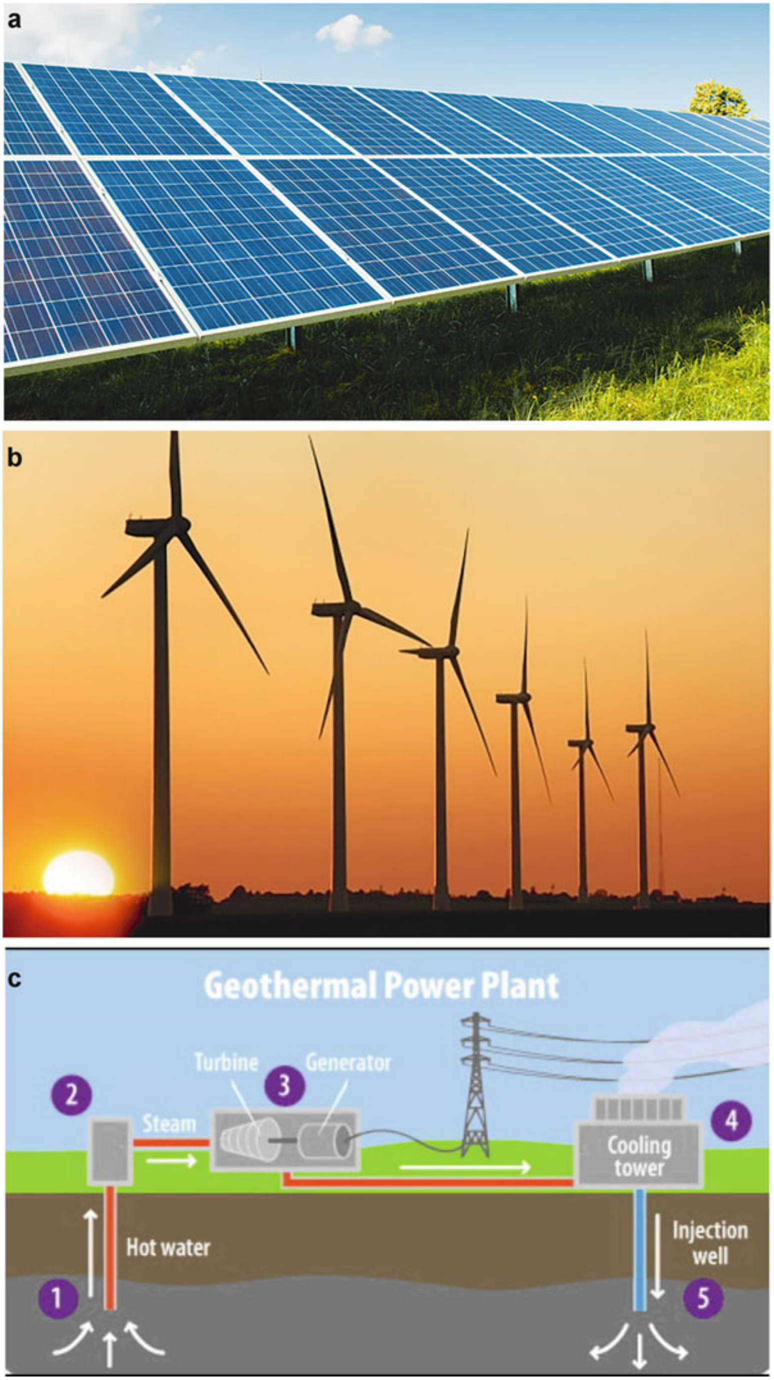


Fig. 7 (**a**) Solar power system. (**b**) Wind turbine. (**c**) Geothermal power

- *Geothermal Power*
  This is moreover a select kind of green innovation. Geothermal plants might produce power that can be used in day-to-day uses like warming and cooling house.

9 Challenges in the Field of Green Computing

The challenges that green figuring is currently facing are listed below:

(a) Return on investment (ROI) is one of the most important factors to consider. The main problem was instilling knowledge in the stakeholders
(b) Performance, energy and temperature aware computing: new optimization techniques
(c) Electronic waste removal
(d) The Indian scenario from a different perspective
(e) It uses a lot of energy
(f) An increase in the amount of energy required

(a) *Return of the investment*: Informing the accomplices about the environmental impact of PCs was the most difficult part. The advantages of a greening project are usually seen over time. As a result, demonstrating rapid returns after applying sustainable IT in the PC community was a major goal in this project [1].
(b) *Performance, energy and temperature aware computing*: new optimization techniques. Because of the remarkable advancement in figuring motion and the growing concern for energy security, energy productivity has become a mechanical issue of paramount importance in PCs. A negotiated settlement between effectiveness, energy and heating rate must be reached in order to obtain the most extreme benefits. It is critical to plan strategies that are optimal in terms of implementation and operation, energy and temperature, taking into account all factors [11].
(c) *Removal of electronic wastes*: Energy yield in PCs became a mechanical problem of paramount importance as a result of remarkable advancements in figuring movement and the growing concern for energy security. To obtain the most extreme benefits, a balance between effectiveness, energy and heating rate must be formed. It is critical to devise tactics that are optimal in terms of execution, energy and temperature, taking all factors into account (Fig. 8).
(d) *Perspective with respect to Indian scenario*: In India, the "green computing" does limit and is defying what is happening as a direct result of various monetary issues and those are associated with bemoaned determined to pull India into the standard improvement of "green computing". The absence of a fundamental investigation drive and a friendly structure has resulted in a halt in the issuance of good licences and the production of locally collected goods.

Fig. 8 Removal of electronic wastes

(e) *Power consuming*: In the field, organizations are looking for a period of high-capability IC chips. They provide superior displays without consuming a lot of energy, so this cycle is not required. It will take a massive attempt, a considerable amount of time and a large lot of extremely qualified people to reach and achieve this goal.
(f) *Increase in energy requirements*: To complete their tasks, some people require or favour any use of potent CPUs. Regardless, these demands necessarily require a substantial amount of computation, making green PCs with equivalent specifications prohibitively expensive (Fig. 9).

10 Conclusion

Green computing has nothing to do with going out and buying biodegradable packaging. As of now, an open door has appeared to consider the effective use

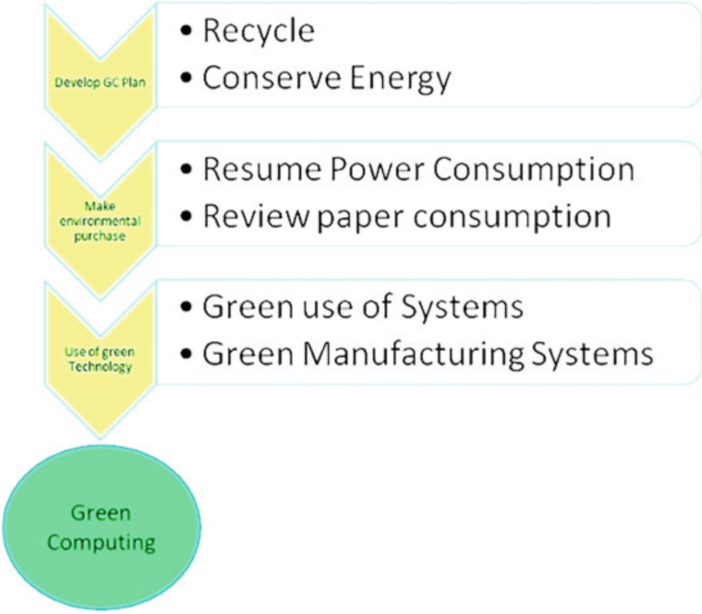


Fig. 9 Power consumption

of PCs and non-unlimited resources. It provides a new avenue for new company visionaries to collect E-squander material and discard PCs. Green computing minimizes an organization's energy consumption, that is, reduces electricity costs and ensures that equipment is made of safe materials that are safe for workers' health. Overall, it protects the country's asset. These green gears will be less expensive in the long run, with virtually no hidden waste costs and improved asset utilization, with no negative impact on precision, execution or lifespan. As a result, it is concluded that appropriate legislation/regulations, user education/awareness and recycling are simple solutions for reducing energy consumption while also minimizing environmental waste. By establishing an appropriate ICT squander of the board body and providing low duty motivating forces to ventures that do not practice green figuring, the government can combat the unreliable removal of PC equipment that requires biodegradability.

References

1. P. Rani, Rekha, D. Singhal, Challenges to the sustainable green computing. Int. J. Adv. Res. Sci. Eng. **2**(2) (2013). ISSN-2319-8354(E)
2. P. Kiruthiga, T. Vinoth Kumar, Green computing – an eco friendly approach for energy efficiency and minimizing e-waste. Int. J. Adv. Res. Comput. Commun. Eng. **3**(5), 356–359 (2014)
3. History of green computing, kamipunyaassignment.blogspot.com

4. N. Vanthana, Green computing. Int. J. Eng. Res. Technol. (IJERT) **6**(7) (2018). ISSN: 2278-0181
5. Climate Savers Computing Initiative (2010). Retrieved http://www.climatesaverscomputing.org
6. Energy Star Program (2010), Retrieved from http://www.energystar.gov/
7. S. Ruth, *Green IT More Than a Three Percent Solution?* (IEEE Internet Computing, Piscataway, 2009)
8. R.K. Modanval et al., Impact of green computing in shaping education, in *Green Computing in Smart Cities: Simulation and Techniques. Green Energy and Technology*, ed. by B. Balusamy, N. Chilamkurti, S. Kadry, (Springer, Cham, 2021). https://doi.org/10.1007/978-3-030-48141-4_9
9. C.M. Rose, A. Stevels, K. Ishii, A new approach to End-of-Life Design Advisor (ELDA), ‖ IEEE International Symposium on Electronics and the Environment, San Francisco (2000)
10. S. Tiwari, Need of green computing measures for Indian IT industry. J. Energy Technol. Policy **1**(4) (2011). ISSN 2224-3232 (Paper), ISSN 2225-0573 (Online)
11. Microsoft: Green IT taking the first step (2010). Retrieved from http://www.microsoft.com/environment/our_commitment/articles/green_guide.aspx

Pattern Recognition Using Graph Edit Distance

Shri Prakash Dwivedi and Ravi Shankar Singh

1 Introduction

The graph is a fundamental and ubiquitous mathematical structure in mathematics, engineering, and computer science. Its strength lies in the flexibility to represent itself as a structural model in the various domains of science and engineering. Graph edit distance (GED) can be stated as the least number of modifications needed to convert one graph into another. The edit operations can add, delete, and substitute nodes and edges. The edit operations were initially used as string edit operations for converting one string into another using the lowest count of edit operations, such as addition, deletion, or substitution of alphabets. Later on, the idea of string edit distance was applied to tree edit distance and further generalized to GED. Due to its flexibility, the edit distance approach is powerful, and it can be applied to different problems.

One of the significant applications of GED is inexact graph matching. In graph matching (GM), we measure the similarity between two objects represented in the form of graphs. It is mainly categorized into two classes, exact and inexact GM. In exact GM, a strict correspondence must be there between the vertices and edges of the two graphs. In the error-tolerant GM, also known as inexact GM, some flexibility or tolerance is allowed during the matching between the two graphs. The limitation of exact GM is that it can be only used to find the strict matching between the two graphs, and therefore, it cannot take into account distortion incurred to the graph

S. P. Dwivedi (✉)
Department of Information Technology, G.B. Pant University of Agriculture & Technology, Pantnagar, India
e-mail: shriprakashdwivedi@gbpuat-tech.ac.in

R. S. Singh
Department of Computer Science & Engineering, Indian Institute of Technology (BHU), Varanasi, India

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4_21

due to the existence of noise in the process of matching. Error-tolerant GM offers flexibility during the process of GM [2]. There are many approaches to error-tolerant GM but GED being very adaptable is one of the most crucial techniques for the GM problem. A comprehensive review of diverse GM methods is described in [4] and [14].

In [15], the authors introduced an approach to present patterns by trees instead of strings. As a tree is a more general high-dimensional structure than string, its use will lead to an efficient description of a higher-dimensional pattern. Then tree system representation of patterns is used for syntactic pattern recognition. A distance measure for attributed relational graph [29] using the computation of the least number of modifications required to change an input graph to the output one is described in [27]. This chapter also considers the costs of recognizing the vertices in the distance computation. In [3], the authors proposed using heuristic information inexact GM of attributed graphs derived from a state-space search. The matching process is generalized to arbitrary graphs, and the edit cost functions are designed so that the GED satisfies the properties of metrics in some situations.

GED is applicable in a broad range of applications as it allows specific edit cost functions to be defined for various applications. A significant limitation of GED is that its computation becomes too costly. It uses exponentially ample execution time to compute GED concerning the number of vertices in the input graph. Graph edit distance problem is shown to be NP-hard in [30]. Since a fast deterministic algorithm is unavailable for this problem, many approximate and suboptimal algorithms have been proposed recently.

The paper [17] proposed an efficient algorithm for GED computation of attributed planar graphs by iteratively matching small subgraphs to optimize structural correspondence. Then it applied the above technique to the fingerprint classification problem. In [19], the authors describe the fast suboptimal algorithm for the GED by reducing the space essential for computing the GED using A^* search technique[16]. They describe the different variants of A^*, such as A^*-beamsearch and A^*-pathlength, to reduce the search space that may not be pertinent to particular classification tasks. The computation of approximate GED by considering only local instead of global edge structure through the optimizing process is given in [23]. In [28], the authors represent the GED as a basis in the label space and use it to define a class of GED cost. They also describe the various characteristics and use of this GED cost. An improvement over the above method by manipulating the initial assignment of the approximation algorithm so that the assignment is ordered based on the individual confidence is provided in [13]. Estimating exact GED considering lower and upper bounds of bipartite approximation utilizing regression analysis is described in [25]. Different search strategies for improving the approximation of bipartite GED computation using the beam search, iterative search, and greedy search, etc., are provided by Riesen and Bunke [24]. The book [21] provides a detailed description of structural pattern recognition and describes various algorithms for structural pattern recognition by using GED.

A novel category of structural pattern recognition using GED is recently proposed [5] that decreases the graph size, reducing search space by ignoring the less relevant vertices using some measure of importance. In [8], the authors presented homeomorphic GED for topologically equivalent graphs and used to perform GM by measuring the structural similarity between two graphs. The proposed technique utilizes the path contraction to remove the vertices having degree two to construct simple paths of input graphs in which every node except first and last has degree two. An extension to GED utilizing the notion of node contraction in which a graph is changed into another by contracting the lesser degree vertices is given in [9]. In [12], the authors proposed centrality GED to perform inexact GM using the various centrality measures for removing the vertices having the least centrality value in the graph. Some other recent works are given in [6, 7, 10, 11, 26].

This chapter is outlined as follows. Section 2 introduces basic concepts and definitions related to GED. Section 3 presents essential algorithms and techniques to compute GED. Section 4 shows some experimental results, and at last, Sect. 5 includes the conclusion.

2 Basic Concepts and Definitions

A description of fundamental concepts and definitions associated with GED is provided in this section. To get an in-depth description, the reader can refer to the texts such as [1, 5, 18].

In computer science, a graph is commonly defined as $G = (V, E)$, where V is the vertex set and E is the edge set in which every edge connects the two vertices. To define a graph in pattern recognition field, we use two additional parameters, node label mapping and edge label mapping, to identify the vertices and links in a graph.

We define a *graph* G as a tuple $G = (V, E, \mu, \nu)$; here, V and E are defined as above, $\mu : V \to L_V$, and $\nu : E \to L_E$. Here, μ is a function that assigns each vertex $v \in V$ a unique label $l_v \in L_V$. Similarly, ν is a function that assigns each vertex $e \in E$ a unique label $l_e \in L_E$.

A graph may be directed or undirected based on its edges; if $\nu(u, v) = \nu(v, u)$, then the graph is undirected since from both directions, edges have the same value, whereas for undirected graphs, $\nu(u, v) \neq \nu(v, u)$. When $L_V = L_E = \epsilon$, i.e., vertex label and edge label sets are empty, the graph G is called an unlabeled graph.

A graph can be converted into another graph using a set of edit operations. A set of edit operations are inserting, deleting, and substituting nodes and edges. A chain of edit operations that convert an input graph into the output one is defined as an *edit path* from the input graph to the output one. To insert a node u, we denote $\epsilon \to u$; to delete a node u, we represent $u \to \epsilon$; to substitute the vertex u by vertex v, we represent $u \to v$. Likewise, to insert an edge e, we denote by $\epsilon \to e$, $e \to \epsilon$ represents deletion of the edge e, and $e \to f$ defines substitution of the edge e by edge f.

In the following definitions, for simplicity, we denote a graph G_i by $G_i = (V_i, E_i, \mu_i, \nu_i)$.

Definition 1 The *GED* from G_1 to G_2 is stated by

$$GED(G_1, G_2) = min_{(e_1,\ldots,e_k)\in\varphi(G_1,G_2)} \sum_{i=1}^{k} c(e_i);$$

here $c(e_i)$ represents the costs of corresponding edit operations of e_i, and $\varphi(G_1, G_2)$ denotes the sequence of edit path to convert G_1 into G_2.

Two graphs are homeomorphic when both graphs are a subdivision of another graph. Subdivision of an edge is the process of inserting an additional vertex along an edge. The subdivision on graph G creates another graph after performing the subdivision on the edges of this graph. We can observe that the subdivision operation on a graph only changes the number of nodes of degree two.

Definition 2 The homeomorphic GED *HGED* from graph G_1 to G_2 is defined as

$$HGED(G_1, G_2) = GED(H_1, H_2) = min_{(e_1,\ldots,e_k)\in\varphi(H_1,H_2)} \sum_{i=1}^{k} c(e_i),$$

where G_1' is the graph obtained from G_1 by doing path contraction from every vertex. Similarly, G_2' is the graph got from G_2 by performing path contraction from each vertex.

Path contraction is the technique of removing every intermediate vertex having degree 2, except first and last nodes. *Node contraction* is the method of removing vertices along with the incident edges, given that it is not an articulation point or cut vertex [9].

A brief explanation of the essential centrality measures [20] follows. Centrality indicates the importance of a node in a network or graph. A node with high centrality will be more significant concerning other nodes in the graph. The *degree centrality* denotes the degree of a node in the graph. When a graph is a directed graph, it will have both the indegree and outdegree centrality. The *betweenness centrality* denotes the number of times a node occurs between other two nodes on their shortest path. The *eigenvector centrality* measures the node's influence using the count of its links to other vertices in the graph. The *PageRank centrality* considers the importance of a vertex proportional to the influence of its neighboring vertices divided by their outdegree.

Definition 3 r-CentralityNodeContraction is the operation to contract the r ratio of vertices from G having minimum centrality scores of a specified centrality identifier.

Definition 4 r-CentralityGraphEditDistance computation between G_1 and G_2 is defined as the GED between these two graphs, where $r.G_1$ nodes of G_1 and $r.G_2$ nodes of G_2 having minimum centrality value are contracted.

Definition 5 In the t-CentralityNodeContraction, t nodes having the minimum centrality score of a specified centrality measure are removed, given that these nodes are not the cut vertex.

Definition 6 t-CentralityGraphEditDistance is the method of computing GED from G_1 to G_2 after deleting the t nodes from both the graphs having the minimum centrality score of a given centering measure.

3 Algorithms

The computation of GED is generally accomplished utilizing tree-search-based algorithms. Tree-search-based techniques will be able to traverse the entire search space for the assignments of vertices and edges from the one graph to another for finding the optimal edit transformation. A commonly used technique is based on A^* search method using an ordered tree to explore the complete search space. The tree's root node will start with the initial solution. Intermediate nodes denote the subsequent partial solution, whereas the leaf nodes represent the complete edit path. A heuristic function is used for selecting the next successor node to be explored at the given level.

Algorithm 1 describes to compute the GED. The inputs to the GraphEditDistance algorithm are G_1 and G_2, where graph G_1 has n vertices and graph G_2 has m vertices. The outcome of this algorithm is the least-cost GED from graph G_1 to graph G_2. The algorithm starts with the empty set A in line 1. Set A includes all the partial edit paths created so far. During the *for* loop of lines 3–4, every vertex of the second graph is substituted by the u_1 vertex of the first graph. Each of these substitutions is then inserted into the set A. Deletion of node u_1 is also appended to the set A. The *while loop* of lines 6–25 computes the minimum cost edit path C_{min} from A. The algorithm uses a heuristic function $g(C) + h(C)$ to select the minimum cost edit path. $g(C)$ represents the cost of optimal edit path from the top root vertex to the current vertex C. $h(C)$ is used to denote the estimated cost from the current node C to the leaf node. The sum $g(C) + h(C)$ provides the total cost assigned to a vertex in the search tree. Here the optimal edit path from root vertex to a leaf vertex will be computed. The *if* loop of line 8 tests to know if the constructed C_{min} is one of the complete edit paths. If this is the case, the algorithm returns the corresponding complete edit path in line 9. If all nodes of first graph G_1 are visited (line 11), then the remaining unvisited nodes of graph G_2 (line 12) are inserted in to the graph (line 13) and A is updated in line 16. Similarly, in the *for* loop of lines 18–22, all the unvisited nodes of G_1 are substituted by each node of G_2 along with the deletion of each node of G_1 in line 19, and finally A is updated in line 22.

GraphEditDistance algorithm is an exact algorithm that explores the complete search space to find the optimal edit path to transform G_1 into G_2. This algorithm is exact, but it is computationally very expensive; therefore, this algorithm may

Algorithm 1 GraphEditDistance (G_1, G_2)

Require: Two Graphs G_1, G_2, where $V_1 = \{u_1, \ldots, u_n\}$ and $V_2 = \{v_1, \ldots, v_m\}$
Ensure: A minimum cost GED between G_1 and G_2

1: $A \leftarrow \emptyset$
2: **for** each $(v_j \in V_2)$ **do**
3: $\quad A \leftarrow A \cup \{u_1 \rightarrow v_j\}$
4: **end for**
5: $A \leftarrow A \cup \{u_1 \rightarrow \epsilon\}$
6: **while** (True) **do**
7: $\quad$ Compute minimum cost edit path $C_{min} = min_{C \in A}\{g(C) + h(C)\}$ from A
8: $\quad$ **if** (C_{min} is a complete edit path) **then**
9: $\quad\quad$ **return** C_{min}
10: $\quad$ **else**
11: $\quad\quad$ **if** (all vertices $(u_i \in V_1)$ are visited) **then**
12: $\quad\quad\quad$ **for** all unvisited $(v_j \in V_2)$ **do**
13: $\quad\quad\quad\quad C_{min} \leftarrow C_{min} \cup \{\epsilon \rightarrow v_j\}$
14: $\quad\quad\quad$ **end for**
15: $\quad\quad\quad A \leftarrow A \cup \{C_{min}\}$
16: $\quad\quad$ **else**
17: $\quad\quad\quad$ **for** (all unvisited vertices $(u_i \in V_1)$) **do**
18: $\quad\quad\quad\quad$ **for** (each $(v_j \in V_2)$) **do**
19: $\quad\quad\quad\quad\quad C_{min} \leftarrow C_{min} \cup \{u_i \rightarrow v_j\} \cup \{u_i \rightarrow \epsilon\}$
20: $\quad\quad\quad\quad$ **end for**
21: $\quad\quad\quad$ **end for**
22: $\quad\quad\quad A \leftarrow A \cup \{C_{min}\}$
23: $\quad\quad$ **end if**
24: $\quad$ **end if**
25: **end while**

not be feasible for the graphs having large sizes. To overcome this disadvantage, several approximate and suboptimal algorithms have been proposed. One of the approximate algorithms for GED is outlined in Algorithm 2. The basic idea of the suboptimal algorithm is to prune the search space using some optimizing and heuristic techniques so that the resulting search space is reduced, thereby reducing the computation time. The steps of ApproximateGraphEditDistance algorithm are the following. The inputs to the ApproximateGraphEditDistance algorithm are G_1 and G_2. The output of this algorithm is the minimum cost approximate GED from graphs G_1 to G_2. The steps of this algorithm are similar to the GraphEditDistance algorithm except for the *while* loop of lines 6–26. Before selecting the next node for consideration, set A is pruned using some optimizing techniques such as beam search in which a fixed number of nodes known as beam width are only explored for selecting the next candidate. Since the heuristic methods view only the partial set of edit operations in the complete search space, this results in an approximate but comparatively efficient solution.

Algorithm 2 ApproximateGraphEditDistance (G_1, G_2)

Require: Two Graphs G_1, G_2, where $V_1 = \{u_1, \ldots, u_n\}$ and $V_2 = \{v_1, \ldots, v_m\}$
Ensure: A minimum cost GED between G_1 and G_2

$A \leftarrow \emptyset$
for each $(v_j \in V_2)$ **do**
  $A \leftarrow A \cup \{u_1 \rightarrow v_j\}$
end for
$A \leftarrow A \cup \{u_1 \rightarrow \epsilon\}$
while (True) **do**
  Prune A using an optimizing/heuristic technique
  Find min. cost edit path C_{min} from A
  if (C_{min} is a complete edit path) **then**
    return C_{min}
  else
    if (all vertices $(u_i \in V_1)$ are visited) **then**
      for all unvisited $(v_j \in V_2)$ **do**
        $C_{min} \leftarrow C_{min} \cup \{\epsilon \rightarrow v_j\}$
      end for
      $A \leftarrow A \cup \{C_{min}\}$
    else
      for (all unvisited vertices $(u_i \in V_1)$) **do**
        for (each $(v_j \in V_2)$) **do**
          $C_{min} \leftarrow C_{min} \cup \{u_i \rightarrow v_j\} \cup \{u_i \rightarrow \epsilon\}$
        end for
      end for
      $A \leftarrow A \cup \{C_{min}\}$
    end if
  end if
end while

3.1 Homeomorphic Graph Edit Distance

As discussed before, we say two graphs to be homeomorphic when both graphs are a subdivision of another graph. During the homeomorphic GED, first, all the nodes of degree 2 are deleted except the first and last nodes along all the simple paths of the graphs, and after that, GED is computed.

Given two graphs G_1 and G_2, homeomorphic edit cost function $\forall$ vertices $u \in V_1$, $v \in V_2$ and $\forall$ edges $e \in E_1$, $e' \in E_2$ is defined as:

$$c(u \rightarrow \epsilon) = x_{node}$$
$$c(\epsilon \rightarrow v) = x_{node}$$
$$c(u \rightarrow v) = y_{node}.||\mu_1(u) - \mu_2(v)||$$
$$c(e \rightarrow \epsilon) = x_{edge}$$
$$c(\epsilon \rightarrow e') = x_{edge}$$
$$c(e \rightarrow e') = y_{edge}.||\nu_1(e) - \nu_2(e')||$$
$$c((u_1, \ldots, u_n) \rightarrow (u_1, u_n)) = z_{path}.||\mu_1(u_1) - \mu_1(u_n)||$$

Here x_{node}, x_{edge}, y_{node}, y_{edge}, and z_{path} are positive values, and $c(u \rightarrow \epsilon)$ and $c(\epsilon \rightarrow v)$ are the costs of deleting vertex u and inserting vertex v, respectively, $c(u \rightarrow v)$ is the charge of substituting vertex u by vertex v, $c(e \rightarrow \epsilon)$ and $c(\epsilon \rightarrow e')$ are the costs of deleting edge e and inserting edge e', respectively, and $c((u_1, \ldots, u_n) \rightarrow (u_1, u_n))$ is the charge of performing path contraction from $(u_1, \ldots, u_n)$ to (u_1, u_n).

The computation of homeomorphic GED is outlined in Algorithm 3. Input to this algorithm is the two input graphs, G_1 and G_2, and the outcome of this algorithm is the least-cost homeomorphic GED from G_1 to G_2. The steps to perform HomeomorphicGraphEditDistance algorithm are follows. The first *for* loop performs the path contraction operations over all the simple paths of graph G_1 to delete all the intermediate nodes of degree 2 except the first and last nodes along the path. The *if* loop checks for all the candidate paths that are eligible for the process of path contraction. Similarly, the second *for* loop performs the path contraction operations over all the simple paths of graph G_2 to delete all the intermediate nodes of degree 2 except the first and last nodes along the path. After performing the path contraction operation, both input graphs G_1 and G_2 are updated along with their modified number of vertices and edges. Finally, the ApproximateGraphEditDistance algorithm is called on the updated G_1 and G_2 to compute the GED.

Algorithm 3 HomeomorphicGraphEditDistance (G_1, G_2)

INPUT: Two Graphs G_1, G_2, where $G_i = (V_i, E_i, \mu_i, \nu_i)$ for $i = 1, 2$
where $V_1 = \{u_1, \ldots, u_n\}$ and $V_2 = \{v_1, \ldots, v_m\}$
OUTPUT: A min. cost homeomorphic GED between G_1 and G_2
for each $(u_i \in V_1)$ **do**
  if (there is a path $(u_i, u_{i+1}, \ldots, u_{i+k})$ such that
$deg(u_{i+1}) = deg(u_{i+2}) = \ldots = deg(u_{i+k-1}) = 2)$ **then**
    $(u_i, u_{i+1}, \ldots, u_{i+k}) \rightarrow (u_i, u_{i+k})$

    $V_1 \leftarrow V_1 \setminus \{u_{i+1}, \ldots, u_{i+k-1}\}$

  end if
end for
for each $(v_j \in V_2)$ **do**
  if (there is a path $(v_j, v_{j+1}, \ldots, v_{j+k})$ such that
$deg(v_{j+1}) = deg(v_{j+2}) = \ldots = deg(v_{j+k-1}) = 2)$ **then**
    $(v_j, v_{j+1}, \ldots, v_{j+k}) \rightarrow (v_j, v_{j+k})$

    $V_2 \leftarrow V_2 \setminus \{v_{j+1}, \ldots, v_{j+k-1}\}$

  end if
end for
Update $G_1, G_2, n \leftarrow n', m \leftarrow m'$
ApproximateGraphEditDistance (G_1, G_2)

3.2 GED Utilizing Centrality Information

To decrease the computing time of GED, first we remove the vertices from the graphs having lower centrality scores prior to computing the GED between the input graphs. r-CentralityGraphEditDistance is an extension to GED to compute the approximate value of GED by ignoring the r fraction of nodes from the input graphs using the specific centrality measure.

Definition 7 r-DegreeCentralityNodeContraction is a process of deleting $r.|G|$ vertices with lowest degree from G, provided these nodes are not cut vertices.

Definition 8 r-BetweennessCentralityNodeContraction is a task of deleting $r.|G|$ vertices having minimum betweenness value from G, provided these nodes are not cut vertices.

Definition 9 r-EigenvectorCentralityNodeContraction is a task of deleting $r.|G|$ vertices having least eigenvector value from G, provided these nodes are not cut vertices.

Definition 10 r-PageRankCentralityNodeContraction is a task of deleting $r.|G|$ vertices having minimum PageRank value from G, provided these nodes are not cut vertices.

The edit cost of r-centrality GED can be defined utilizing an extra cost $c(a \rightarrow \epsilon) = 0$, for $r.|G|$ nodes of the G with the least value of the specified centrality indicator.

r-GED uses the Euclidean's distances to assign a fixed cost for inserting, deleting, and substituting the nodes and edges. Suppose graphs G_1 and G_2 have nodes $a \in V_1, b \in V_2$ and links $e \in E_1, f \in E_2$, the modified edit cost function can be specified as given below:

$c(a \rightarrow \epsilon) = p_{node}.$
$c(\epsilon \rightarrow b) = p_{node}.$
$c(a \rightarrow b) = q_{node}.||\mu_1(a) - \mu_2(b)||.$
$c(e \rightarrow \epsilon) = p_{edge}.$
$c(\epsilon \rightarrow f) = p_{edge}.$
$c(e \rightarrow f) = q_{edge}.||\nu_1(e) - \nu_2(f)||.$
$c(a \rightarrow \epsilon) = 0$, when a is one of the $r.|G|$ vertices with minimum centrality score, and it is not an articulation point.

$p_{node}, q_{node}, p_{edge}, q_{edge}$ are positive constants.

The steps to perform r-CentralityGraphEditDistance (G_1, G_2) are outlined in Algorithm 4. The r-CentralityGraphEditDistance algorithm's inputs are G_1 and G_2 and the parameter r. The outcome of this algorithm is minimum cost r-CentralityGraphEditDistance between G_1 and G_2. Line 1 of the algorithm calls the procedure r-CentralityNodeContraction that removes $r.|G|$ vertices of minimum

Algorithm 4 r-CentralityGraphEditDistance (G_1, G_2)

Require: Two Graphs G_1, G_2, where $|V_1| = n$ and $|V_2| = m$, a constant r
Ensure: A min. cost r-GED between G_1 and G_2
1: $G_1' \leftarrow r$-CentralityNodeContraction $(G_1, \lceil r.n \rceil)$
2: $G_2' \leftarrow r$-CentralityNodeContraction $(G_2, \lceil r.m \rceil)$
3: **ApproximateGraphEditDistance** (G_1', G_2')
4: **procedure** r-**CentralityNodeContraction**$(G, \lceil r.|G| \rceil)$
5:     **for** $(i \leftarrow 1$ to $\lceil r.|G| \rceil)$ **do**
6:         Choose the node v having least centrality value
7:         **if** (v is not an articulation point) **then**
8:             $V \leftarrow V \setminus \{v\}$
9:             $E \leftarrow E \setminus \{(u, v) | (u, v) \in E$ for every $u \in G\}$
10:        **end if**
11:    **end for**
12:    **return** G
13: **end procedure**

centrality value from G_1. Similarly, line 2 of the algorithm calls the procedure r-Centrality node contraction that removes $r.|G|$ nodes of least centrality value from G_2. The procedure r-CentralityNodeContraction is described in lines 4–13. This procedure uses *for* loop in lines 5–11 to delete $\lceil r.|G| \rceil)$ vertices from a G with minimum centrality value of specified centrality criteria. It selects the node u with minimum centrality value (line 6), verifies that it is not a cut vertex (line 7), after that the corresponding node is deleted in line 8, and the associated edges are removed in line 9. The preprocessed graph is returned in line 12. Finally, line 3 executes ApproximateGraphEditDistance and computes minimum cost edit path that also satisfies for complete edit path.

4 Results and Discussion

Pattern recognition is among the significant applications of GED. Especially in structural pattern recognition, where the underlying pattern has structures that fixed dimensional vectors cannot represent, graphs can be utilized to represent such structures. When a pattern represents a graph, pattern recognition is usually known as GM. GED is a crucial technique for GM. In Sect. 3, we discussed a few important algorithms for computing the GED between two graphs. Since Algorithm 1 explores the complete search space to find the optimum edit path, it takes an exponential amount of time to output this GED. Due to its computationally exponential complexity, this algorithm cannot be used for the graph exceeding 10–15 nodes. Algorithm 2 computes the inexact approximate GED between two graphs by pruning the search space using some heuristic function for finding the successive node at every levels of the search tree. Algorithm 3 uses the inexact GED after performing the path contraction of every simple path of both input graphs. Algorithm 4 also uses the idea of Algorithm 2 to prune the search space

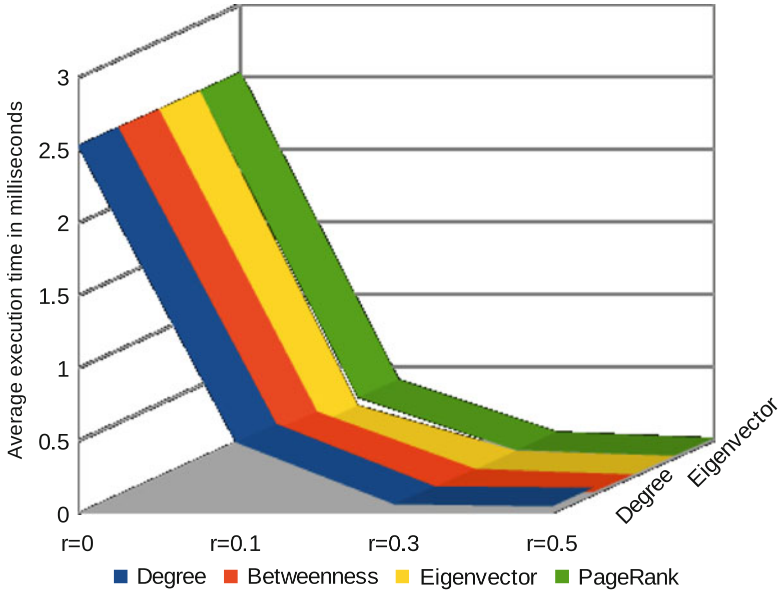


Fig. 1 Computation time for letter A datasets

after performing the node contraction on both input graphs using a given centrality measure.

In this section, r-CentralityGraphEditDistance algorithm's computation time and accuracy to perform the GM are observed. This section utilizes letters and AIDS datasets from IAM graph database [22] to perform the GM. Computation time is the execution time of the algorithm to perform the GM using r-CentralityGraphEditDistance for a given centrality measure. Accuracy of the algorithm is described in terms of classification accuracy on the test set after performing training of the algorithms using a training set of the given datasets.

The computation time taken by Algorithm 4 in milliseconds to perform the GM on the letter A datasets using the four centrality measures, degree, betweenness, eigenvector, and PageRank, is shown in Fig. 1. This figure shows the difference in time taken by this algorithm using the various centrality measures.

Figure 2 shows the computation time used by Algorithm 4 in milliseconds to perform GM for the active AIDS datasets utilizing the various centrality criteria. To prune the search space and to select the successive node in the search tree, we have used beam search heuristic technique.

Accuracy of Algorithm 4 on letter A datasets of high distortions utilizing the given centrality measures for four different values of $r = 0, 0.1, 0.3$, and 0.5 is provided in Fig. 3. This figure demonstrates that the accuracy ratio using the eigenvector centrality is usually more than other centrality indicators. From this, we can infer that eigenvector centrality can be more suitable to perform GM on the letter datasets.

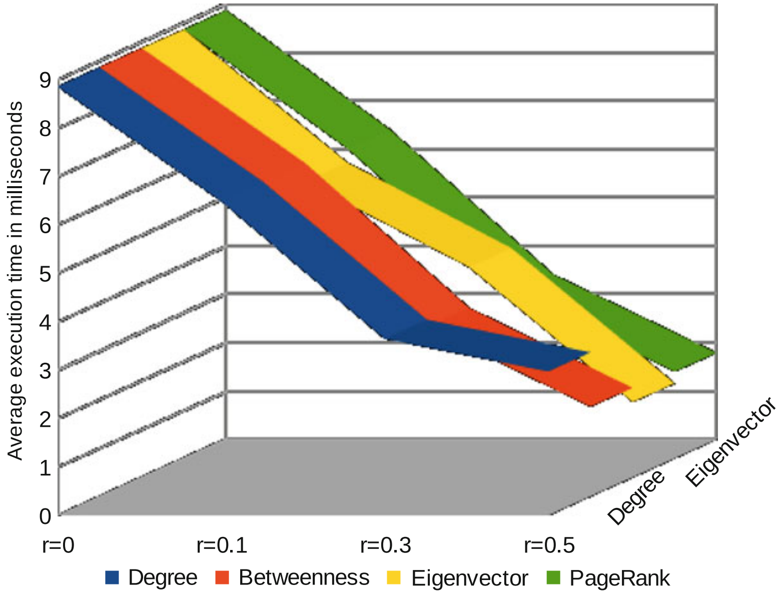


Fig. 2 Computation time for active AIDS datasets

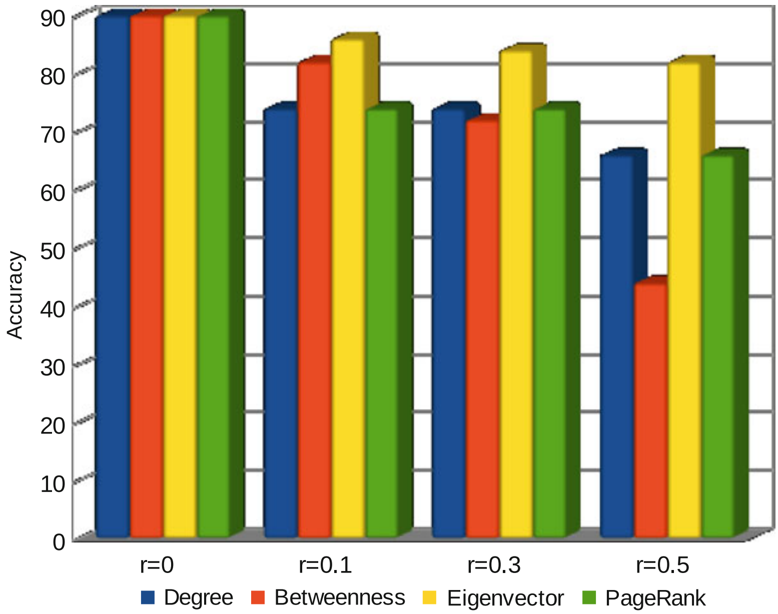


Fig. 3 Accuracy ratio of letter A datasets

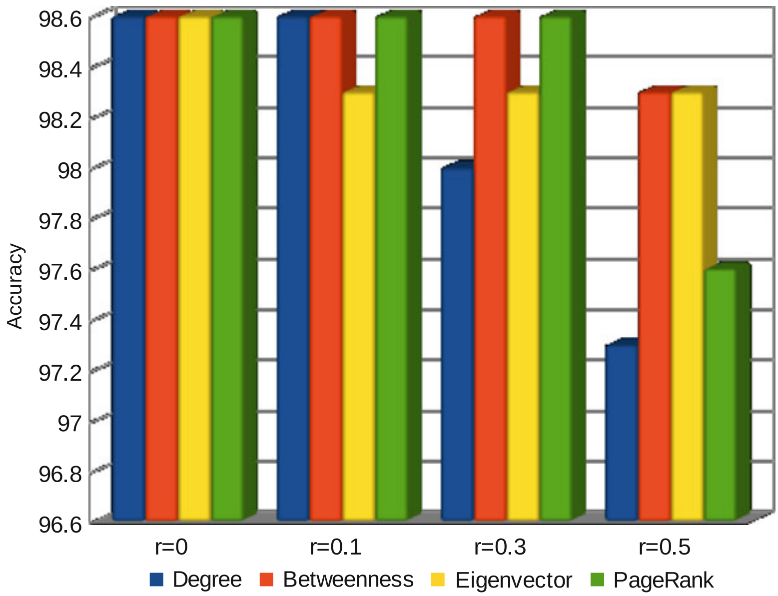


Fig. 4 Accuracy ratio for active AIDS datasets

Classification accuracy of r-CentralityGraphEditDistance to perform GM for the active AIDS datasets using the four centrality measures for four different values of $r = 0, 0.1, 0.3$, and 0.5 is provided in Fig. 4. This figure indicates that the classification accuracy utilizing betweenness criteria is generally more than other centrality indicators.

5 Conclusion

This chapter discussed the GED-based techniques for structural pattern recognition. GED is a crucial technique to measure the similarity between two graphs. We presented the important techniques and algorithms to compute the GED. We also described the various extensions and advancements to compute the GED that can be used for the trade-off between efficiency and accuracy consideration.

References

1. C.C. Aggarwal, H. Wang, Managing and mining graph data, in *Advances in Database Systems* (Springer, Berlin, 2010)
2. H. Bunke, Error-tolerant graph matching: A formal framework and algorithms, in *Advances in Pattern Recognition, International Workshop on Structural, Syntactic and Statistical Pattern Recognition (S+SSPR)*. Lecture Notes in Computer Science (Springer, Berlin, 1998)

3. H. Bunke, G. Allerman, Inexact graph matching for structural pattern recognition. Pattern Recog. Lett. **1**, 245–253 (1983)
4. D. Conte, P. Foggia,C. Sansone, M. Vento, Thirty years of graph matching in pattern recognition. Int. J. Pattern Recog. Artif. Intell. **18**(3), 265–298 (2004)
5. S.P. Dwivedi, Some algorithms on exact, approximate and error-tolerant graph matching. PhD Thesis, Indian Institute of Technology (BHU), Varanasi, 2019. arXiv:2012.15279
6. S.P. Dwivedi, Inexact graph matching using centrality measures (2021). *arXiv:2201.04563*
7. S.P. Dwivedi, Approximate bipartite graph matching by modifying cost matrix. Lecture Notes Electr. Eng. **837**, 415–422 (2022)
8. S.P. Dwivedi, R.S. Singh, Error-tolerant graph matching using homeomorphism, in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (2017), pp. 1762–1766
9. S.P. Dwivedi, R.S. Singh, Error-tolerant graph matching using node contraction. Pattern Recog. Lett. **116**, 58–64 (2018)
10. S.P. Dwivedi, R.S. Singh, Error-tolerant geometric graph similarity, in *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Lecture Notes in Computer Science, vol. 11004 (Springer, Berlin, 2018), pp. 337–344
11. S.P. Dwivedi, R.S. Singh, Error-tolerant geometric graph similarity and matching. Pattern Recog. Lett. **125**, 625–631 (2019)
12. S.P. Dwivedi, R.S. Singh, Error-tolerant approximate graph matching utilizing node centrality information. Pattern Recog. Lett. **133**, 313–319 (2020)
13. M. Ferrer, F. Serratosa, K. Riesen, Improving bipartite graph matching by assessing the assignment confidence. Pattern Recog. Lett. **65**, 29–36 (2015)
14. P. Foggia, G. Percannella, M. Vento, Graph matching and learning in pattern recognition in the last 10 years. Int. J. Pattern Recog. Artif. Intell. **28**, 1450001.1–1450001.40 (2014)
15. K.S. Fu, B.K. Bhargava, Tree systems for syntactic pattern recognition. IEEE Trans. Comput. **22**, 1087–1099 (1973)
16. P.E. Hart, N.J. Nilson, B. Raphael, A formal basis for heuristic determination of minimum cost paths. IEEE Trans. Syst. Sci. Cyber. **4**, 100–107 (1968)
17. M. Neuhaus, H. Bunke, An error-tolerant approximate matching algorithm for attributed planar graphs and its application to fingerprint classification, in *International Workshop on Structural, Syntactic and Statistical Pattern Recognition (SSPR and SPR)*. Lecture Notes in Computer Science, vol. 3138 (Springer, Berlin, 2004), pp. 180–189
18. M. Neuhaus, H. Bunke, *Bridging the Gap Between Graph Edit Distance and Kernel Machines* (World Scientific, Singapore, 2007)
19. M. Neuhaus, K. Riesen, H. Bunke, Fast suboptimal algorithms for the computation of graph edit distance, in *Proceedings of the 11th International Workshop on Structural and Syntactic Pattern Recognition*. Lecture Notes in Computer Science, vol. 4109 (Springer, Berlin, 2006), pp. 163–172
20. M.E.J. Newman, *Networks–An Introduction* (Oxford University Press, Oxford, 2010)
21. K. Riesen, *Structural Pattern Recognition with Graph Edit Distance, Approximation Algorithms and Applications* (Springer, Berlin, 2015)
22. K. Riesen, H. Bunke, IAM graph database repository for graph based pattern recognition and machine learning, in *International Workshop on Structural, Syntactic and Statistical Pattern Recognition (S+SSPR)*. Lecture Notes in Computer Science, vol. 5342 (Springer, Berlin, 2008), pp. 287–297
23. K. Riesen, H. Bunke, Approximate graph edit distance computation by means of bipartite graph matching. Image Vision Comput. **27**(4), 950–959 (2009)
24. K. Riesen, H. Bunke, Improving bipartite graph edit distance approximation using various search strategies. Pattern Recog. **48**(4), 1349–1363 (2015)
25. K. Riesen, A. Fischer, H. Bunke, Estimating graph edit distance using lower and upper bounds of bipartite approximations. Int. J. Pattern Recog. Artif. Intell. **29**(2), 1550011 (2015)

26. A. Robles-Kelly, E. Hancock, Graph edit distance from spectral seriation. IEEE Trans. Pattern Analy. Mach. Intell. **27**(3), 365–378 (2005)
27. A. Sanfeliu, K.S. Fu, A distance measure between attributed relational graphs for pattern recognition. IEEE Trans. Syst. Man Cyber. **13**(3), 353–363 (1983)
28. A. Sole-Ribalta, F. Serratosa, A. Sanfeliu, On the graph edit distance cost: properties and applications. Int. J. Pattern Recog. Artif. Intell. **26**(5), 1260004.1–1260004.21 (2012)
29. W.H. Tsai, K.S. Fu, Error-correcting isomorphisms of attributed relational graphs for pattern analysis. IEEE Trans. Syst. Man Cyber. **9**, 757–768 (1979)
30. Z. Zeng, A.K.H. Tung, J. Wang, J. Feng, L. Zhou, Comparing stars: on approximating graph edit distance. PVLDB **2**, 25–36 (2009)

Index

S. Awasthi et al. (eds.), *Sustainable Computing*,
https://doi.org/10.1007/978-3-031-13577-4