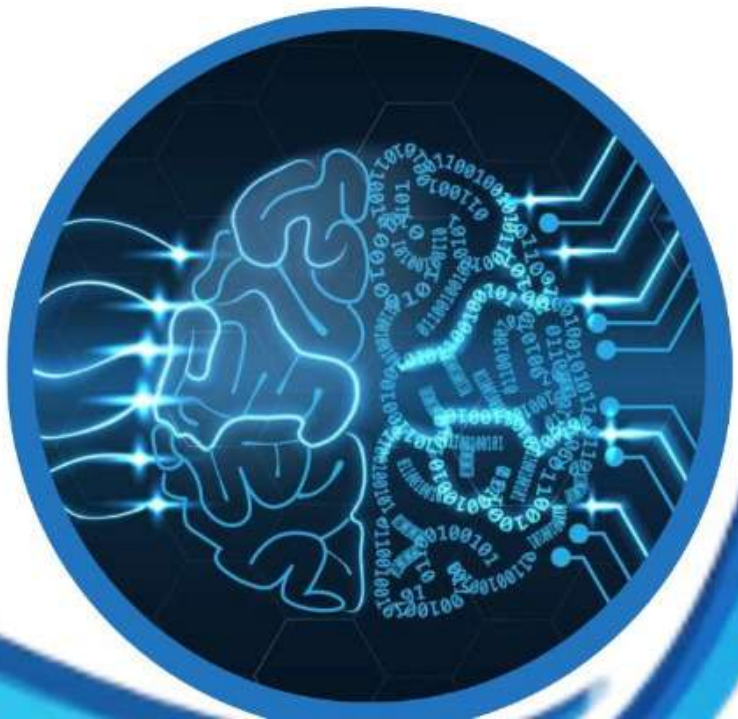
Application of Artificial Intelligent and IoT

DR. ABDELGADIR ELAMIN ELTOM



Application of Artificial Intelligent and IoT

EMPYREAL
PUBLISHING HOUSE

India | UAE | Nigeria | Uzbekistan | Montenegro | Iraq | Egypt | Thailand | Uganda | Philippines | Indonesia

www.empyrealpublishinghouse.com

Application of Artificial Intelligent and IoT

Edited By:

Dr. Abdelgadir Elamin Eltom
Associate Professor
Department of Clinical Chemistry

First Impression: January 2023

Application of Artificial Intelligent and IoT

ISBN : 978-93-93810-31-1

Rs. 650/- ($18)

DISCLAIMER

Information contained in this book has been published by Empyreal Publishing House and has been obtained by the editor from sources believed to be reliable and correct to the best of their knowledge. The author is solely responsible for the contents of the articles compiled in this book. Responsibility of authenticity of the work or the concepts/views presented by the author through this book shall lie with the author and the publisher has no role or claim or any responsibility in this regards. Errors, if any, are purely unintentional and readers are requested to communicate such error to the author to avoid discrepancies in future.

Published by:
Empyreal Publishing House

Dedication

To the sole of my Father

The greatest pleasure in life is doing what people say you cannot do …

I dedicate this work to …

Mother …

Family …

Sisters and Brothers

PREFACE

With the deployment of IoT, the business world is changing rapidly. IoT is assisting in the significant capture of a massive amount of data from many sources. Investment in new technologies will be required to realise the future and full potential of IoT devices. The combination of AI (Artificial Intelligence) with the Internet of Things (IoT) has the potential to reshape the way industries, businesses, and economies operate. IoT powered by AI generates intelligent technologies that results in intelligent behaviour and assist in decision-making with little or no human intervention.

Combining these two streams serves both ordinary people and experts. While IoT is concerned with devices connecting with each other over the internet, AI is focused with devices learning from their data and experience.

Businesses have been built or optimized using IoT devices and their data capabilities, ushering in a new era of business and consumer technology. Now the next wave is upon us as advances in AI and machine learning unleash the possibilities of IoT devices utilizing “artificial intelligence of things,” or AIoT.

Consumers, businesses, economies, and industries that adopt and invest in AIoT can leverage its power and gain competitive advantages. IoT collects the data, and AI analyzes it to simulate smart behavior and support decision-making processes with minimal human intervention.

ACKNOWLEDGEMENT

This book is a compilation of the research work conducted in the field of Artificial Intelligent and IoT. It is an attempt to integrate their knowledge and expertise. I would like to express my profound gratitude to all my Teachers who have been encouraging me to do things differently. I am grateful to all the researchers for contributing their quality research work. Wiithout their contribution and support, this book would not have been possible. I also extend my heartfelt gratitude to Empyreal Publishing House for their wholehearted support throughout the publication process. I am also grateful to my family members as they all kept me going and were very patient and supportive during the entire editorial process.

Dr. Abdelgadir Elamin Eltom

Table of Contents

ALGORITHM BASED ON RFID TECHNOLOGY FOR CORONAVIRUS TRACKING

Oscar Camilo Fuentes Amín[1] Mariano Esteban Romero Torres[2] Guido de Jesus Vida Ayala[3] and helmer Muñoz Hernández[4]

School of Basic Sciences, Technology and Engineering. - National Open and Distance University, Sahagun, Colombia[1,2, 3]

Faculty of Administrative and Accounting Economic Sciences - Universidad Del Sinú[4]

ABSTRACT

it is shown the design of an algorithm based on RFID (Radio Frequency Identification) Technology, able to identify related people to coronavirus sufferers. The device is composed by an Arduino board, an RFID Reader, four RFID cards and the algorithm. To test the algorithm functionality a simulation was made using Octave Online, with twelve recorded data from RFID tags, four locations and three different dates. Based on the proposed control strategy, it was possible to identify the related users to a possible positive case of Covid-19. The prototype could be used for Coronavirus tracking, allowing the gradual return to economic activities and the control of flow of people in the commercial sector.

Keywords—coronavirus tracking, people monitoring, Algorithm, rfid technology.

I. INTRODUCTION

On 31 December 2019, the World Health Organization (WHO) was informed of cases of pneumonia unknown etiology detected in Wuhan City, Hubei Province of China, the causal agent was not identified. Next, on 7 January, 2020, Chinese authorities identified a new type of coronavirus, which was isolated, the positive cases continued to rise covering several countries [1] [2] [3]. It is really worrying that coronavirus can be spread from person to person, according to a mathematical model it was estimated that the expected number of secondary infections that result from introducing a single infected individual into an otherwise susceptible population was 3.58 [4]. According to this, the WHO has asked every country in the world to implement some suggestions in order to mitigate as much as possible the spread of the virus [5]. There have been developed many laws and decrees related to some new social behavior practices, since the most effective measure to slow the spread is through social distancing [6]. In this respect, all countries rulers in the world have had to implement curfews, airport closures and to ban all social acts, as the best measures to ensure distancing.

The crisis caused by the coronavirus has had socio-economic implications, because people cannot leave their houses, supermarkets have had to limit shoppers, there are many travel restrictions, schools have closed down, to mention a few issues [7]. Because of the confinement, it has become necessary to work from home, so many labor sectors have had to develop or strengthen technologies that allow this new form of work. In this order of ideas, the use of technology has increased and these technological advances have been added to everyday life.

It should be noted that nowadays contemporary societies are undergoing accelerated changes by technological advances. The new industrial age is based on technological devices with internet connection, that field is called Internet of Things (IoT) and it covering many platforms as hardware and software to facilitate daily activities [8]. One of the recently used technologies is

called Radio Frequency Identification (RFID), which is accurate to assign electronic tags to objects for further identification. This technology has had many applications in different sectors [9] [10] [11].

Since the socio-economic implications of the coronavirus are known and knowing the features of some technologies of Internet of Thing, an algorithm based on RFID technology is proposed for coronavirus tracking as a possible solution to return to the economic activities, since this method couldassist the rulers for sectoring the population at risk of infection and isolate them, while the non-riskedpeople could be able to keep their economic activities.

II. MATERIALS AND METHODS

A. Coronavirus Tracking

The coronavirus tracking consists in the installation of a device composed by an RFID Reader, WifiModule and an Arduino, as shown in figure 1. The device must be installed on the door of each socialarea, for example the supermarkets. In addition each person must have an RFID tag and finally each device must record data on a cloud.

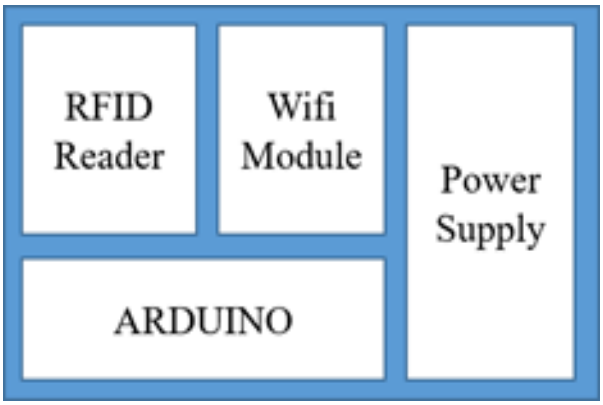


Fig. 1. Design Implementation proposal.

The proposed algorithm is able to build an array with tag, zone, year, month and day, which is record when an RFID Card passes by the reader. In addition, the algorithm shows all registers made by a particular user, who has informed about a possible positive case of infection. Finally the algorithmshows the tag of related users.

It is shown the design of an algorithm based on RFID technology and Arduino Nano, used to simulate a particular situation below.

B. Hardware

Radio Frequency Identification system is a technology able to identify objects through electromagnetic field, this technology uses radio-frequency waves to transfer identifying information between tagged objects and readers without requiring line of sight [12]. An RFID system consist of two stages, a card and a reader that creates an RF field for detecting radio waves. A tag has two parts, a small chip and an antenna, the chip contains the information while the antenna transmits it. The chip isused to store a unique identifying number and it is able to obtain energy from radio-frequency waves supplied by the reader, then the tags do not need power supply [13].

Thanks to the features described about RFID technology it was possible to design an electronic circuit, in order to simulate the flow of people in four different zones, with their respective dates. The system was composed by an Arduino board, an RFID reader, four tags and three buttons to simulate thezones and the alert command, see Figure 2.

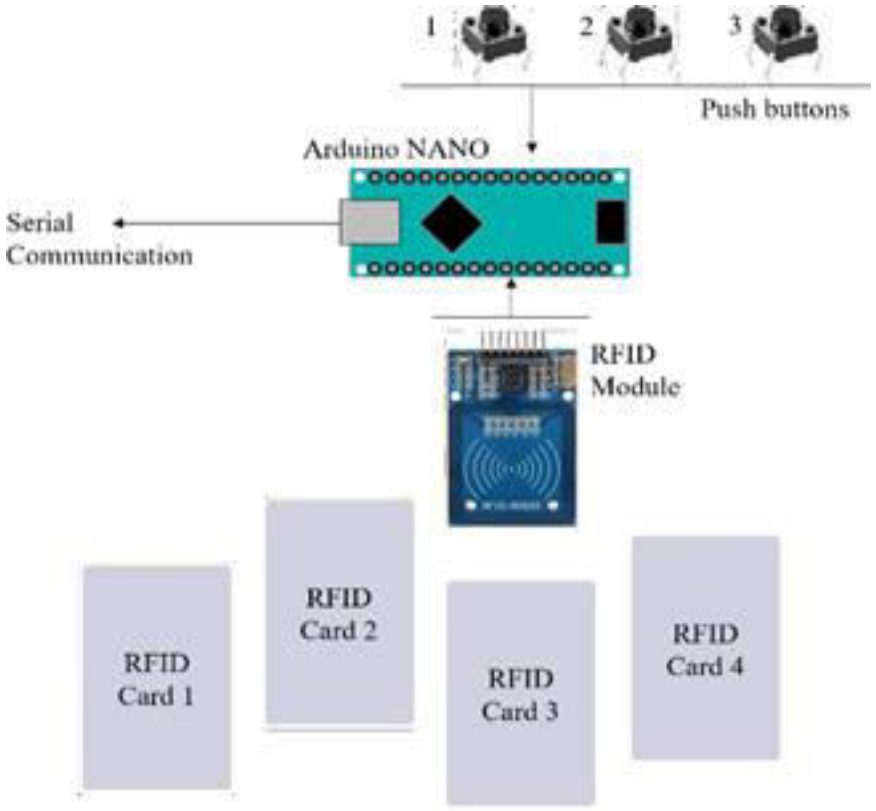


Fig. 2. Electronic circuit design.

Arduino is responsible to read the RFID tags through the Reader. According to the tag, Arduino assigns an ASCII code, 'A', 'B', 'C' or 'D'. On the other hand, the state of "push button 1" is read by Arduino, which send two possibilities, 'true' or 'false', in order to simulate an alert of positive case of COVID-19. The "push button 2" and "push button 3" are used to simulate different zones, '00' corresponds to zone 1, '01' corresponds to zone 2, '10' corresponds to zone 3 and '11' corresponds tozone 4. Taking into account the electronic circuit described, Arduino is able to send an array to a computer when a tag passes by the reader, and such Array has the following format:

Array: 'tag', 'zone', 'year', 'month', 'day'.

C. Algorithm

The algorithm was designed with the purpose of identifying the related users to a possible positive case of Covid-19 when the positive user reports an alert. The process begins with the registration of each user when a tag passes by electronic circuit, once this process is finished the system checks the status of the "push button 1", if the "push button 1" is 'false', then the algorithm waits for a new tag, ifthe "push button 1" is 'true', then the algorithm identifies the related users. The Figure 3 shows the Process Flow Diagram.

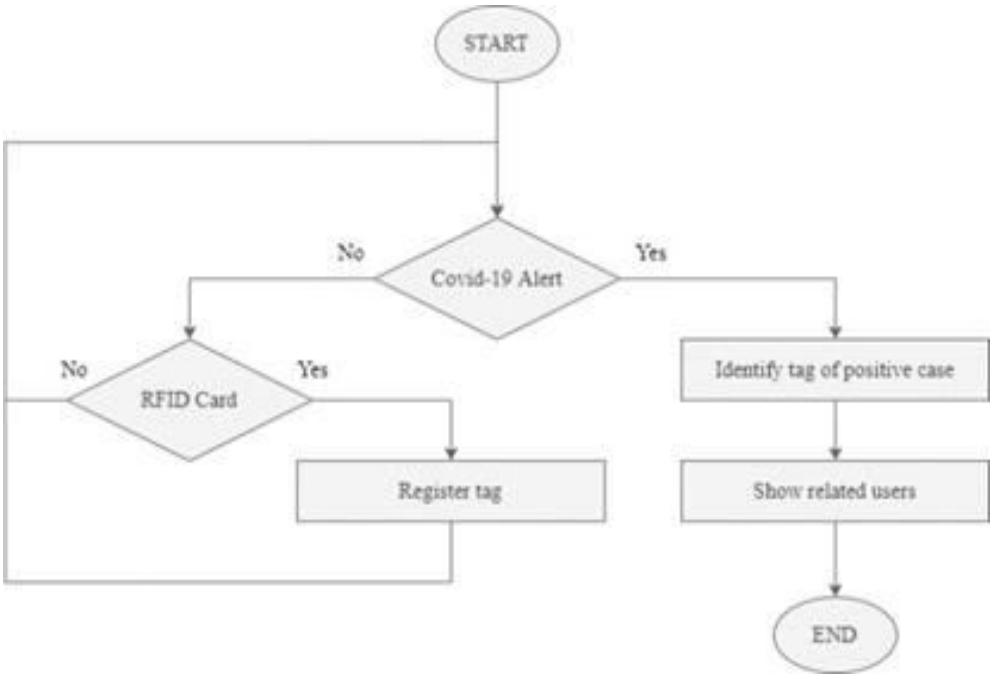


Fig. 3. Process flow diagram for Coronavirus tracking.

III. RESULTS

A simulation was made to test the algorithm, which was based on four RFID tags, four zones and three dates. The electronic system was programmed to send an array with "User tag", "Zone", "Date"and "Alert". Table 1 shows twelve random recordings of data in order to simulate a particular situation.

Table I. Registration of People Simulation.

Assigned user	**Registration of people**			
	Zone	**Year**	**Month**	**Day**
65	1	2020	05	01
66	1	2020	05	01
66	2	2020	05	01
67	2	2020	05	01
68	4	2020	05	01
65	2	2020	05	02
66	1	2020	05	02
67	1	2020	05	02
68	4	2020	05	02
65	1	2020	05	03
66	2	2020	05	03
68	4	2020	05	03

The numbers 65, 66, 67 and 68 correspond to the letters A, B, C and D respectively. According to the simulation, the user assigned to ASCII code 'A', was alerted as a possible positive case of infection.The algorithm showed all registered users through an array called "S", see Figure 4.

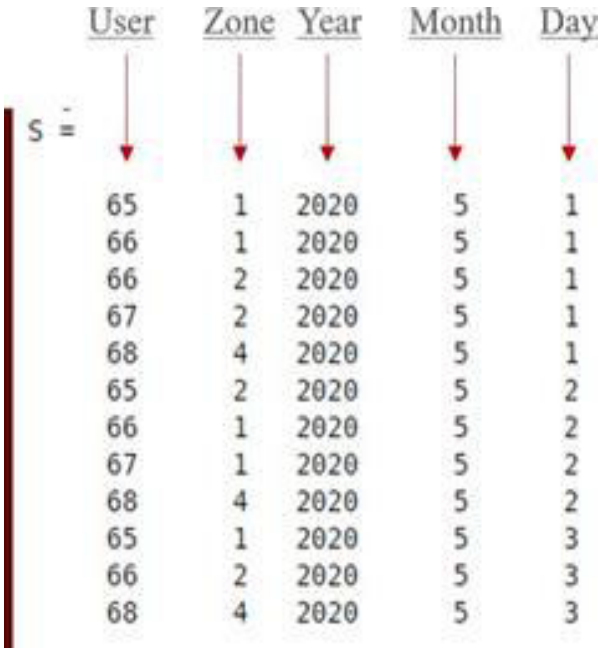


Fig. 4. All registered users.

Next, the algorithm showed the recording of the user "65", through an array called "COV", see Figure 5.

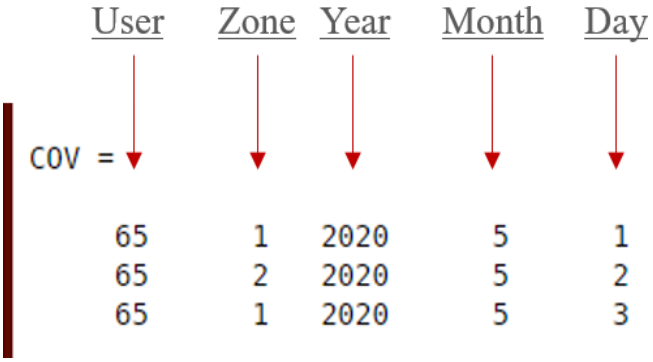


Fig. 5. User number 65.

Finally the algorithm developed was able to show the related users to "65", through a vector called"Rel", see Figure 6.

Fig. 6. Related users.

Taking into account the result showed in figure 6, User B could be at risk, since according to simulation, on May 1, 2020, B was in the same zone than A. It must be stated that the proposed trackingstrategy, based on RFID technology, was capable to identify the user at risk precisely.

IV. CONCLUSIONS

The design of an electronic system able to record data through RFID technology was raised as a possible solution for Coronavirus tracking. To achieve that purpose, it was design an algorithm able to get data from an electronic system composed by RFID technology and Arduino. As it was seen in thesimulation, the algorithm developed was capable to identify all registers made by a particular user andthe same algorithm shows the related users, who might be at risk.

Through the system designed it will be possible to assist the rulers for sectoring the population at risk of infection and isolate them, thus the non-risked people could be able to keep their social activities.

ACKNOWLEDGMENT

The author gratefully acknowledge the academic support provided by National Open and DistanceUniversity.

REFERENCES

[1] World Health Organization, "Novel Coronavirus (2019-nCoV) Situation Report - 1," 2020.

[2] L. Hongzhou, C. W. Stratton and Y. Tang, "Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle," Journal of Medical Virology, vol. 92, no. 4, pp. 401- 402, 2020.

[3] C. Wang, P. Horby, F. G. Hayden and G. F. Gao, "A novel coronavirus outbreak of global health concern," The Lancet, vol. 395, no. 10223, pp. 470 - 473, 2020.

[4] T.-M. Chen , J. Rui, Q.-P. Wang,, Z.-Y. Zhao , . J.-A. Cui and L. Yin , "A mathematical model for simulating the phase-based transmissibility of a novel coronavirus," Infectious Diseases of Poverty, vol. 9, no. 24, pp. 1 - 8, 2020.

[5] World Health Organization, "Population-based age-stratified seroepidemiological investigation protocol for coronavirus 2019 (COVID-19) infection," 26 May 2020. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/332188/WHO-2019-nCoV-Seroepidemiology- 2020.2-eng.pdf.

[6] W. Smith and D. Freedman, "Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019- nCoV) outbreak," Journal of Travel Medicine, vol. 27, no. 2, pp. 1 - 4, 2020.

[7] M. Nicola, Z. Alsafi, C. Sohrabi, A. Kerwan, A. Al-Jabir, C. Losifidis, M. Agha and R. Agha, "The socio-economic implications of the coronavirus pandemic (COVID-19): A review," International Journal of Surgery, vol. 78, pp. 185-193, 2020.

[8] A. Zanella, N. Bui, A. Castellani, L. Vangelista and M. Zorzi, "Internet of Things for Smart Cities," IEEE Internet of Things Journal, vol. 1, no. 1, pp. 22-32, 2014.

[9] K. Domdouz, B. Kumar and C. Anumba, "Radio-Frequency Identification (RFID) applications: A brief introduction," Advanced Engineering Informatics, vol. 21, no. 4, pp. 350-355, 2007.

[10] E. Welbourne, L. Battle, G. Cole, K. Gould, K. Rector, S. Raymer, M. Balazinska and G. Borriello, "Building the Internet of Things Using RFID: The RFID Ecosystem Experience," IEEE Internet Computing, vol. 13, no. 3, pp. 48-55, 2009.

[11] D. Zhang, L. Yang, M. Chen, S. Zhao, M. Guo and Y. Zhang, "Real-Time Locating Systems Using Active RFID for Internet of Things," IEEE Systems Journal, vol. 10, no. 3, pp. 1226-1235, 2016.

[12] Q. Z. Sheng, X. Li and S. Zeadally, "Enabling Next-Generation RFID Applications: Solutions and Challenges," IEEE Xplore, vol. 41, no. 9, pp. 21-28, 2008.

[13] X. Zhu, S. K. Mukhopadhyay and H. Kurata, "A review of RFID technology and its managerial applications in different industries," Journal of Engineering and Technology Management, vol. 29, pp. 152-167, 2020.

INTERNET OF THINGS FOR INDUSTRY 5.0 INTEGRATED ARTIFICIAL INTELLIGENCE

Dr. Shipra Yadav[1], Dr. Keshao D. Kalaskar[2] and Dr. Pankaj Dhumane[3]
[1]Assistant Professor, IICC, RTM Nagpur University Nagpur, (MH), India
[2]Associate Professor, Dr. Ambedkar College, Chandrapur, (MH), India
[3]Assistant Professor, Sardar Patel College, Chandrapur, (MH), India

ABSTRACT

The Industry 5.0 vision, a precursor to the next industrial revolution and an improvement over Industry 4.0, envisioned the new aims of resilient, sustainable, and human-centric approaches in a variety of emerging applications, such as future factories and the digital society. AI-based technologies are the key element of the Internet of Things in the Industry 5.0 paradigm. Through precise production automation and the use of critical thinking abilities, Industry 5.0 demonstrated a substantial connection between intelligent systems and humans in the majorityof applications. Additionally, industry 5.0 introduces a number of capable technologies that assist businesses in working inexpensively and changing rapidly without making any initial investments. The Internet of Things (IoT) ecosystems have greatly improved in recent years due to significant advancements in smart devices, wireless connectivity, and sensor nodes. The Internet of Things (IoT) is a system of interconnected, digitally devices that can collect and send data over a wireless network without requiring human input. People feel more at ease and relieved as a result of the implementation of Internet of Things. IoT devices provide users the opportunity to generate reports without constraints and access information even in remote locations. Moreover, they accurately direct people with wise judgments using communication technologies, as was already mentioned. Preprocessing is necessary since many linked devices gather a lot of raw sensed data. However, it only becomes something important since IoT devices require sufficient resources for Edge computing. The essential tools for information inference in edge computing are AI-based algorithms. Moreover, the sensed data gathered by IoT applications is typically unstructured and needs further analysis, where AI-based models assist in extracting pertinent data. Additionally, there is a possibility for malicious attacks when data is transmitted from device to device. As a result, this chapter examines Industry 5.0, IoT architecture, and AI-based IoT. We also examine the technical aspects of IoT networks and discuss communication-enabling technologies. Next, we covered a variety of AI-based IoT technologies, AI-based tools, Edge computing, and trust models for IoT devices. Future smart cities hold the key to satisfying the residents' increasing demands. Improvements ininformation and communication will enable better utilization of the available resources. The world's metropolitan environment will ultimately determine its fate. The percentage of people who live in urban regions is increasing and is now a larger portion of the population overall. This sudden flood of people opens up possibilities but also presents difficulties. Governments around the world are challenged with the challenge of creating sustainable, reasonable space in the world's continuously growing cities. The Smart City concept is emerging, where the rights and welfare of the residents are guaranteed, the industry is active, and urban development is evaluated from an environmental perspective. This study provides an overview of future technologyanalysis and smart city requirements. We conduct thorough investigation to find and

examine the most recent technological developments, which serve as the cornerstone of the succeeding prosperous period. Deep learning (DL), machine learning (ML), the internet of things (IoT), mobile computing, big data, block chain, sixth-generation (6G) networks, WiFi-7, industry 5.0, robotic systems, heating, ventilation, and air conditioning (HVAC), digital forensic, industrial control systems, connected and automated vehicles (CAVs), electric vehicles, product recycling, flying cars, pantry backup, calamity backup, and crucial integration of cyber security to maintain the integrity of the system are some examples of these technologies. We provide a comprehensive analysis of the current application frameworks for future smart cities. We also talk about the different technological issues that smart cities of the future will face. In order to create smart cities that set the standard for smart living, we define the future dimensions of smart cities.

Index Terms: Survey, Current State, Smart cities, Internet of Things, cloud computing, intelligent sensors, transportation, data mining, and smart education are all terms used to describe future cities. Urban modeling, smart mobility, smart health, real-time systems, applications, challenges, and industry standards Cyber-Physical System, Industry 5.0, Internet of Things, Artificial Intelligence, Edge Computing, Dew Computing, Sensor, Machine Learning,Deep Learning, Augmented Reality, Virtual Reality, and Mixed Reality

1 . INTRODUCTION

Today's industry and academic buzzwords include wireless communications, IoT devices, intelligent sensors, industrial IoT, mobile-edge computing, and communication protocol. The Internet of Things (IoT) generally operates by implanting short-range movable transceivers into a variety of gadgets and ordinary objects, enabling unique communication channels among people, objects, and things. IoT would therefore provide information and communication a new dimension. IoT devices link to one another via innovative communication technologies includingRFID, Wi-Fi, GSM, Bluetooth, and many more, which can raise people's standards of living [1-4]. According to the most recent survey, more than 60 billion IoT devices, including sensors, laptops, game consoles, and embedded devices, are anticipated to exist in 2025. IoT knowledge has generally evolved in a manner that is surprisingly comparable to the development of the modern world, in which people and objects are practically integrated into information systems via wireless sensor technology. Information sharing is the primary goal of IoT integration, allowing intelligent environments to recognise items and then extract information. In the Internet of Things, embedded devices are essential since they primarily link with intelligent sensors to gather data. These sensor nodes are used by embedded systems to interact specifically with the physical environment [3-5]. To increase the operating efficiency of new appliances, the IoT platform now provides advanced management and monitoring functions. In the early years of 2000, a renowned researcher named Kevin Ashton defined and used the term"Internet-Of-Things" (IoT). According to Kevin Ashton's theories, the Internet of Things (IoT) is a structure or system of real objects connected to the internet via smart sensors. Moreover, Ashton invented the RFID technology, which is widely used for automated transportation tracking. Now, a variety of definitions are accessible based on their particular application in the real world. For instance, according to IEEE, "the Internet of Things (IoT) is a framework of connected technology that enable for new appliances and services to enable interaction between both the physical and virtual worlds through M2M communication]"[3].

According to the definition given by the Internet Architecture Boards (IAB), the Internet of Things (IoT) is the networking of "smart objects," which refers to numerous devices that can rationally communicate using internet protocol but are not directly controlled by people. Instead, these devices are parts of structures, vehicles, or the environment. [6]. As was previously mentioned, the majority of IoT systems are growing more dynamic, diverse, and complex; as a result, it might be difficult to organise such an IoT system or model. Additionally, in order to draw in more abusers, such IoT systems' solutions need to be more efficient and flexible. By utilising changes to computing technologies, Artificial Intelligence (AI) has recently achieved significant success in several fields [5]. Another innovative technology used on IoT for enhanced services ismachine learning (ML), a branch of artificial intelligence. IoT is widely acknowledged to require both AI and ML to create intelligent network management and operations. By utilising AI and ML in pattern recognition, Natural Language processing, object detection, and network sharing, many different types of research project obtained excellent results. Consequently, the IoT sector can profit from utilizing AI and ML technology. There are lots of opportunities to improve well-organized smart real-world appliances by using AI and ML-based models with IoT [6-7].

Everyone needs to look at this and understand how an IoT functions and impacts our daily lives before learning about the technical research trends of IoT. Each researcher and data scientist seeks to import and comprehend IoT foundations in accordance with their visualization and subsequent needs. After all, the requirements for IoT representation do not have an universal definition. IoT representations give rise to the terms Internet of Everything (IoE), Internet of Things (IoT), and Web of Things (WoT), each of which has a defined set of operational protocols. However, the Internet of Things (IoT) is based on the integration of different standards and supporting technologies with different sensor, processing, connective, and storage capabilities. When it comes to the Internet of Things, the integration standards being used devices create significant difficulties. Since they are essential to the continued growth of IoT projects, integration challenges in IoT devices are regarded as critical IoT concerns [7-9]. IoT expansions, modernization, and proper setting of things are currently the object of extensive standardization administrations, associations, researchers, and manufacturing industries. A broad context with integrated ethics under one IoT is still lacking, unfortunately. Every field has been adequately penetrated by the Internet and its linked technologies. The advantages of integrating the technology into daily life have now transcended its impact on the individual level to benefit the masses at the community and metropolitan levels. IoT gadgets, which have become a reality to enable us to live in smart cities, are a key component of these new internet-enabled technologies. By bringing about such a paradigm change and opening up new ways to integrate technology into daily life, these IoT gadgets redefine human-computer interaction (HCI). Nowadays, HCI technology is more focused on people than it was in the past, which would have been computer-centric. Due to the HCI's endurance and coherence over the past ten years, there are currently more dimensions where it can be used. Countries began implementing cutting-edge procedures and technologies to manage and control their cities more effectively in order to gain from smart governance and management. For a city to be considered genuinely smart, nearly all necessary aspects of management and government must incorporate these technological advancements into related activities. Making a city truly smart, which is a massive effort with good social, economic, human, and environmental effects, necessitates a

transformation at the design level. The ISO standard database also contains a conceptual model of a smart city, known as ISO/IEC 30182:2017(en) Smart city concept model.It is regarded as the organization's initial move toward regulating the many elements of smart cities in the future. Making a city truly smart, which is a massive effort with good social, economic, human, and environmental effects, necessitates a transformation at the design level. The ISO standard database also contains a conceptual model of a smart city, known as ISO/IEC 30182:2017(en) Smart city concept model. It is regarded as the organization's initial move toward regulating the many elements of smart cities in the future.

2. Industry 5.0 Paradigm

The majority of fields have undergone digitalization in the twenty-first century. We must acknowledge, though, that businesses find it challenging to implement industry 4.0, IoT, and AI technology in their operations. In addition to the aforementioned technologies, industry 5.0, the next phase of the Industrial Revolution, will appear in the future years [2, 6, 10–14]. Early in 2015, the phrase "industry 5.0" became popular; however, it was actually referred to as the "fifth industrial revolution," which had a significant impact on many areas, particularly day-to-day business, due to the speed at which new technologies were being added to the mix and how they were being integrated with human processes [15–18]. At the end of the 18th century, the first Industrial Revolution, often known as Industry 1.0, began. It represented the industrialization of mechanical systems using human, animal, water, and stream power. The second industrial revolution, also known as Industry 2.0, started in the final quarter of the 19th century and was characterized by the utilization of electrical energy for mass production [19–22]. A few characteristics of Industry 2.0 include the development of the telephone, mass production, the telegraph, assembly lines, and industrialization. Early in the 20th century, the third industrial revolution, often known as Industry 3.0, began. In the industrialized field, it introduced computerization and eventually microelectronics. Robots, information technology, and microprocessors are used to achieve a higher level of automation; most of these 20th-century endeavors are intimately tied to information and communication technology (ICT). Among the industries benefiting from the third revolution are computer integrated manufacturing, computer-aided processing planning, computer-aided design, and flexible manufacturing systems. The fourth industrial revolution, often known as Industry 4.0, began in recent years with the introduction of Cyber physical Systems (CPS), which has revolutionizedthe manufacturing industry. CPS, cloud computing, big data analytics, augmented reality, Internet of Things, simulation, and intelligent devices were the main characteristics of Industry 4.0. That is to say, it seeks fully integrated solutions and entirely focuses on end-to-end digitization and embracing digital industrial ecosystems [20–24].

Additionally, it placed a lot of emphasis on IoT devices that connect to the industrial plant. The fifth industrial revolution is more fascinated by cutting-edge human-machine interfaces through human-machine interaction because Industry 5.0 places an emphasis on collaboration between different types of machinery. The primary goal of Industry 5.0 is to advance Industry 4.0 to a more advanced state. It introduces the idea of collaborative robots, or cobots, for this. Cobots will meet the demand for businesses that manufacture individualized items in the present with successful integration [20–24]. Therefore, Industry 5.0 is more familiar in

manufacturing and medical than other related fields because to improved production, software tools, the internet of everything, and robotics leveraging technological advancements.

It provides users the option to experience mass customization through global group collaboration. There is a need for customer goals because technological advancements do not take the organization's foundation into consideration. Industry 5.0 adheres to the following concepts to achieve client goals: Mass customization refers to telling clients how easy and affordable it is to customise different goods or services. Customer-centric focuses on the needs of the customer and attempts to overcome barriers to corporate growth through re - engineering; green computing also emphasises the environment. Cyber-Physical Systems (CPS) prepares an intelligent system from the human servicing the consumers by extracting the greatest benefits from the human with machine intelligence [16–18, 20–24].

Comparison Table from Industry 1.0 to 5.0

Phase	Period	Description	Identification	Key Point
Industry 1.0	1780	Industrialmanufacture based on stream andwater machines	Mechanization Water and stream	First mechanical Loom
Industry 2.0	1870	Mass production with electrical energy	Electrification Division of labor Mass production	First assembly line
Industry 3.0	1970	Automation with electronic and ITsystem	Automation ElectronicsIT systems	The first programmablelogic controller
Industry 4,0	2011	The connected device,data analytics, computerizedmachinery programs to automate the industry production	Globalization Digitalization IoT, Robotics, Big data, Cloud computing	Cyber-physicalsystems
Industry 5.0	Future	Cooperation among human intelligence witha machine to improve products and services	Personalization Robotics and AI Sustainability	Human-robot coworking Bio-economy

Top enable technologies for Industry 5.0
Additive manufacturing, Multi-agent systems, Smart manufacturing, Digital Eco-system, Collaborative Robotics, Internet of Everything, Mixed reality, Industrial Blockchain, Drones, 5G and Beyond

Reasons for adopting Industry 5.0 in Manufacturing

From the discussions mentioned above, Industry 5.0 will advise clients or find a solution for the problems presented by removing human workers from various manufacturer procedures. To

support the industry 5.0 producer, however, innovative technologies are required [16–18, 20–24]. Advances in digital twins and multiscale modeling and simulation have made it more difficult to manage manufacturing sites' profitability. While creating the policies for managing and customizing real products rather than product blueprints, visualization tools are vital. Utilization of compatible small sensor data From smart buildings to autonomous factory cobots and distributed intelligent systems, sensor nodes have greatly increased. The next industrial revolution will inevitably benefit from the real-world raw data that these intelligent sensor nodes sense and collect. Energy optimization, quick and efficient customization, selecting alocal agent for data pre-processing, and developing highly simulated distributed intelligence in IoT, Industry 4.0, are still unresolved research questions. With the ongoing development of Big data and AI-based robots, it is now even more possible to create digital twins that are more realistic in virtual reality. It is appropriate for professionals in the sector to enable less waste in the process flow and system design. Therefore, the digital twin with cutting-edge visualization methods will significantly boost throughput across all sectors. Real-time trackers will facilitate customers' sales orders with manufacture orders and additional material, enhancing real-time production tracking. When a trainee or teacher is in a different area but is learning a particular job in a virtual or simulated environment, virtual training will be helpful. The costs and time for both parties are significantly reduced with this kind of training. Artificial intelligence models andintelligent autonomous systems are very useful for autonomously managing production lines in the manufacturing sector. Modern AI-related ML and DL models successfully alter intelligent systems and solutions that support scenarios requiring decision-making. The aforementioned schemes are securely and progressively guided by transfer learning - transfer learning rules in industry 5.0.

In addition to sensory abilities, computer vision has showed considerable potential in reproducing primitive vision using DL and RL and GPU-based computation. However, cobots' proficiencies need to be increased subtly in order to achieve industry 5.0's advanced performances.

Industry 5.0 issues and restrictions

Most of the industrial problems brought on by eliminating human labour from various processes are solved by Industry 5.0. However, since humans might eventually add cutting-edge manufacturing talents, it must also include other forward-thinking abilities. There are many abilities that are still being developed, some of which are mentioned in this section.

1. There is a need to understand how an autonomous system can incorporate ethical standards before sophisticated talents are incorporated into industrial management.
2. In the autonomous system model, ethical behavior needs to be properly verified and validated.
3. The overproduction phenomenon may be impacted by implementation operation transparency and quick, competent manufacture.
4. In an autonomous scheme, the outcome results must be transparent ethical behaviour solutions. Industrial specialists in particular are having trouble adapting and implementing new ideas.

5. Validation and tuning will prevent certain moderately important issues between firms, society, experts, stockholders, and technology.

3. Elements of IoT

Understanding IoT building pieces can help visualise and provide a clearer understanding of the IoT's actual meaning than functionality, as we indicated in the opening [23]. We identified six core components of IoT, which are shown in Figure 1.

Identification

Identification is a crucial concept in any network used for data or communication. The IoT framework relies on exact identification to name services and connect them to their claims. The IoT system makes it difficult to address object ID and its corresponding IP address. A device's name is indicated by its ID, and its location inside the network is indicated by its address. As identification models are not unique and since objects may experiment using public IP addresses while connected to the network, differentiation between object identification and addresses is necessary. Therefore, the developed models must overcome the difficulties outlined above and accurately identify each object within the network.

Sensing

IoT setup aims to assemble data from a specific area or region using sensing devices. Sensors, wearable sensors, and actuators used primarily for sensing purposes are examples of things or devices that collect real-world information from the device and send it back to a database or the cloud for further processing. To understand IoT products, for instance, single-board computers (SBCs) like the Arduino Yun and Raspberry PI that have sensors, integrated TCP/IP, and safety capabilities are a natural choice. To deliver the necessary data to clients, such devices typically connect to a central management portal.

Communication

In general, the majority of IoT objects have sufficient resources; yet, due to these resources' limitations, connections with heterogeneous devices and objects sometimes include lossy, noisy signals. Wi-Fi, Bluetooth, NFC, RFID, and IEEE standards are some IoT communications; a quick explanation was given in the following section to help.

Computing

Another crucial IoT concern is the hardware's processing power.The computation-related parts of an appliance, such as microprocessors, microcontrollers, and software-oriented appliances, serve as its brain. Hardware platforms made for Internet of Things devices include Arduino, Rasberry PI, UDOO, MULLE, and Gadgeteer. Other platforms include Cloud platforms for processing large amounts of data in real-time, Real-Time Software Operating Systems (RTSOS) for real-time IoT tasks, and Tiny-OS for light-eight operations. Nevertheless, the research community has concerns about several of the computational components.

Services

IoT provides a huge range of services. Most of them are classified into identity-based services, which include most real-time appliances, and information aggregative services, which collect raw sensor data from the actual world and connect it to the right IoT applications. Aiming to depict collaborative systems that may operate anytime clients need them, collaborative-aware services utilize the gathered data for data analytics for decision making. However, the

abovementioned services are not yet at or reachable to a comfortable stage; numerous difficulties in addition to challenges need to be resolved.

Semantics

The IoT uses semantic operations to effectively abstract information from various objects. Finding resources that boost model performance is comparable to knowledge extraction. Some of the well-known semantic technologies used in IoT systems are Resource Description Framework (RDF), Wide Web Consortium (W3C), Efficient XML Interchange (EXI), and Web Ontology Language (OWL).

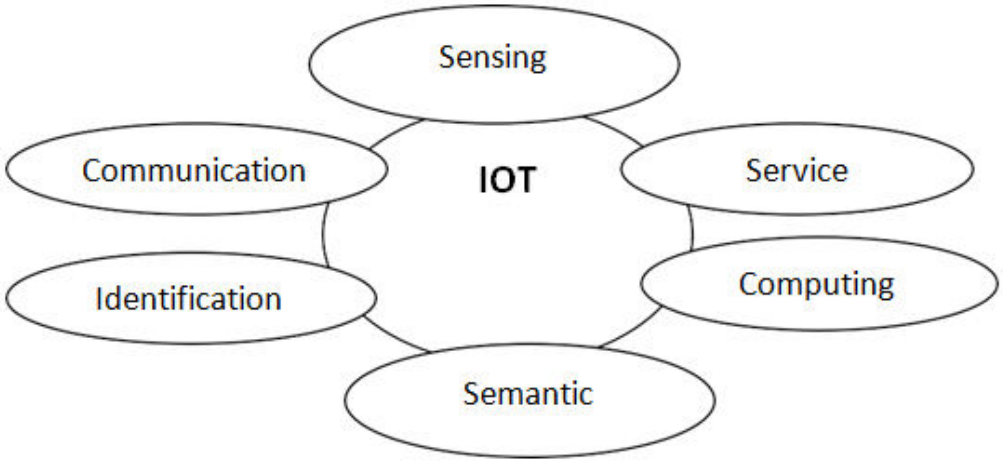


Figure 1: Internet of Things Elements

4. IOT Architecture

By integrating IoT and its variants into a variety of disciplines, businesses will improve their products' or employees' productivity. However, as there are so many different devices, protocols, and working environments, these proposals are severe and challenging to put into practise in the actual world. In other words, it is certain that during this stage, the issue of developing an uniform IoT architecture will arise. Understanding the variables that influence

IoT behaviour is better before creating IoT architecture since it makes it simpler to locate dependable IoT solutions. Additionally, it will cut down on the resources used in IoT design. It is essential to understand what this concept entails before unveiling the problems and offering a clear architecture of this creativity [23–25]. IoT architecture is essentially a combination of excellent fundamental network tools. It was measured as a system of interconnected worldwide networks made up of various allied devices that rely on equipment for information processing, networking, and communication. Look at Figure 2.

4.1 Perception Layer

IoT is a type of physical global network in which devices can connect and be remotely measured. The perception layer, which is like a bridge between the physical and digital worlds, is thought of as the first stage for IoT initiatives. recognized as a sensing layer in some circumstances. The perception layer spends the majority of its time talking about wireless intelligent devices such smart sensors, tags, and actuators. These wireless systems with tags or sensors are now capable of automatically sensing and exchanging information between various

devices. Devices can range in size and operation, from small to improvised vehicles. Sensors gather environmental data, convert it to electrical impulses, and then transmit it to IoT systems. Electrical signals gathered from an IoT system are transformed into physical actions by an actuator.

4.2 Connectivity Layer

Since it handles all communications between the various IoT devices, systems, and cloud computing facilities, the connectivity layer is regarded as the second stage of the IoT strategy. TCP/UDP or software/hardware modules can be used to connect the physical layers to cloud data centres. Ethernet links fixed IoT devices, Wi-Fi provides ubiquitous wireless connectivity used in residential IoT setups, NFC transmits data between two devices, and Bluetooth transfers tiny data files but is ineffective for transferring large data files. IoT uses message-oriented protocols depending on the application's need for data communication in some uncommon cases. Among them are Message Queue Telemetry Transport (MQTT), Data Distribution Service (DDS), Constrained Application Protocol (CoAP), Advanced Message Queuing Protocol (AMQP), and others.

4.3 Edge or fog Computing Layer

For IoT systems to handle the growing number of connected devices and real-world applications, edge/fog computing is essential. Edge/fog computing is intended to store and pre-process sensed data as quickly as feasible and as close to its sources as is practical. As a result, itcan help IoT devices save time and resources. It will also cut down on system latency, which can increase performance accuracy. Typically, edge/fog computing occurs on gateways, which are regional computers dispersed throughout the network.

4.4 Processing Layer

All of the data from the IoT schemes is gathered by the processing layer. To employ abstract information for decision-making or to make data available for any further processes, preprocessing models are applied. APIs are used to monitor real-time data, which functions as a hub for both event-based and query-based data intake. Only other connected devices can currently understand the data after it has been collected from multiple devices using data abstraction techniques.

4.5 Application Layer

Software programmers were used for data analysis in order to bind appropriate responses to the application layer's key business queries and requirements. Numerous IoT requests range from simple to complex and function, requiring various skill sets and working paradigms. Various apps are now built directly on top of IoT platforms, suggesting software-related advanced setups through ready-to-use tools for data mining, pattern recognition, and forward- thinking analytical abilities.

Commercial Layer

Systems for problem-solving and making decisions can only benefit from the data gathered and pre-processed in IoT schemes. For this reason, the business layer is clearly defined as a separate stage that is complex and difficult to represent in a single application layer.

Security is a crucial concept in any network-related application. The security layer, which protects all the services and levels outlined above, is essential in the Internet of Things. It is

challenging to discuss IoT security issues in a single paragraph or section. In IoT systems, there are various security levels:

Device security - IoT-related devices require chips that can boot procedures, physical metal shields, low-resource authentication services, and authentication services to prevent unauthorized code. IoT data transit is usually done over wireless channels, making it simple for hackers to steal or manipulate the data. As a result, data delivered through a network or device must be in an encrypted format. Cloud security - sensing data stored in the cloud must be encoded to reduce risks of disclosing sensitive data to intruders. To ensure that security is high at all levels, from the smallest devices to complex analytical systems, always pay attention to security protocols.

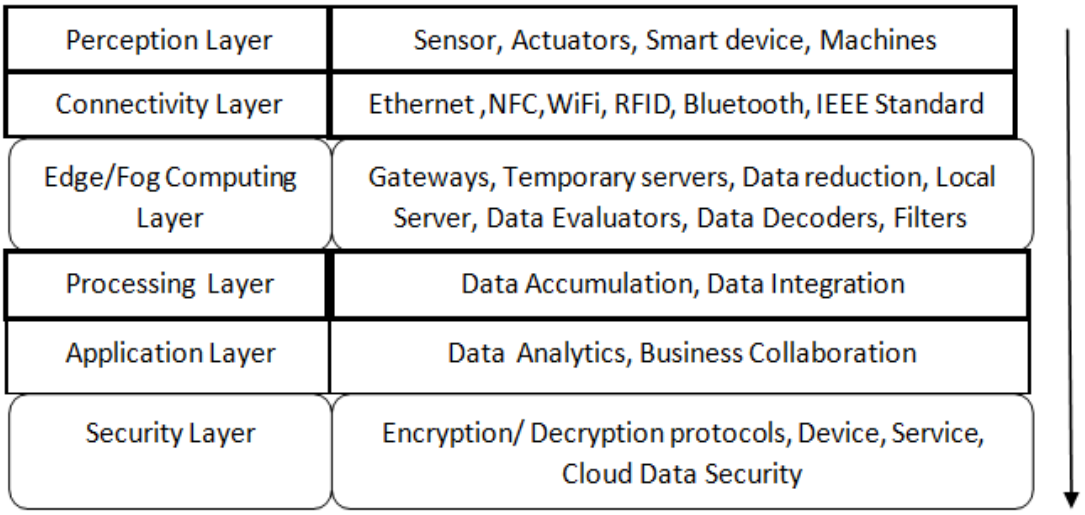


Figure 2: Internet of Things Architecture

5. ENABLING TECHNOLOGIES

5.1 The Radiofrequency Identification (RFID) Technology

RFID communication technology, which consists of tags and scanners, was created specifically for tracking transportation. RFID is regarded as an automatic identification system that automatically matches the target tag signal with relevant information. As a result, it was widely used in a variety of hazardous and passive environments. The RFID structure is complete with tags and readers, as we discussed earlier. Address bars that are connected to objects serve as the tag, which is handled by the antenna via a tiny microchip. To transfer and retrieve data records from an entity over a tag, the electro-magnetic pitch is being used. Only when both tags and readers are situated at a specified angle or range may readers read or abstract the data records stored on a tag. by transmitting the proper data, signal The antenna on the patch receives the signal that the reader sends in order to read the information from the tag and acknowledges it by providing the necessary data. Three RFID communication tags are listed below: passive tag - collects signals from battery-operated tags, Active tags do not require batteries because they derive their energy from reader signals. The last option is active reader active tag, which operates on both high and low frequencies. Since RFID tags automatically track payments, goods or baggage, inventory management, product monitoring, and product lifecycle

supervision, they are expertly applied to real-world appliances. The tags then update the information without the involvement of a third party or a human. RFID technologies can be used in a variety of fields to build and improve the usability and effectiveness of models and systems. The majority of IoT WSN appliances are built in difficult conditions where the signals are disrupted, intercepted, and there is a possibility for the entire device to crash, thus there are some disadvantages to the adoption of RFID.

5.2 Actuator

Actuators are used in specialized appliances, and they are most effective when the appliances are moving. It generates a range of motions, including rotary, spherical, linear, and oscillator motions, and then converts them into kinetic energy to produce electricity. Actuators come in three different varieties: pneumatically based, which uses compressed air, hydraulically based, which uses hydraulic fluids and electrically based, which uses motors.

5.3 Near Field Communication

When two objects are close to one another, NFC technology is used to transmit data and build up a small communication arrangement. Although it uses touching or two items that are closer to the exact position, it is comparable to radio communication. The size of the object's antenna affects the NFC's communication range. As a result, NFC technology is typically not advised for remote areas. It is also not safer due to its flaws, which make it easily exposed to attackers.

5.4 Wireless Fidelity (Wi-Fi)

A good fit for data-intensive IoT-based applications, Wi-Fi is a well-known wireless network capability. With an intelligent transportation system, it has high wireless access for a compact area. It has several collective versions, including: A transmission rate of 54 Mbps is provided via IEEE 802.11a, and data rates of up to 2.4 GHz are also available.

5.5 IEEE 802.15,4

The MAC layer is a sublayer of IEEE 802.15.4, often known as Low-Rate Wireless Personal Area Networks (LRWPANs). It offers efficient communication for high security, low cost, low power consumption, and supports several sensor nodes simultaneously. Numerous communication technologies, including ZigBee, Z-Wave, Bluetooth, etc., are based on the IEEE802.15.4 specifications. Although this is an interesting subject to explore, it does not deliver QoS.

5.6 Bluetooth LE

Information communication between stationary and mobile devices over a short distance using Industrial, Scientific, and Medical (ISM) bands is known as Bluetooth or IEEE 802.15.1. The smart home, smart city, healthcare, security, military appliances, fitness, and sectors all strongly included it. The most recent versions of collecting and aggregating sensed data from IoT-based sensor nodes include Bluetooth SIG, BluetoothBLE, Bluetooth 4.0, and Bluetooth 5.0. Short-range monitoring devices were a perfect fit for Bluetooth technology.

5.7 ZigBee

The main goal of Zigbee's innovation is to broaden the IoT and WSN application domains. It is a unique sort of adaptable wireless networking technology that performs better with short-range products including intelligent home automation, medical equipment, and industrial equipment.

In addition to having a four-layer architecture made up of the physical, MAC, network, and application layers, the ZigBee design incorporates MAC and IEEE protocols.

6. Artificial Intelligence (AI) in the Internet of Things (IoT)

The "Internet of Things (IoT)" is rapidly replacing the "Internet of Computers (IoC)" as the primary operating system for the internet. To enable intelligent internet communications, it is necessary to think carefully about the significance of AI techniques. Because of their fantastic remote monitoring of occurrences in areas like healthcare, weather reports, seawater levels, event predictions, etc., wireless sensor networks are currently a popular study issue. Additionally, intelligent sensors were widely used in electronic-based mobile devices, smart cities, and household appliances [22–25]. The IoT concept is "radio-frequency identification (RFID) tags, sensors, actuators, mobile phones, etc. are only a few examples of the numerous items or things that are constantly present around us. They may communicate with one another and work with their neighbours thanks to special addressing systems "[36]. As a result, the concept of IoT is constantly evolving and will have an impact on people's lives all over the world. IoT robots can be built with the use of AI techniques that enable them to develop roles devoid of constant human control [24–25].

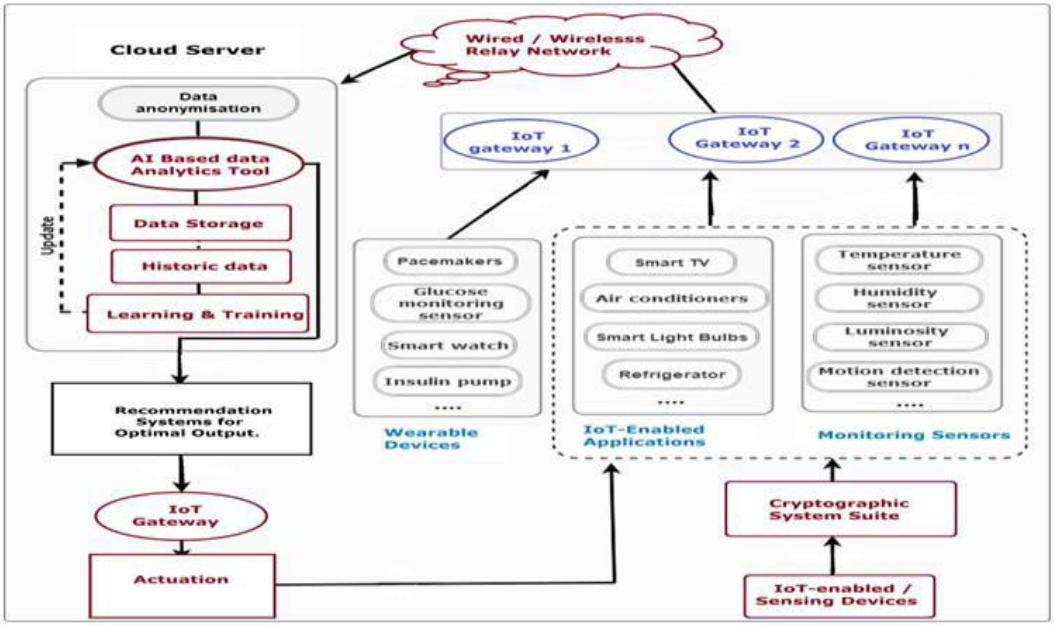


Figure [3]: IoT Data-workflow diagram

The IoT data-flow diagram is shown in Figure [3]. Data first comes from IoT-capable hardware and IoT appliances, and then it travels through IoT-Gateway to a cloud-based server. Here, data has been studied using a variety of analytical tools, as well as training and learning techniques. Then, recommendation algorithms enter the scene to suggest the best course of action; actuators are present to direct the flow toward IoT Appliances for additional processing. But wecan assert that new IoT standards are essential for transferring from today's networks ofintelligent sensors are allowed in systems with actuation-type machinery. The "Internet of Intelligent Things (IIoT)" will be used in these types of emerging projects. These networking advancements over time have led to the availability, durability, and intelligence of today's internet connections.

Giving familiar things the ability to interpret their contexts and draw conclusions under their own appears crucial [22–23]. At the moment, there is no requirement to submit judgments or conclusions to central decision-making nodes. The IIoT is enabled to respond to enhanced time-critical conditions since the conclusions are reached in a non- centralized manner, which is made possible by the huge intelligence of sensors and allowing them the ability to turn by providing to the incentive professed by sensors.

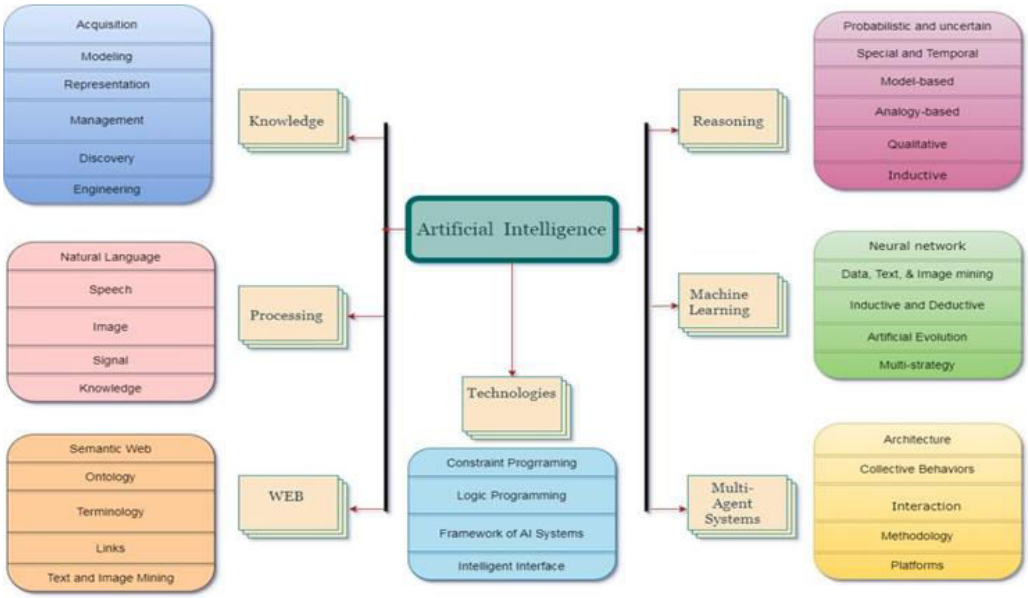


Figure. [4]: Artificial Intelligence Classification

Data collection is made easier by the Internet of Things since it acts as a catalyst, making it one of the most useful resources available right now. As demonstrated in Figure [4], enormous amounts of unstructured data can be quickly sorted through to produce industrial insights and well-informed judgments when combined with another accelerator, artificial intelligence.

6.1. Intelligent Sensing Using Artificial Intelligence

When discussing AI for intelligent sensing, ML models are used to identify important trends or projections in the data gathered by the intelligent sensors. For instance, class volumes identified by the model are dynamically increased by active sensor learning. A specified technique for data collecting must come after episodic retraining or thorough querying because the data is composed in a real-time setting [21–22]. Active earning systems that try to categories events without prior knowledge using a variety of diverse ambient sounds are best suited for unsupervised models.

6.2. An Iot Decision Tree

By using sorting strategies based on appropriate features, DT addresses categorization challenges. Information gain and the Gini index are two of the many DT algorithms used to identify the most desirable feature that divides the training instances most effectively. The entire process of a DT goes as follows: first, pre- and post-pruning are used to shrink the tree. Second, corrected space between the objects was searched. To remove the duplicate features, an optimized search model was used as the third step. Fourth, the structure of the resulting tree

was converted into a suitable data structure, similar to a collection of rules. DT successfullyused real-time IoT applications such pattern recognition, decision making, environment monitoring, security parameter detection, healthcare management, etc.

6.3 Iot Random Forest

The supervised learning models category includes Random Forest (RFs). RF is made up of a lotof trees that are built randomly and then skilled to cast a vote for a good class. The ultimate classification outcome is determined by which class received the most votes. The outcome classification is the average of the DT results because decision trees are typically used by RF to generate subgroup rules for voting. Additionally, where it requires the fewest input parameters, RF computational accuracy defeats feature selection, but it is inapplicable in real-world applications. IoT devices across many areas could benefit greatly from RF models.

6.4 Clustering

The fundamental objective of K-means is to group unlabeled data features into K clusters or sets; in this case, data points fitting to the same cluster must have some similarity. K-means is often a very scalable and quick machine learning algorithm. In other instances, used MapReduce to evaluate the various small datasets and then gives a cluster solution for a high dimensional of small data based on the K-means method [19–20]. The K-Means cluster was created by researchers, who subsequently categorized consistent travel patterns.

6.5. Support Vector Machine of One Class (Ocsvm)

The semi-supervised technique known as OC-SVM is an addition to SVMs. When the new data after some operations lie outside the boundary line and are noted as an outlier or anomaly, it forms a boundary line among the training data. Due to the nature of their operation, OC-SVMs are beneficial for assessing the results of IoT-based machines, WSN anomaly detection, and network intrusion detection.

6.6 Ensemble Learning Models in Iot

Combining different fundamental classification techniques, ensemble learning (EL) generates a group's effective output. Research on EL trials demonstrates that different learning models are used for different applications. So, in order to increase precision, the research community begins merging multiple disparate classifiers. Additionally, EL models employ a variety of learning strategies that reduce variation and are resistant to over-fitting. In IoT-based environmental datasets, EL has been successfully employed for online intrusion and anomaly detection, as well as for assessing real-time datasets for precise IoT device decision-making.

6.7 Neural Network

Nonetheless, finding the best weights for an FFNN falls under the category of an NP-complete task. Adaptive moment estimation, stochastic gradient descent, adaptive slope, Nesterov's accelerated gradient, adaptive delta, and RMSprob are only a few of the learning techniques used by the model. In the Internet of Things, FFNN can be used as a solution for issues including energy management, decision-making, feature selection, and computing complexity reduction.

6.8 Support Vector Machine (SVM) in Uiot

By calculating the distance's data properties, SVMs conduct classification by constructing the splitting hyperplane between two different classes. Large datasets with several feature attributes

and few sample points are best suited for SVMs. The fundamental benefit of SVMs is their ability to detect intrusions in real-time and then dynamically inform training patterns. Many security applications, including intrusion detection and outlier identification, use SVM versions including QS-SVM, CESVM, and SVDD, which are also efficient in-memory storage solutions with low time complexity.

6.9 Principal Component Analysis

The primary subspace is an L-dimensional linear subspace that Principle Component Analysis (PCA) orthogonally plans data facts into. Data visualisation is one of the applications of PCA, which works with high dimensional datasets based on reiterative expectation expansion method and data compression. Therefore, PCA is regarded as the most significant pre- processing step in ML. The PCA version known as Canonical Correlation Analysis (CCA) deals with two or more variables. Here, identifying a pair of extremely cross-correlated linear subspaces that are consistent is the major objective. As a result, each factor and a single component from the other subspace are related within one of the subspaces.

6.10 Bagging

The goal of bagging is to first reduce over fitting while improving the accuracy and stability of ML-based approaches. Using this technique, training datasets are created by randomly choosing data points from the one and only training set. An ML technique is therefore trained on each originally created training set. The bagging method improves the results for many ML algorithms, including DT, RF, and neural networks.

6.11 Neural Network

The models that process new data instances the quickest are neural networks (NN) with condensed representations. NN features a variety of NNs with unique structures and equipment thanks to advances. The most prevalent kind of neural network used in functional appliances is the feed-forward neural network (FFNN), also known as multilayer perception. The nonlinear function, also known as the active function, in FFNN determines the activity of each layer. With enough hidden units, an FFNN with at least two hidden layers can estimate a random mapping from a finite input space to a finite output space. However, finding the best weights for an FFNN falls under the category of an NP-complete task.

6.12 Machine Learning for Analytical Abilities (Iot)

For many years, many business organisations have employed analytical talents; today, many firms are attracted to planning their AI capabilities. Organizations and businesses combine their expertise for effective data use, statistical analysis, and quantitative processes to advance decision-making from the last several decades. But at the moment, those businesses are mostly focused on developing and using AI to complement one another. Like ML and DL, which swiftly increase supremacy in addition to demand, AI is not statistical [22–25]. Administrations with analytics-oriented clusters may choose to focus their attention primarily on these machines or learn new abilities in non-statistical areas. The innovation analytics has evolved into several different forms, some of which are included here. Analytics 1.0 is a time of artisanal expressive talent, the advent of examination, and the invention of writing implements. Years of commercial analytics dominance throughout this era, with price mostly based on goal of internal decision provision rather than developing analytical skills Analytics 2.0: At this

point,big data analytics are comparable to Hadoop, and the emergence of data gurus is a result of information-based breakthroughs like Google and Facebook. Analytics 3.0: Data is created by massive organisations, followed by analytics-based productions and finally analytical events utilising various ML models. Analytics 4.0: Numerous firms have heavily incorporated AI and cognitive based models to increase analytical sophistication. It adopts multiple model accuracy levels, applies AI models, and makes excellent use of self-rule in the execution of automated ML approaches. A combination of talents and internal collaborations require AI, which is one reason for its adoption into analytics [23–24]. For instance, comprehension of the embedded learning data models is necessary for computer knowledge. Another factor is the need for advanced statistical models, massive data processing, and reliable data analytics. When values are created and their outcomes are known, supervised learning is used in machine learning. Because AI has the capacity to change business models, it will play a significant part in analytics

4.0. As a result, the impact of analytics 4.0 will likely be more disruptive than previous automation evolutions. Additionally, businesses that adopt analytics 4.0 do so more quickly than those that do not use any AI models. The first step in comprehending AI achievement is to analyse how AI will affect creativity, new abilities, and what kind of practical act policy should be used. Businesses that have control over their current analytical capabilities can implement AIconsiderably more quickly and actively.

6.13 Deep Learning for analytics, (IoT)

IoT-based devices gather enormous amounts of sensed data from the surroundings in which they are deployed thanks to the development of numerous networks and small technologies. Additionally, these IoT devices and objects will produce quick and real-time data streams depending on the applications. Here, using analytical models on such large data streams is essential for locating unique information, predicting future structures, and then taking control of outcomes. It turns IoT applications into a well-intentioned business standard and a skill that improves quality of life. The majority of IoT devices have been developed over the last few years with input from several research areas, including the military, smart cities, healthcare, and agriculture. Intelligent learning techniques for prediction or data analysis models are responsible for these applications' success. With the examination of various ML strategies, DL has been aggressively applied in many IoT appliances today. Future applications are expected to make extensive use of the DL-IoT combo. Because it swiftly handles the emergent analytical services required in real-time IoT equipment, DL is primarily used in classical ML [20–24]. The expansion of IoT requires stakeholders to define its purpose, foundation, capabilities, and problems in addition to its variant's big-data derivatives. IoT and bigdata work closely together because IoT is an important information provider for bigdata.

To better understand IoT-based data analytics requirements, it needed to determine the features of IoT data and how they dissimilar from those of big standard data [20-24]. Some of them mentioned below:

1. Large-Scale Streaming Data: When IoT is implemented using a large number of dispersed devices, a large amount of streaming data is produced because IoT applications collect a lot of data.

2. Heterogeneity: The Internet of Things (IoT) is a heterogeneous network of linked devices, andas a result, different IoT data acquisition device assemblies produce heterogeneous data.
3. Relationship between time and space: At the moment, the majority of IoT devices are based in the real world, where sensor devices are involved in a specific place and have a position and time-stamp for every single data substance.
4. High noise data: Before applying them to any decision-making systems, they need to be eliminated due to dynamic environment changes, minute mistake bits, and noisy data producedin IoT requests. If not, it will damage the outcome results. While using big data to gather privateinformation is a clever way to enhance the quality of our lives, it is not an easy task. To handle such complex and thought-provoking jobs, it is necessary to move beyond the capabilities of out-of-date inference learning models and adopt cutting-edge abilities, practises, and infrastructures.

6.14 Edge Computing in IOT

IoT connected objects typically produce enormous amounts of data, which is collected and processed at one of the appropriate/suitable objects to create meaningful information. Since the courts of enormous sensing then boost information kept in IoT, big data support IoT applications. As a result, today's entire IoT setups employ substantial data operations. Additionally, due of the heterogeneous connections, IoT gathers unstructured, multivariatedata that requires additional analysis to be able to isolate the useful information[23–24]. The Internet of Things (IoT) is quickly emerging as the next digital revolution, but it will require extensive data processing, storage, and analytical skills. Utilizing the application platform as a service to send the data to the cloud is one likely solution. These days, cloud computing is a well-known technology that provides computing resources or online data storage. Cloud services are offered by IT firms like Google Cloud, Amazon Web Services, and IBM CloudAnalytics. The capacity to store and subsequently exploit sensed data information is only one benefit that cloud computing offers. However, IoT-related appliances must be capable of storing, managing, and analysing data from massive numbers of objects spread across a wide geographic area. However, new situations will arise as cloud computing is used in IoT. Cloud computing cannot outperform fog or edge computing in terms of speed of adoption or help with the aforementioned problems. In short, fog/edge computing provides processing then storage facilities to devices at the edge of the system as a replacement for executing whole computational activities at the core of the cloud. Any network node or item having fog computing capabilities may carry out data storage, computation, and heterogeneous network connectivity tasks effectively. These tools/items are used anywhere on the network to link IoT objects and devices to apps [23–25]. Typically, various types of data are gathered from IoT items and transferred to the appropriate object or place for further analysis based on the application requirements. While integrating IoT with fog/edge computing, there are restrictions and tribunals in addition to the benefits. The most important step is setting up fog/edge computation and allocating enough resources to IoT devices. IoT devices only ever use a small subset of services, hence each fog/edge service node has insufficient connection, processing, and storage capacity. Every fog/edge computing node should be configured and accomplished optimally in this context so that IoT devices can supply the requested service resolutely. It can be difficult to figure out how to change how a fog/edge node allocates its resources. This

indicates that the current hot study area in IoT fog/edge computing is managing resources between the fog/edge computing, Therefore, it is necessary to confirm the many requirements, such as energy consumption, node cot, and service availability, before introducing fog/edge computing nodes to a specific service. Problems with security and privacy in the fog/edge computing architecture are equally serious issues.

6.15 Federated Learning

In federated learning (FL) [14–24], an algorithm is trained across a number of distributed edge devices or servers that store local data samples without exchanging them. Graph. [5]

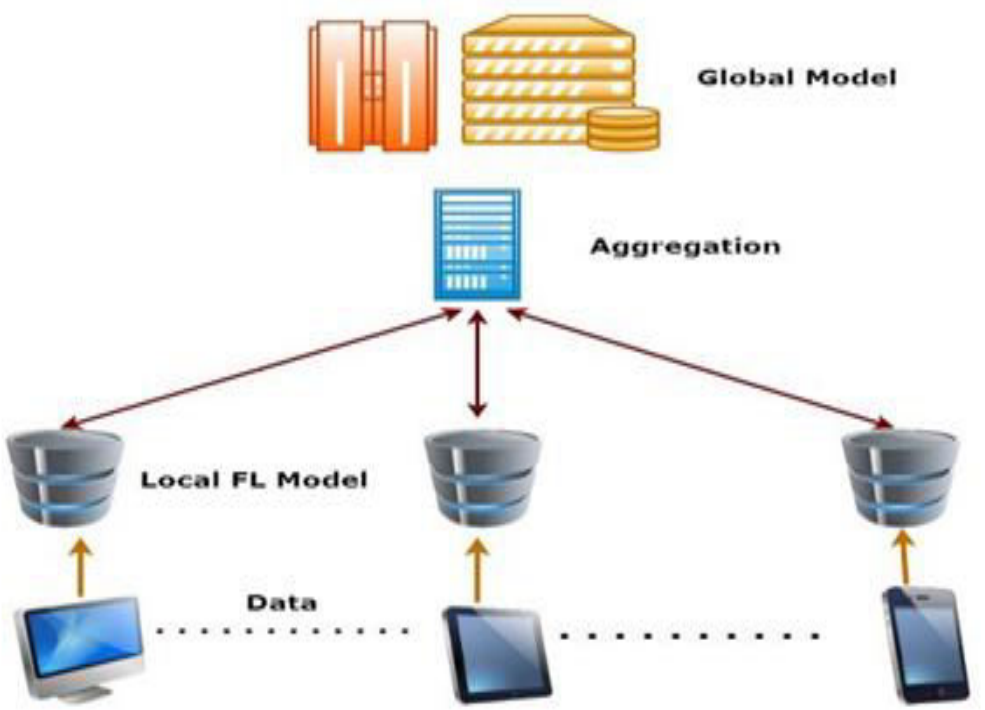


Figure. [5]: Fundamental FL Architecture

FL Architecture Fundamentals Users train local models using local data to update the global model at the base station, as depicted in Figure [5]. The local models are trained using the aggregated data from the global model. Up until the global model converges, these processes keep happening. Devices can collectively learn from a common model thanks to federated learning. The shared model is initially trained on a server using proxy data. Each device then downloads the model and updates it using federated data. The system trains the model with data that is readily available locally. The updates created by the model's adjustments aredelivered to the cloud. The training information and individual updates are stored on the device[18–19]. To ensure quicker downloads of these changes, the model compresses using random rotations and quantization. The models combine to create a single model when they are communicated to the server by the devices. Until the model was of high quality, it was repeated several times.

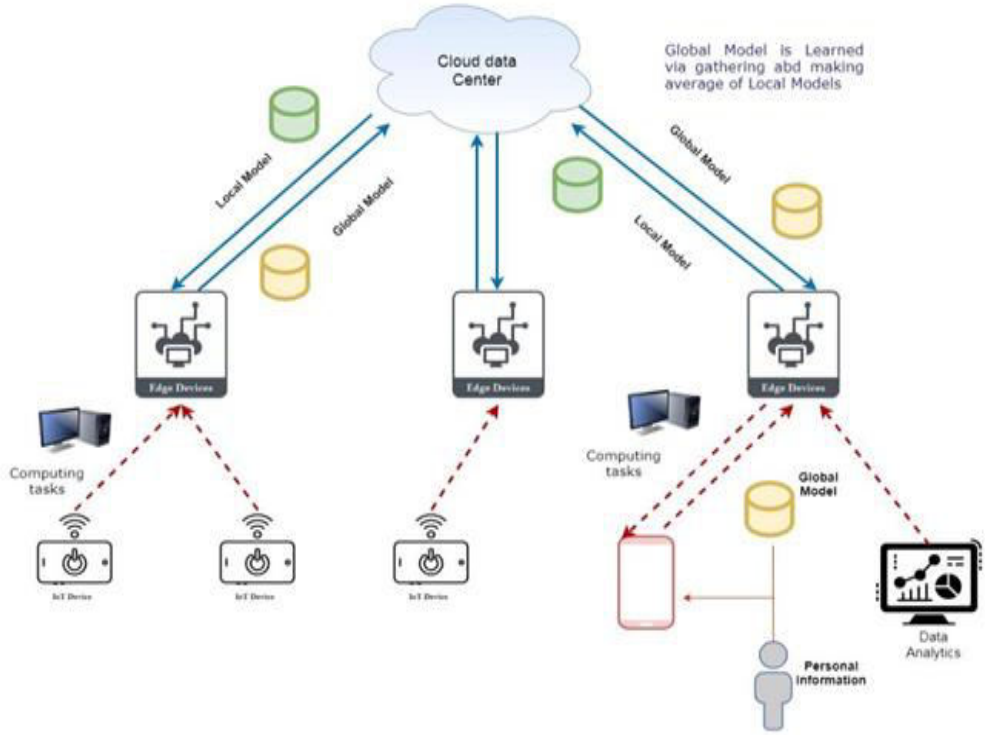


Figure [6]: Process of Federated Learning

The method for federated learning as displayed in Figure [6] is as follows:

1. The devices will receive a training model.
2. The equipment is set up to learn from local data.
3. The devices provide encrypted updates on the parameters to the server.
4. The server groups the devices. To update the current model for each group, the server combines the updates it receives from each set of devices.
5. A new round of training is conducted when the new updated model has been sent to the devices for on-device testing (again, the idea of decentralization is at work here and fresh set rules follows.

Advantages of Federated Learning

Here, we'll talk about a few FL advantages [24]. First of all, FL enables mobile devices to jointly build a prediction model, saving the training data locally rather than uploading and storing it on a central server. Second, it extends model training to cutting-edge technology, including smart phones, tablets, the Internet of Things, and even "organizations" like hospitals that are subject to strict privacy laws. Maintaining personal information local has important security advantages. Thirdly, because the forecast is made by the device itself, it enables real-time prediction. FL removes the delay that results from transferring raw data to a central server and then returning the results to the device. Fourthly, even without an internet connection, the prediction process can go on because the model is stored locally on the device.

7. IoT Systems' AI-based Trustworthiness

Many of us believed that the integration of IoT makes human life more comfortable and stress-free. Because it contains malware, copyrights, spam, and other unwanted content, some experts have claimed that IoT stands for "Internet of Garbage." But it grows with greater communication, better handling, strict moderation, and efficient community management. The most important step, after gathering data from the trash liked network, is determining acceptable value. It is common knowledge that IoT is expanding quickly and has new requirements. Big data analytics, real-time monitoring using streaming data, and another crucial topic that has come up in the discussion above are effective communication capabilities and ensuring security requirements in such a vast network. The software programmes that have been implemented should have secure network connections as well. Since their data is available on a network, customers and employees of smart IoT products will be particularly vulnerable. Data confidentiality, privacy, and trust are the three main problems with IoT devices and services. Before beginning data exchange and service access, the IoT object or device must obtain authorization from a body or individual. Cybersecurity refers to a strategy for protecting IoT systems and the parts that make them up. IoT-based cyber security systems primarily prevent attackers from stealing critical data, making cybersecurity protocols particularly crucial when working with small devices. There are various cybersecurity techniques, including secure socket layers, firewalls, antivirus software, and intrusion detection systems. IoT schemes can greatly benefit from the application of ML, DL, Blockchain, and quantum-resistant crypto methods for improved security. Additionally, there have been some recent problems, such as the collection of user data by small IoT wearable devices that connect to the databases of device providers [23–25]. The user's consent is not obtained before these device providers sell

the user data they have collected to other commercial enterprises. Business organisations send that specific person ongoing notifications and adverts via social networks based on the information. The biggest difficulty, other from security constraints, is how to prevent these forms of data ethics in IoT-based programmes.

8. AI tools for IOT

S.No	AI Tools	Features
1	iFogSim [53]	1. Network Communication canbe done. 2. Mobility and edge-processingcan be simulated. 3. Some parameters like energy efficiency, network protocols, and heterogeneity cannot be exhibit.
2	IoTSim [53]	1. In IoTSim, IoT Devices are modeled, and performance analysis realized. 2. But, Edge devices, energyefficiency, mobility, communication protocols cannotbe model.
3	IoTSim-Edge [54]	1. It allows researchers to modelmobile IoT devices. 2. It allows researchers to modela variety of IoT protocols. 3. It is in favor of a high-energy-consumption profile. 4. It allows for the abstraction ofgraph modeling.

9. Internet of Things applications

Applications for the Internet of Things (IoT) bring enormous value to our daily lives, as was said in the introduction. Every day, a new IoT-enabled product claims to have the cutting-edge advancements in wireless networks, sophisticated sensors, and revolutionary computing capabilities. IoT applications aim to provide billions of common objects and things connectivity and intelligence. This section makes an effort to provide an overview and discussion of a wide range of IoT-related domains, including intelligent homes, structural health monitoring, the environment, logistics, agriculture, health, and industry domains.

9.1 Augmented Reality

Augmented reality (AR) improves how people need, understand, and present information without interfering with the real environment. Mobile Augmented Mobile Reality (MAR) adds value and improves the interface with reality by superimposing virtual objects over basic materials on the screen. By enabling personnel to access the most essential sensor data in the control panel, such as the display option, it can improve productivity and manufacture services. A helmet that may shield employees from falling objects and help them avoid mistakes was created by US-based DAQRI. In addition, the DAQRI equipment excels at diagnostics in addition to detecting dangers via thermal vision. The manufacturer of heavy equipment, Caterpillar, employs AR technology to look at the machine and quickly see a visual overlay that indicates when certain mechanisms need to be replaced and when filter operations need to be performed as well as how much gasoline is required. The Bosch firm used AR to generate overlay text, films, and augmented 3-D simulations over a piece of equipment because user manuals and technical papers are notoriously boring.

9.2 Virtual Reality

The number of devices linked to the Internet is significantly rising due to the ongoing expansion of virtualization. In contrast to standard television systems, virtual reality (VR) has a greater potential to disrupt the industry. However, VR has ultra-high-definition with obvious, dynamic changes that pose substantial hurdles for fulfilling such promise. Virtual reality is a key component of smart cities, and China has already developed VR-based smart cities with virtual and physical rules for emergency department fire monitoring. Japan simultaneously launched the Tokyo Virtual Lab, which integrates street and traffic data to mimic traffic problems while also assisting drivers in emergency situations.

9.3 Mixed Reality

Mixed reality (MR) mixes VR and AR services to provide real-time interactions between physical and virtual things. According to recent studies, corporate investments in MR will surpass 4.4 billion dollars by the end of 2020. Microsoft HoloLens with MR capabilities as well as wearable holographic computers are employed in the teaching and training stages. Professionals may easily shape their work in a shared virtual environment with 3D modelling with MR. In the field of healthcare, MR has several uses for training and teaching, including procedures that are remotely taught by experts as they perform them in real-time.

9.4 Smart Lock

Operators are now able to abandon conventional locks and generate interest in smart locks thanks to IoT in smart home security. A substitute operator can unlock the doors using

biometric data such as iris scans, fingerprints, and facial mapping as smart locks don't need a physical key to operate.

9.5 Smart Factory

Enterprise asset management is a component of smart factories. IoT-based power-driven asset management improves operational efficiency, optimizes resources, and more effectively manages the sales lifecycle, compliance processes, and welcoming bright environments. One IoT smart factory platform that allows for energy management to enhance organizations' energy use is WebNMS.

9.6 Smart Grid

IoT is creatively used in the production of electric power to track the energy output of multiple power plants. Additionally, IoT-based systems are successfully used to monitor substations, towers, electricity use, and dispatch lines. By measuring various elements and networks, IoT devices also help clients with intelligent meters. IoT devices with high processing capacities can improve the performance of the intelligent grid in processing, disaster recovery, dependability, and alerts.

9.7 Intelligent Robotics

The Internet of Robotic Things (IoRT) has several uses and can use the facts gathered from intelligent sensors to monitor and optimize machine performances in real-time. Logistics distribution, rescue, agriculture, security, health and military, and entertainment are all areas where service and humanoid robots are used. But as the current pandemic situation has demonstrated, IoRT technologies still need to progress significantly.

9.8 Near Field Communication (NFC) Payment

Nowadays, every retail transaction involves NFC since it enables customers to make contactless payments using their NFC-enabled smart devices. It shortens the payment process and improves the payment's security and indemnification**.**

9.9 Internet of Underwater Things with AI

IoUT with intelligent sensors, autonomous undersea vehicles useful for spotting enemy submarines and underwater loot. IoUT also aids in the metal, mineral, coral, and reef detection processes. In general, sensors with video capturing devices that work with IoT schemes are needed for locating undersea resources**.**

9.10 UAV Intelligent

The high range agility and autonomy of UAVs allow them to provide a variety of comforts to IoTs. For crowded surveillance with face recognition and mobile edge computing with insufficient energy capacity and dimensions, UAV-based IoT strategies effectively apply.

Additionally, UAVs with theory-based gaming platforms are often used in auto spacing to locateterrestrial stations. Remote operators can navigate around explicit scenes of interest using UAVs equipped with AR, VR, and MR technology. UAVs are also helpful in reducing transmissionpower and clustering IoT devices optimally.

9.11 Forensic applications powered by IoT

With the resources at hand, various models for IoT security and privacy have been implemented; nonetheless, the topic is currently under active investigation. Recently, there has

been some attention paid to IoT digital forensics. IoT security vulnerabilities are very likely because IoT is still in its early stages of development. To trace assaults and locate trustworthy digital evidence to identify offenders, active digital forensics methods must be built in equal measure with security justifications. The VitalPatch can be examined to reveal any linked forensics items of a person, such as heart rate, ECG trends, activity tracking, port scans, timeline records, etc.

9.12 Iot-Based Intelligent Healthcare Systems

Wearable Internet of Things (IoT) devices enable continuous physiological monitoring, which is more helpful for ongoing health than fitness tracking. A tool to track and elevate our mood during the day is called Moodables. Moddables are head-mounted devices that elevate our mood by delivering low-intensity electricity to the brain. Small, ingestible sensors keep track of the medication inside our bodies and alert us if they detect any irregularities, providing diagnostic patients with early warnings. Additionally, it can be used to shorten wait times for emergency rooms, improve drug management, track patients, and guarantee staff availability for key hardware.

9.13 Smart Disaster Management

The potential damage from forthcoming disasters is reduced with the use of intelligent disaster management. Additionally, ensure prompt and appropriate advice is given to the victims for quick recovery. The Internet of Things (IoT) skill has progressed in recent years and is likely to be highly useful in disaster situations. Risk reduction is aided by IoT systems that utilise satellite communication and geographic information systems. To avoid awareness-raising, relief and reaction efforts, and missing person searches, it gives recommendations for prevention, issues alerts in advance, and makes use of social media.

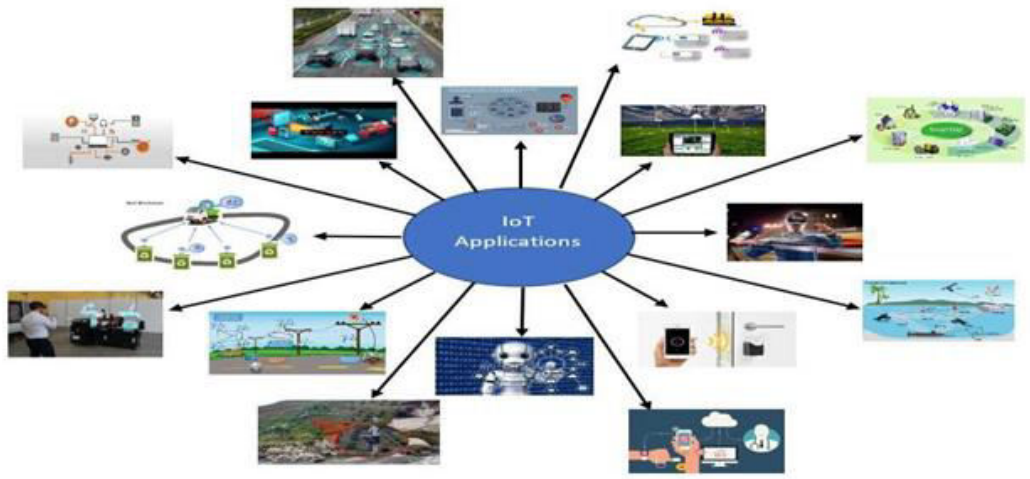
9.14 Music

The majority of IoT applications today are made for energy optimization, industrial manufacturing, environmental monitoring, and energy monitoring. Google's Universal Orchestra and MIT's Patchwork are two significant instances of IoT-based Music developments. Some of the remote performances created by well-known international corporations include SoundWIRE and JackTrip. Rhythmic vibration helps the user feel the rhythm of the music by having actuators collect data before beginning to vibrate with a rhythm and intensity that correspond to the music being played. Instrumentalists can perform auto-tuned instruments over the partnered device. The implanted gadget has sensors that can detect marks on a device's screen. Instrumentalists can then play tools remotely over the gadget by auto-initiatinga similar movement on crucial musical instruments.

9.15 Agriculture

The demand for food sources has greatly increased as the world's population has grown. The employment of cutting-edge techniques by agriculturalists to increase food production is supported by developed countries and research institutes. One of the IoT's most rapidly expanding industries is smart farming. Here, farmers are utilizing artistic interpretations of the data to provide a more favorable return on investment. Smart irrigation makes it simple to manage water usage and control traditional peats by using Internet of Things (IoT)-based sensors to measure the amount of moisture in the soil and releasing water through irrigation

pipes. The internal meteorological data of a greenhouse could be tracked and managed to harvest the most delicate conditions for growing plants. They examine and then enhancevarious control strategies using the sensible data that was collected from several sensors and kept in a centralized server. Different Applications have described in Figure [7]



9. Open Research Challenges for AI-based IoT Systems

Obstacle 1: How is AI-based Industry 5.0 handling processing power?

Computing power issues are affecting both technical and industrial companies. IoT devices generate enormous amounts of data that must be consistently consumed in order to construct AI models that can analyse this big data with ML and DL. Manufacturing businesses and new businesses are in serious danger. The amount of power required for a learning algorithm typically drives the developer away. Although they require more cores than GPUs, ML and DL are exceptional AI components with excellent precision and efficiency. The researcher has used many different concepts and plans to advance ML and DL models in a number of gadgets.

Obstacle 2: How AI addresses IoT's technological issues ?

many research-focused models for IoT are now planned. IoT, which consists of several links to contact/communicate with various networks, still has many practical jobs to do like a heterogeneous network, nonetheless. There are effective, well-known platforms for hiding network arrays at the moment, but they are complicated and use a lot of power. IoT confronts significant difficulties with service-oriented architecture due to cost and performance constraints. The number of connected devices rises, scalability, data pre-processing, service provisioning, and networking become complicated depending on the applications. Particularly, communication issues and repeated delays are also brought on by the transmission of sensed data over heterogeneous networks. High automated standards are required so that data received from various devices may be quickly accessed and transformed effectively inside IoT systems. The correct handling of linked device collaboration between many entities is more crucial; identification and optimization are still unresolved research problems. A unified data structure is one of the appropriate integrated processes that an IoT needs because it is marginally impacted by newly implanted items. AI-based methods make use of useful IoT data properties, but they were burdened by complexity and power consumption

Obstacle 3 How Industry 5.0 will handle various Implementation approaches?

Based on the recent advancements in numerous network technologies, AI has the power to digitally alter any industry or business field. The absence of innovative implementation ideas, however, is a major problem with AI. In order to prosper in a corporate setting, a strategic strategy is required, such as identifying progressive areas, objects with benefits, persistent failures, etc. Company managers, supervisors, and technical teams need to be well-versed inthe capabilities, limitations, and challenges currently facing AI in order to understand the aforementioned issues. Companies can easily identify the areas that AI can improve if they keep up with these developments on information working styles connected to AI.

Obstacle 4 What were the implementation issues with IoT data analytics, and how were they resolved?

We covered IoT analytics using ML and DL methods in the section above. In some circumstances, addressing time series-based data structures presents challenges for IoT analytics deployment. Finding justifications for extrapolative analysis is challenging since IoT- based intelligent sensor nodes continuously collect large static data samples over an extended period of time. Furthermore, it is challenging to locate adequate storage space and then analyze the collected data quickly. Since data is sensitive, it can only be used in precise and appropriate ways to provide useful information about any company's product. Therefore, data scientists are needed by data scientists, developers, data processing specialists, and database specialists with in-demand skill sets.

Obstacle 5 How AI addresses the security and system complexity issues posed by wireless networks and IoT ?

Complexity is a key component of any network, and the integration of AI into communication systems has increased the complexity of schemes. In industrial network case-specific ML, DL methods are used to accomplish a single goal while ignoring the objectivities that are still there. Additionally, IoT devices with sufficient resources communicate data to higher levels without doing any preprocessing calculations. Therefore, the embedded AI methods in WSNs and IoT must maximize single goals while supervising other constraints like storage, connectivity, processing, and latency. The AI strategy used in one layer may aid or benefit from optimization in another layer. The major goal of using AI for IoT is to monitor network flow and identify security flaws like incursion before making decisions. AI aids in the creation of high-quality data sets that include details on assault categories and outlines. Therefore, IoT security using AI methods based on top-notch datasets is highly improbable. Being aware that getting a dataset for IoT safety training is more difficult than in other fields is important. In the age of AI-capable technologies, privacy deals with the moral, ethical, and legal issues that are anticipated to come to light. There are several methods for information to leak; occasionally, AI will do so while carrying out multiple jobs at once. Encoding, decoding, blockchain, quantum models, and shuffling algorithms will undoubtedly handle privacy issues in WSNs and IOT.

Obstacle 6 How can AI enable UIoT to have high-speed intelligent communication?

Dynamic AUVs, underwater magnetic inductions, and acoustic networks are all components of the UIoT, a complex heterogeneous network. An unanticipated problem for the UIoT system is the ocean's unstable environment, which affects the topology and localization accuracy underwater due to tides, wind, and temperature changes. Every network component of the UIoT

is optimised and coordinated for communication by using AI-based methods, which subsequently makes the greatest use of the deployed network.

Obstacle 7 : How will product management change as a result of Industry 5.0's AI?

Many industrial sectors use Radio Frequency Identification (RFID) tags for supply-chain, product tracking, and delivery management. RFID readers are placed along the entire route to monitor and RFID tags are implanted on delivery packages. IoT devices enable greater reader position flexibility while enabling ongoing interoperability amongst RFID-based devices employed by various performers. Retail businesses use these IoT devices largely to indicate product availability and correct stock data. More crucially, a combination of sensors, biosensors, and RFID technology may allow the food industry to control manufacturing processes and product value.

Obstacle 8: Human Resources

No matter how large a firm is, how well it uses AI in the areas where development is required matters. Since AI is an emerging technology, there is a demand for people that can manage and use the AI models. In order to hire and train an AI expert, businesses allocate more funds.

Obstacle 9 : Educating AI

If the deployed AI system gathers enough data, it is prepared to go on to the training stage. Every AI-based strategy differs in terms of model type, data structure, results, and decision-making. In AI, there is no perfect model that is appropriate in every situation, thus we must try every model; depending on the results, we decide which works best among all of them. Therefore, if we want to encourage machines to learn more, we must first examine more of the natural world by giving them new features in the form of intelligent sensors. Additionally, run suggested models frequently with each new option so that AI models would automatically learn to perform better.

Obstacle 10 Connected Devices

Almost every industry in the world underwent a digital transformation using connected devices. It is still difficult to connect basic IoT devices, like intelligent sensors, to the internet since they use Bluetooth or Zigbee to communicate with other sources. Since connecting every gadget to the internet requires reinstallation, replacement with new equipment, and sophisticated hardware machines, it is not a simple operation. Not a straightforward process, AI-based MQTT is a simple illustration of how to link gadgets with the internet.

Obstacle 11: Effective Sensing

Numerous factor combinations make up the urban environment; heterogeneous multiple sensing models are required for effective monitoring of those factors. They require a broad framework for the collection of sensed data and the ability to present it in terms of its spatial

and temporal features using both stationary and mobile sensing equipment as well as random, constant sampling. For example, to recognise urban noise and air pollution zones, fixed intelligent sensor nodes must continuously sample noise and air quality data. Compressive wireless sensing (CWS) uses synchronised signals to reduce each signal's broadcast strength and transmit loudly. samples to a central node for data aggregation from quality of air measurement predicted data

10 Final Thought (Conclusion)

This chapter described a path that began with comprehending the IoT skills with AI-enabled visual patterns and ended with how it can be useful in many areas. Additionally, it aids researchers and industry professionals that identify design, construction, and AI algorithms using IoT and cutting-edge IoT defenses. It provides a thorough analysis of a functional framework, followed by a knowledge hierarchy for IoT, object recognition, intelligent sensors, learning, and analytics in Intelligent IoT-enabled Systems. This chapter examines IoT paradigms with AI capabilities that will be used to advance civilization in the future. In particular, it will be seen that the extensive references of different works followed by implementations are crucial collections for engineers and administrations. Figure 7 Internet of Things Application Domain.

REFERENCES

1. Hwangnam Kim, Suk Kyu Lee, and Mungyu Bae (2017). IoT Networks in the Future: A Survey, Appl. Sci. 2017, 7, 1072, doi:10.3390/app7101072.
2. Juan Manuel Cueva Lovelle, B. Cristina Pelayo-Gbustelo, Vicente Garcia-Diaz, Edward Rolando Nunez-Valdez, Cristian Gonzalez Garcia, and others (2018. International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 5, No. 4, "A Review of Artificial Intelligence in the Internet of Things."
3. Mohammed Aledhari, Mehdi Mohammadi, Ala Al-Fuqaha, Mohsen Guizani, and Moussa Ayyash (2015). Ieee Communication Surveys & Tutorials, Vol. 17, No. 4, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Application
4. In 2017, Jie Lin, Wei Yu, Nan Zhang, Xinyu Yang, Hanlin Zhang, and Wei Zhao published their research. Ieee Internet of Things Journal, Vol. 4, No. 5, "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications."
5. Jörg Henkel, Lars Bauer, and Farzad Samie (2019). An Overview of Machine Learning in the Internet of Things, from Cloud to Things June issue of Ieee Internet of Things Journal, Vol.
6. Altexsoft, 2020, IoT Architecture: The Road from Physical Signals to Business Decisions, IoT architectural layers and components can be found at https://www.altexsoft.com/blog
7. G. B. Chander and G. Kumaravelan (2018). WirelessSensor Networks: An Introduction. Wireless Sensor Networks and Soft Computing, 1.
8. Chander, B., and Kumaravelan, G. 8. (2020). foundation of the internet of things. Internet of Things (IoT) Ecosystem: Fundamentals and Insight Paradigm (pp. 3-33). Cham: Springer.
9. K. Gopalakrishnan (2020). Traditional wireless sensor networks in the Internet of Things include security flaws and problems. Internet of Things (IoT) Ecosystem: Fundamentals and Insight Paradigm (pp. 519-549). Cham: Springer.
10. Yu Chuan Lin 10. (2017). IEEE Robotics and Automation Letters, vol. 2, no. 1, pp. 1809–1816, Development of Advanced Manufacturing Cloud of Things (AMCoT) – A Intelligence Manufacturing Platform Chen Chao Chun, 11. (2018). IEEE Robotics &

Automation Letters, vol. 3, no. 3, pp. 1378–1385, "A Novel Automated Construction Scheme for Efficiently Developing Cloud Manufacturing Services." URL: 10.1109/LRA.2018.2799420

11. Rajkumar Buyya, Soumya K. Ghosh, Harshit Gupta, and Amir Vahid Dastjerdi (2015). Journal Software: Practice and Experience, Volume 47, Issue 9, Pages: 1275–1296. iFogSim: A toolbox for modelling and simulation of resource management approaches in the Internet of Things, Edge and Fog computing context
12. B Zhou and R Buyya (2018). A taxonomy, survey, and future prospects for augmentation technologies for mobile cloud computing, ACM Computing Surveys (CSUR). Pages 1–38 of Volume 51, Issue 1.
13. R. Ranjan, O. Rana, S. Nepal, Z Wen, P. James, M. Yousif (2018). The Internet of Things and data science integration, as well as IEEE Cloud Computing, are the upcoming major issues.
14. Rajkumar Buyya, Soumya K. Ghosh, Harshit Gupta, and Amir Vahid Dastjerdi. (2015). iFogSim is a modelling and simulation toolbox for resource managers. 16. A.V. Ungureanu (2020). Industry 4.0 to Industry 5.0 Transition. The four Cs of the current economic climate. 16th Economic International Conference NCOE 4.0 2020, in C. Nastase (vol. ed.), Lumen Proceedings: Vol. 13, pp. 70–81.
15. Sherali Zeadally, Zubair Baig, Erwin Adi, Adnan Anwar (2020). Neural Computing and Applications (2020) 32, pp. 16205–16233; Machine learning and data analytics for the Internet of Things.
16. Deepti Kiran, Itisha Sharma, and Illa Garg (2020). European Journal of Molecular & Clinical Medicine, Volume 07, Issue 08, pp. 2515-8260, "Industry 5.0 And Smart Cities: A Futuristic Approach."
17. Saeid Nahavandi (20). (2020). Sustainability 2019, 11, 4371; 10.3390/su11164371; Industry5.0: A Human-Centric Solution.
18. Tie Qiu, Zhao Zhao, Tong Zhang, Chen Chen, and C. L. Philip Chen are in the list of 21.
19. IEEE Transactions on Industrial Informatics, 1551-3203 (c) IEEE. Underwater Internet of Things in Smart Ocean: System Architecture and Open Issues.(2019)
20. Vural Ozdemir and Nezih Hekim are number 22. (2018). The Internet of Things, Next-Generation Technology Policy, and the Birth of Industry 5.0: Making Sense of Big Data, OMICS Volume 22, Number 1, 2018 of A Journal of Integrative Biology by Mary Ann Liebert, Inc. DOI: 10.1089/omi.2017.0194
21. Chander, B., and Kumaravelan, G. 23. (2021). AI and cyber security: Part 1. An End-to-End Panorama of In The "Essence" of Network Security (pp. 147-171). Singapore's Springer.
22. Skobelev PO and Borovik SY are 24. (2017). From digital production to a digital society, weare transitioning from Industry 4.0 to Industry 5.0. Int Sci J. 2017;2(6):307e311.

23. T. Pflanzner, A. Kertesz, and 25 (2018). EAI Endorsed Transactions on Internet of Things, Volume 3, Issue 12, "A Taxonomy and Survey of IoT Cloud Applications"

24. T. H. Davenport & J. Kirby (2016). Winners and losers in the age of intelligent machines: only humans need apply. Harper Business, New York.

25. T. H. Davenport, V. Mahidhar (2018). What cognitive approach do you use? Harvard Business Review (Summer). Retrieved from "whats-your-cognitive-strategy" at https://sIoanreview.mit.edu

CNN BASED IMAGE QUALITY IMPROVEMENT IN HANDHELD ULTRASOUND DEVICES

S. Ashlin Lifty and R. Deepa
Department of Computer Science and Engineering, Prince Dr K Vasudevan College of Engineering and Technology, Chennai 600100

ABSTRACT

As ultrasound methods seem to be more precise, they are now increasingly popular in the healthcare industry; yet, the quality of the image in portable ultrasound tools is relatively poor. Convolutional Neural Networks are used in the proposed method to increase the visual norm in mobile handsets to higher views. Convolutional Neural Networks were proposed to preprocess in portable gadgets, leading to high sights. The median filter is being used in histogram equalisation to decrease unwanted noise and keep the features whilst ensuring a high spectral response.The histogram equalisation method is also used to change the dynamical value's histogram. To boost picture sharpness, it stretches out the most prevalent pixels or expands out all the image's intensity value. Unsharp masking, despite to its name, is a technique for sharpening an object. When post-processing any digital photographs, sharpening is vital as it tends to emphasise information. A Convolutional Neural Network is frequently used to achieve better precision.CNN was designed specifically to deal with pixel data. It's a supervised classifier that creates a structure, analogous to a pipeline, but then just generates a fully-connected surface in which all the layers are joined and also the result is analyzed. It could provide highly designed and programmed training as well as provide a greater restoration vision of minute details, shape, and dispersion by using CNN.

Index Terms-- Hand-held ultrasound device, Convolutional Neural Network ,Image enhancement Median filter , Histogram Equalisation ,Unsharp masking.

I. INTRODUCTION

Image processing is a method of improving or obtaining data from a shot by performing actions on it. Classical and digital image are indeed the two main forms of image analyses. Traditional data preprocessing mainly deals with 2D signals and uses analog inputs. The approach of employing a software program to modify bitmap images that used a technique is referred to as digital image analysis.Ultrasound is a kind of painless analytic imaging. It makes constant photographs or video of inward organs or different tissues, for example, blood supply routes, utilizing high-recurrence sound waves.The transducer discharges sound waves into your body, assembles them, and moves them to a PC, which produces illustrations. PC related gadgets, hand-conveyed (HCU), and handheld (HHUSD) frameworks are the three kinds of convenient ultrasound gadgets. In this model, the Convolutional Neural Network method is utilized.; it connects with the utilization of man-made reasoning calculations to observe designs in informational collections with information focuses that are neither classed,labelled.In the field of picture processing,the most critical obligation is picture recuperation. The picture is regularly debased, causing noise to show up in the picture. The middle channel is regularly used to lessen the presence of such commotion, but it functions admirably for pictures with generally 20% clamor force. Along these lines, to accomplish a superior picture reclamation, we can utilize

one more picture rebuilding approach known as versatile middle separating, which is compelling for commotion forces more noteworthy than 20%. The benefit of a versatile channel over a middle channel is that it doesn't obliterate the picture's edges or fine elements. The versatile channel works in two stages: first, it tracks down the piece's middle worth, and afterward it inspects in the event that the ongoing pixel esteem is a motivation or not. In the event that a pixel's worth is adulterated, it replaces it with the middle worth, or it keeps the dark scale pixel's worth. The middle channel is the most generally utilized. It is a nonlinear separating approach. The center can be n × n pixels in size and is planned to add or float over a mxm contorted picture. The middle worth of the nxn part on the picture is procured during this strategy, and afterward the worth of a given pixel is subbed with it. [1].Magnetic Resonance Imaging (MRI) is a clinical imaging strategy that is utilized to break down and analyze messes like malignant growth and cerebrum cancers. Doctors need solid difference checked pictures procured from MRI to inspect these illnesses for better treatment purposes since it contains the most data about the infirmity. Since MRI pictures are low-contrast, diagnosing them may be troublesome, thusly better picture pixel localisation is required.Adjustment approaches for histograms help in the improvement of pictures, bringing about expanded visual quality and a clear cut issue. The difference and splendor have been helped so that the first data has not been lost and the brilliance has been kept. We looked at the few leveling methodologies that are basically examined in this research[2][9].Infrared and noticeable picture combination has been utilized in an assortment of utilizations, including military, observation, imaging.Wavelet change is utilized to show the need of unsharp veiling for visual combination. The disintegration of information photographs is finished with DWT (infrared, apparent). The coefficients are approximated and indicated. Unsharp concealing was utilized to improve Splendor utilizing assessment factors. from that point forward,the ,sifting utilizing calculation is used tocombine the assessment boundaries that were acquired after unsharp covering. To consolidate nitty gritty boundaries, the ideal consolidation rule is utilised.Finally, IDWT is utilized in the making of a combined picture. When contrasted with the outcomes acquired utilizing the proposed combination approach, the proposed combination strategy gives great difference as well as better execution brings about terms of mean, entropy, and standard deviation.Unsharp covering is a picture upgrading procedure that supports the differentiation of edges to work on the details.Whenever the foggy picture is deducted from its genuine picture, the outcome could be a crisper picture. Permit f(l,k) to be a symbolism, fs(l,k) to hone variety of f(l,k), and f(l,k) to be a cloudy type of f(l,k) to be a dim release of f(l,k) (l,k). Utilizing the grouping calculation, the picture pixels strategy integrates different source information . This approach doesn't utilize any changes. The combination rule is straightforwardly applied to the pixels of the images[3][6][10].To survey the presentation of various kinds of CNNs, a scope of picture informational indexes are given. An ImageNet dataset, as well as CIFAR10, CIFAR100, and MNIST picture informational collections, are standard benchmark datasets for estimating the presentation of a convolutional brain organization. The outcomes got from notable organizations, Alex Net, GoogLeNet, and ResNet50, is inspected in this review. Since evaluating a channel's proficiency on a specific dataset doesn't uncover its entire potential and cutoff points, we picked three of the most conspicuous pieces of information for our investigations: ImageNet, CIFAR10, and CIFAR100. Recordings are utilized as assessment tests instead of training tests, it ought to be noted. [4][5][7][8]. CNNs (Convolutional Neural

Networks) were grown basically for pixel information processing.The reason for picture improvement is to make a picture more helpful for a specific movement, for example, making it all the more emotionally appealing for human sight. Upgrades are used to work on visual understanding and appreciation of visuals more straightforward. Advanced photography enjoys the benefit of permitting us to alter the computerized pixel values in a picture. Contrast upgrade is a basic part of any emotional picture quality appraisal that is utilized to work on the general nature of a clinical picture for highlight perceivability and clinical estimation.

II. BACKGROUND

A. Plane Wave Imaging

Plane wave imaging, a typical ultrafast clinical ultrasound imaging methodology, accomplishes a high casing rate by producing a solitary plane wave without focusing.However, when contrasted with the consistently utilized centered line examine mode, the imaging quality endures incredibly. Versatile pillar arrangement, as it is known, can further develop imaging quality to the detriment of extra handling. Profound brain networks are utilized to achieve PWI improvement while holding a high casing rate . The PWI signal from a high accuracy objective is the network input, in the mean time the result signal is the concentrated output answer from a similar position.

B. Frequency Domain Processing

The recurrence area RF information and reactions from the concentrated transmission are used training.The input is in the time series, which is moved to the fourier change, then, at that point, DNN is taught, and the recurrence and seriousness is gotten back to the time area.. These organizations make an interpretation of PWI RF information into suitable centered reactions, which can then be utilized to produce a high-goal B-mode ultrasound picture.

III. METHODS

A. Method Overview

CNN is many times used to fill in for advanced mobile phones' low quality pictures. A convolutional brain organization (CNN) is utilized to perform picture handling that is well defined for pixel esteem. A Convolutional layer, a Pooling layer, and a Fully connected layer make up the three layer network models.It's a managed system in light of machine learning.The middle channel is in many cases utilized in computerized picture handling to diminish bending from harmed information. Middle separating, frequently alluded as salt-and-pepper sound decrease, is a backward system to limit impedance. The impact change of the cell being assessed in a middle channel, that movements along the edge, is equivalent to the typical pixel power of the pictures inside the casing. To achieve the luminance and brilliance goal, histogram balance is utilized. Histogram Equalization is a strategy that changes the power of an image by utilizing the dispersion. To support the picture's splendor, it circulates out the most normal pixel esteem or grows out the picture's dynamic reach. Honing of pictures Increasing the worth outcomes in a more keen picture while utilizing unsharp masking.A moderate (boisterous) series of pictures of a situation can be converged to shape a somewhat raised picture or picture series, as per the thought hidden super-goal. As an outcome, it endeavors to remake unquestionably the most elevated scene picture from an assortment of moderately low noticed photographs.

B. Denoising

Denoising is a term used to portray picture separating strategies in which both the info and result pictures are force pictures at the most minimal degree of reflection. Eliminate clamor is a strategy to increment picture information by eliminating undesirable disfigurements or expanding a part of the picture that is significant for post handling. Picture pre - handling the techniques exploit the significant redundancy in pictures. One of the pre-handling approaches is picture cleaning. Channels are utilized to change or intensify visual attributes as well as get experiences from photos, like limits, edges, and masses. Deeply, which is a little framework treated to each adjoining pixels inside an image. Middle channels are helpful for decreasing encompassing clamor when the unsettling influence extent likelihood thickness work has enormous furthest points and redundant trends.The middle sifting process is done by floating an edge over the picture. The handled picture is shaped by picking the midpoint of the info window's information and putting it in the focal point of the source pictures. The midpoint is the greatest probability evaluation of position in an Interpolation sound sign. The middle channel functions admirably in the drawn out looked at commotion since it gauges the variety for connected with the right areas. At the point when an edge is crossed, one side of the window dominates, and the result unexpectedly flips between the qualities. Thus, the edge doesn't distorted.Using the pixel replication approach, each line and segment of the picture is duplicated, and a vacant middle veil is flowed across each line and column.It is meant as, g (x,y)=h(x,y)+n(x,y)

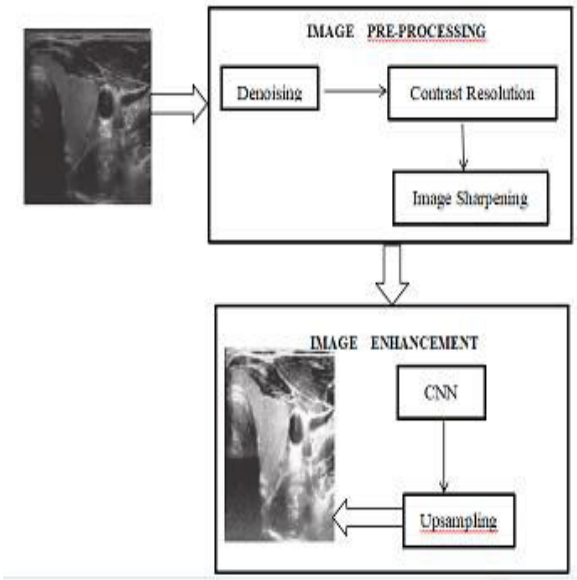


Fig 1: System Architecture

A kind of advanced picture inspecting that includes expanding the quantity of pixels in a picture without adding any information or detail.Typically, the source pictures are being utilized to assess the obscure the new shaded cells. When photographs are amplified along these lines, the visual quality is regularly compromised.

C. Contrast Resolution

The capacity of any imaging modalities to segregate between contrasts in picture force is alluded to as difference goal in radiology. The inherent differentiation nature of a computerized

picture is characterized by how much particular adjoining pixels,which would be determined as how much pieces per data of interest. Unreasonable splendor debases picture quality.The graphical portrayal of a picture's pixel power values is known as a histogram. It's the information structure that monitors the frequencies of all the picture's pixel power levels.This is achieved utilizing factual coordinating, which permits the picture's diminished regions to arrive at more noteworthy force. Histogram adjustment is utilized to compute the scale factor while keeping up with picture quality. The techniques portrayed can be utilized to change the luminance:

G(x, y) =h'(x, y) +b

G(x, y) =h'(x, y)-b

In a histogram adjusting, the x hub shows the dim scale pixel power, while the y hub mirrors the impact of different levels. The thickness of a cell is its luminance. As opposed to the others, the outline delineates number of times picture pixels are at a specific degree of power. Pictures are put away in a PC as a variety of numbers, which are suggested as contiguous pixels. These board values demonstrate the luminance of each board. Highly contrasting are addressed by the numbers 0 and 255, likewise.

D. Image Sharpening

It's a strategy for improving a picture's clear sharpness. Photoshop can't mystically fix additional subtleties once a picture has been

Caught; the genuine goal stays fixed. As such, expanding acutance is the main technique to work on evident sharpness. You ought to increment edge difference to your picture in the event that you believe it should look more keen. Fine detail is accomplished by underlining the picture's corners and specialized viewpoints. To recognize any edges, it takes away an obscured (unsharp) duplicate from the first picture. This edge detail is utilized to make a veil. The impact is then applied to the first picture, with the differentiation expanded at the edges.Sharpening eliminates material from the edge to make a new, sharp edge, while sharpening keeps up with the sharpness of the sharp edge by taking the blade's edge back to the centre.Unsharp concealing (USM) is an advanced picture upgrading strategy originally utilized in bad imaging and presently regularly utilized in picture handling method. The method's name comes from the way that it makes a veil of the first picture utilizing an obscured, or "unsharp," negative picture. From that point onward, the unsharp veil is applied to the information solid picture, creating in a very hazy picture. Despite the fact that the picture seems more clear, this may not be a genuine portrayal of the thing.

G (unsharp) =h'(x, y) +Gmask(x, y)

Gmask(x, y) =h'(x, y)-g(x, y)

Unsharp Mask makes a mutilated portrayal of the picture that is then deducted from the genuine picture utilizing an Image channel.

E. Image Enhancement

Nature of a picture is the strategy of fortifying the capacity of better review to understand or significant considering in pictures while at the same time giving 'better' information than those

other mechanized picture examination methodology. The central objective of picture improvement is to alter the characteristics of a picture to make it more OK for a predefined reason and onlooker. As a result of this method, one or many picture ascribes are adjusted. The qualities picked and how they are changed are task-explicit. CNN was made in view of pixel information. The CNN involves a progressive model that forms an organization looking like a pipe and afterward yields a completely associated layer.Typically, a CNN has three layers: Convolutional layer , Pooling layer, Fully associated layer.

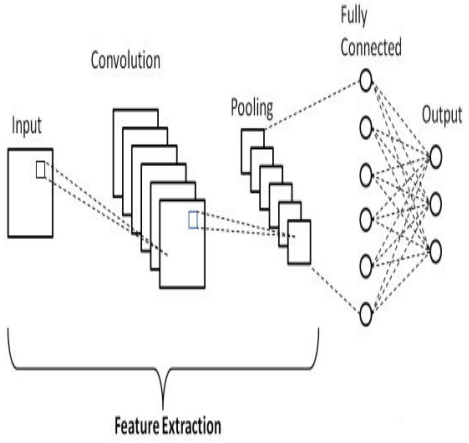


Fig: 2: CNN layer Classification

1) Convolutional layer

A network with the accompanying parts is taken care of into a CNN: (measure of data sources) x (source level) x (source width) x (source channels). The picture is disconnected to a component map, otherwise called an enactment map, in the wake of going through a convolutional layer, with the accompanying shape: (measure of information sources) x (include map level) x (highlight map width) x (include map channels).The input is convolved by convolutional layers, The result is in this manner gave to the following level. This is comparable to a neuron's reaction to a solitary upgrade in the visual cortex.Each convolutional neuron just cycles information for the open field it is alloted to. Albeit completely connected feedforward brain organizations can be utilized to recognize faces and classify information, this engineering is unsatisfactory for greater data sources like high-goal photographs. Because of the huge size of the contribution of pictures, where every pixel is a significant information trademark, it would require an exceptionally enormous number of stowed away neurons, even in a shallow design. For instance, every cell in the second layer of a convolution layer for a (little) example picture 100 by 100 has 10,000 loads. Convolution, then again, decreases how much pieces of information, taking into account a more profound network.Furthermore, since spatial connections between various highlights, convolutional brain networks are proper for information with a lattice like engineering.

2) Pooling layer

The component guides' aspects are decreased by utilizing pooling layers. Subsequently, the arrangement of factors to learn is decreased, similar to the quantity of calculation done in the organization. The elements contained in a space of the dataset created by a convolution layer are summarized by the pooling layer. Therefore, as opposed to exactly situated highlights made by

the convolution layer, following activities are led on summed up highlights. Thus, the model is more impervious to changes in the place of components in the source pictures. Pooling that chooses the most Note worthy part from the region of the dataset covered by the channel is known as max pooling. Subsequently, following max-pooling layer, the result would be an element map involving the main highlights from the first layer.

3) FULLY CONNECTED LAYER

In a Brain Organization, Completely Connected

layers are those where all of the result from one level are interconnected to each enactment square of the following layer. The last couple of layers in most run of the mill AI models are full associated layers that accumulate the information recovered by going before layers to create the last result. The result of the completely associated layers shows the likely highlights of the information. Albeit the secret layers could be leveled and associated with the result neuron, including a thoroughly layer permits you to learn non-direct combinations of these attributes for a (generally) minimal expense. The last layers of a Convolutional NN are totally associated layers. A layer is associated layer's neurons are totally connected with the excitement of the resulting layer's neurons. Then again, these totally associated layers can acknowledge one bunch of information. To switch our 3D information over completely to 1D, we utilize Python's compliment work . Our muti volume is successfully decreased to a single vector.

IV. RESULT AND DISCUSSION

The use of a convolutional mind association to overhaul the idea of the image of hand-held ultrasound devices with a faster response time has been proposed. By decreasing upheaval and altering contrast in the image, the proposed system hopes to outfit more exact readiness with high precision and make a significant standard modifying picture with postprocessing plan and speck. Moreover, the image is improved by adding more pixels and supporting the image's objective. The denoising framework can be done using an adaptable center channel over a center filter.Traditional progressed channels can't accomplish some sign taking care of endeavors, however flexible channels can. Convolutional cerebrum associations (R-CNN) areas can be proposed for zeroing in blundering on the tortured locale and dealing with.

V. CONCLUSION

The utilization of a convolutional brain organization to upgrade the nature of the picture of hand-held ultrasound gadgets with a quicker reaction time has been proposed. By lessening commotion and modifying contrast in the picture, the proposed strategy looks to furnish more precise preparation with high exactness and make a high goal rebuilding picture with postprocessing design and dot. Furthermore, the picture is improved by adding more pixels and supporting the picture's goal. The denoising system can be finished utilizing a versatile middle channel over a middle filter.Traditional advanced channels can't achieve some sign handling undertakings, though versatile channels can. Convolutional brain organizations (R-CNN) districts can be proposed for focusing erring on the tormented region and handling.

ACKNOWLEDGMENT

I would like to express great fullness to P.G Department of Computer Science Engineering, Prince. Dr.K.Vasudevan College of Engineering and Technology, ponmar.

REFERENCES

[1] Hetvi Soni, Darshana Sankhe, "Image Restoration using Median Filtering," Volume: 06 Issue: 10 ,Oct 2019.

[2] Omprakash patel,Yogendra P.S Maravi,Sanjeev Sharma,"A Comparative Study of Histogram Equalization Based Image Enhancement Techniques for Brightness Preservation and Contrast Enhancement," Signal & Image Processing An International Journal,November 2020.

[3] Sumanth Kumar Panguluri, Laavanya Mohan,"Discrete Wavelet Transform Based Image Fusion Using Unsharp Masking,"December 2019.

[4] Di Fan, Shuai Fang, Guangcai Wang,Shang Gao, Xiaoxin Lin," The visual human face super-resolution reconstruction algorithm based on improved deep residual network," EURASIP Journal on Advances in Siganl Processing, November 2019.

[5] Tao Qiu, Chang Wen , Kai Xie , Fang-Qing Wen, Guan-Qun Sheng, Xin-Gong Tang ,"Efficient medical image enhancement based on CNN-FBB model," Volume 13, Issue 10, 22 August 2019.

[6] Kuo Shiuan Peng, Fang Cheng Lin, Yi-Pai Huang, Han Ping D. Shieh,"Efficient super resolution using edge directed unsharp masking sharpening method,"Proceedings - 2019 IEEE International Symposium on Multimedia, ISM 2019.

[7] Sushma L , Dr. K. P. Lakshmi," An Analysis of Convolution Neural Network for Image Classification using Different Models,"Volume 9, Issue 10, october 2020.

[8] Robin Zatta, Ritesh Jain, Janusz Grzyb, Ullrich R. Pfeiffer," Resolution Limits of Hyper-Hemispherical Silicon Lens-Integrated THz Cameras Employing Geometrical Multi-Frame Super-Resolution Imaging ,"IEEE Transactions on Terahertz Science and Technology 2021.

[9] Kaeser Md. Sabrin, Md. Haider Ali," An Intelligent Pixel Replication Technique by Binary Decomposition for Digital Image Zooming,"Image and Vision Computing ,December 2020.

[10] Wei Ye, Kai-Kuang Ma," Blurriness-Guided Unsharp Masking," IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 01 Sep 2019.

[11] P. M. Corry et al., “Human cancer treatment with ultrasound,” IEEE Trans. Ultrason., Ferroelect., Freq. Control, vol. 31, no. 5, pp. 444-456, Sep. 1984.

[12] B. G. Ewigman, “Effect of prenatal ultrasound screening on perinatal outcome,” N. Engl. J. Med., vol. 329, no. 12, pp. 821-827, Sep. 1993.

[13] C. A. Linte et al., “Evaluation of model-enhanced ultrasound-assisted interventional guidance in a cardiac phantom,” IEEE Trans. Biomed. Eng., vol. 57, no. 9, pp. 2209-2218, Sep. 2010.

[14] J. M. Chang et al., “Experimental system prototype of a portable, low-cost, C-scan ultrasound imaging device,” IEEE Trans. Biomed. Eng., vol. 55, no. 2, pp.519-530, Mar. 2008.

[15] B. J. Kimura et al., "Usefulness of a hand-held ultrasound device for bedside examination of left ventricular function," Am. J. Cardiol., vol. 90, no. 9, pp. 1038-1039, Nov. 2002.

[16] C. Martinoli et al.,"Point-of-care ultrasound in haemophilic arthropathy: will the HEAD-US system supplement or replace physical examination?" Haemophilia, vol. 22, no. 1, pp. 20-21, Dec. 2016.

[17] K. F. Stock et al., "Comparison of a pocket-size ultrasound device with a premium ultrasound machine: diagnostic value and time required in bedside ultrasound examination," Abdom. Imaging, vol. 40, no. 7, pp. 2861-2866, Oct. 2015.

[18] B. P. Nelson and A. Sanghvi, "Out of hospital point of care ultrasound: current use models and future directions," Eur. J. Trauma. Emerg. Surg., vol. 42, no. 2, pp. 139-150, Apr. 2016.

[19] N. Kolbe et al., "Point of care ultrasound (POCUS) telemedicine project in rural Nicaragua and its impact on patient management," J. Ultrasound, vol. 18, no. 2, pp. 1-7, Jun. 2015.

[20] D.M. Becker et al., "The use of portable ultrasound devices in low-and-middle-income countries: a systematic review of the literature," Trop. Med. Int. Health, vol. 21, no. 3, pp. 294-311, Dec. 2015.

[21] A. Riley et al., "Utility of hand-held echocardiography in outpatient pediatric cardiology management," Pediatr. Cardiol., vol. 35, no. 8, pp. 1379-1386, Dec. 2014.

[22] S. Ahn et al., "Smartphone-based portable ultrasound imaging system: prototype implementation and evaluation," in IEEE International Ultrasonics Symposium, Taipei, Taiwan, Oct. 21-24, 2015, pp. 1-4.

[23] R. Henderson and S. Murphy, "Portability enhancing hardware for a portable ultrasound system," U.S. Patent 9629606. Apr. 25, 2017

[24] G. R. Lockwood et al., "Real-time 3-D ultrasound imaging using sparse synthetic aperture beamforming," IEEE Trans. Ultrason., Ferroelect., Freq. Control, vol. 45, no. 4, pp. 980-988, Jul. 1998.

[25] G. Matrone, "The delay multiply and sum beamforming algorithm in ultrasound B-mode medical imaging," IEEE Trans. Med. Imaging, vol. 34, no. 4, pp. 940-949, Apr. 2015.

[26] S. Huang et al., "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," IEEE Trans. Image Process. vol. 22, no. 3, pp.1032-1041, Mar. 2013.

[27] S. Fu, "A compound anisotropic diffusion for ultrasonic image denoising and edge enhancement," in IEEE International Symposium on Circuits and Systems. Kobe, Japan, May. 23-26, 2005, pp. 2779-2782.

[28] C. Wang and Z. Ye, "Brightness preserving histogram equalization with maximum entropy: a variational perspective," IEEE Trans. Consum. Electron., vol. 51, no. 4, pp. 1326-1334, Nov. 2005.

[29] D. Wu et al., "Deep dynamic neural networks for multimodal gesture segmentation and recognition," IEEE Trans. Pattern. Anal., vol. 38, no. 8, pp.1583-1597, Aug. 2016.

[30] G. Wu et al., "Scalable high-performance image registration framework by unsupervised deep feature representations learning," IEEE Trans. Biomed. Eng., vol. 63, no. 7, pp. 1505-1516. Jul. 2016.

[31] J. Shi et al., "MR image super-resolution via wide residual networks with fixed skip connection," IEEE J. Biomed. Health, 2019, Preprint.

[32] J. Shi et al., "Super-resolution reconstruction of MR image with a novel residual learning network Algorithm," Phys. Med. Biol., 63(8): 085011. Apr. 2018.

[33] J. Hatvani et al., "Deep learning-based super-resolution applied to dental computed tomography," IEEE Trans. Radiat. Plas. Med. Sci., Dec 2019

SURVEY ON EFFICIENT AUTOMATED EPILEPTIC SEIZURE DETECTION USING A DEEP LEARNING TECHNIQUE

R. Manjupriya and A. Anny Leema
VIT University, Vellore 632014, VIT University, Vellore 632014

ABSTRACT

Uncontrolled neuron activity in the brain is the source of the neurological disorder known as epilepsy seizures that substantially reduce patients' quality of life and frequently risk their lives. Research on automatic seizure identification is essential for the field of seizure treatment and is necessary for seizure prevention. For these applications to be successful, the classification and detection algorithms must be accurate. Numerous methods have been studied, developed, and introduced over time. This review describes a variety of seizure detection techniques from the last six years, including both traditional techniques and more current deep learning algorithms. The benefits and drawbacks of each of the algorithms under investigation were compared in terms of performance. Our survey revealed that recent research has mainly examined the strength of deep learning methods in the identification and classification of seizures, which are used in various disciplines like image processing and categorization. Hybrid deep learning has also been studied. Finally, we conclude by discussing potential directions for future study in this area as well as unanswered issues.

Keywords: Epileptic seizure, Deep learning, seizure Detection methods, Neural Network, EEG.

1 INTRODUCTION

The sudden onset of signs and symptoms associated with epilepsy is brought on by improper brain synchronization and rapid neural processes. Epilepsy is a most common neurological chronic disorder that affects 50 million people around the world, it is characterized by the repetitive occurrence of abnormal electrical activity in the brain area over time and it is also called an epileptic seizure [1]. A seizure is a single occurrence of a sudden surge of electrical activity in the brain for a short period whereas epilepsy is the repetitive occurrence of a seizure. An epileptic seizure is also called a primary seizure due to abnormal electrical activity. on the other hand, there is a non-epileptic seizure also called a secondary seizure. Non-epileptic seizure is not caused by abnormal electrical activity in the brain. It can have some physical causes like heart conditions, diabetes, emotional pain, and brain injuries. Non-epileptic seizure has a period of seizure-like activity without central nervous system problems [2].

It is challenging, time-consuming, and prone to error for a neurologist to manually observe and identify epileptic seizures. The quantity of long-term EEG recordings that can be evaluated by neurologists must be reduced by the development of efficient computer-aided technology to help patients and neurologists identify and recognize epileptic seizures [3]. Several machine learning algorithms have been developed to detect epileptic seizures using statistical, temporal, frequency, time domain, and dynamical factors. Traditional machine learning algorithms use a trial-and-error approach to choosing features and classifiers. One needs a solid understanding of data mining and signal processing methods to create a precise model. For little amounts of data, these models work effectively. Machine learning approaches may not work well in the modern era due to the increased availability of data [4]. DL techniques, which are cutting-edge

techniques, have therefore been used. Unlike traditional machine learning approaches, DL models need a lot of data during the training process [5]. This is because these models include numerous feature spaces and encounter the overfitting issue in the absence of data.

The monitoring technique used to capture the electrical activity taking place at the surface of the brain is called an electroencephalogram (EEG). EEG recording tests are used by a neurologist to diagnose many neurological disorders of a person's brain and most importantly for an epilepsy diagnosis. EEG waveforms are generally classified according to their frequency(0-60Hz), amplitude into five bands namely delta (0.1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), gamma (30–60 Hz). Each EEG signal subband captures different brain activity [6]. Table I summarizes these frequency ranges.

Table 1: EEG frequency bands

Brand	Frequency (Hz)	State
Delta	0.1-3.5	Deep sleep
Theta	4-7	Drowsiness (also early stages of sleep)
Alpha (Low range)	8-9.25	Relax with eyes closed
Alpha (high range)	10-11.75	Thinking normally
Beta (Low range)	12-15	Relax but focus
Beta(mid-range)	16.5-18	Thinking and interacting with the outside world
Beta (high range)	20-28	alertness
Gamma	Above 30	More awareness, high mental activity

2 Dataset

Datasets are crucial in the development of an accurate and reliable automatic seizure detection model. Several EEG Datasets, namely CHB-MIT, Bern-Barcelona, Bonn, Kaggle, and Freiburg are available to construct epileptic seizure detection systems that are automated.

2.1 CHB-MIT Scalp EEG database

Database was collected at the Children Hospital Boston containing long-term EEG recording of 22 pediatric subjects which include 5 males (age 3-22) and 7 females (age 1.5-19). Each patient case has 23 EEG signals, with 24 or 26 signals occasionally present.

2.2 The Bern-Barcelona EEG database

Bern-Barcelona dataset is the intracranial EEG recording of five epileptic patients and it is provided by the University of Bern, Switzerland. These EEG recordings are part of the diagnosis before the surgery. The database contained two types of EEG signals: focal and additional focal EEG signals. Each data includes 3750 combinations of simultaneously captured signals that have a length of 20 seconds and a sampling rate of 512 Hz.

2.3 Bonn database

EEG database is an online available dataset produced by the University of Bonn. It has five subtypes, identified by the letters A, B, C, D, and E. Each dataset consists of 100 single-channel EEG recordings with a sampling frequency of 173.61Hz and each has a period of 23.6 seconds.

2.4 Temple University Hospital

The TUH EEG Corpus has more than 25,000 EEG records from more than 14,000 individuals, making it the largest openly accessible repository of clinical EEG data. The EEGs in the TUH EEG Corpus comprise 24 to 36 channels of signal data in addition to an annotation channel with labels identifying events that are important to doctors and technicians. Signals are captured with a resolution of 16 bits per sample at a rate of 250 Hz.

2.5 Freiburg

The EEG database consists of invasive EEG data from 21 people with focal epilepsy that is medically incurable. The data was collected at the Freiburg University Hospital's Epilepsy Centre during intrusive pre-surgical epilepsy surveillance. The EEG data were obtained using a 16-bit analog-to-digital converter, a 128-channel Neurofile NT digital video EEG system, with a sampling rate of 256 Hz .

3 Epileptic seizure Detection Methods

The automatic Epileptic seizure detection technique generally consists of two key stages. The first stage proceeds with preprocessing, feature extraction, and its selection from EEG signals. The second stage involves designing and training classification systems for the identification and prediction of an epilepsy seizure using these extracted features. Fig 5 shows the flowchart of the seizure detection system.

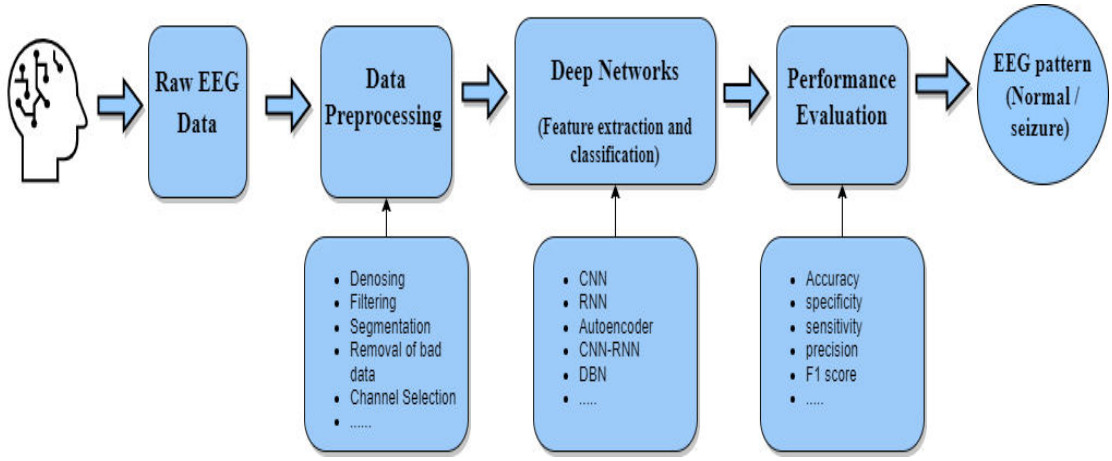


Fig 1. An epileptic seizure detection system

3.1 Pre-processing

Preprocessing includes noise removal, normalization, and input preparation while building automatic seizure detection models with deep learning. When removing undesirable noise levels, filters with finite impulse response (FIR) or infinite impulse response (IIR) are frequently used. Finally, signals are prepared for deep network deployment using various time domain, frequency, and time-frequency approaches.

3.2 Deep Learning Algorithm for Classification

Deep Learning can learn high-level features from raw EEG data using deep architecture. The architecture of a deep neural network has more hidden layers than two. This increase in network size results in a significant rise in the number of model parameters, necessitating the use of the proper learning techniques as well as safety measures to avoid overfitting the trained network.

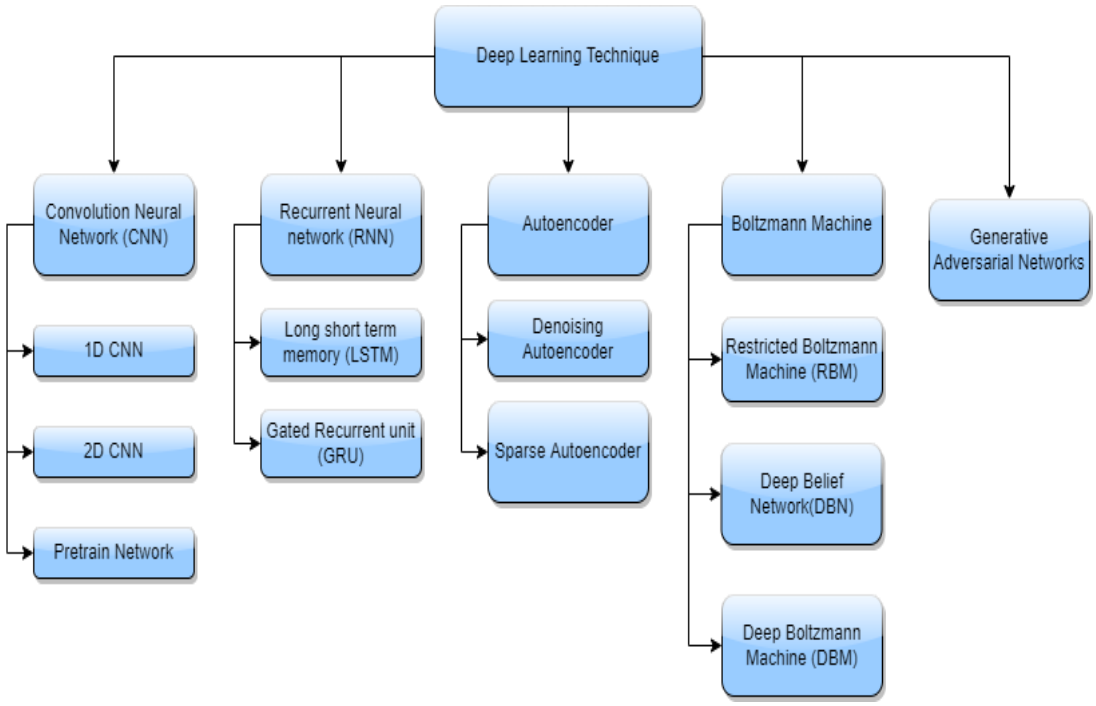


Fig 2. Various deep learning techniques for seizure detection

3.2.1 Convolution neural network

Convolutional networks are a type of DL network that is quite popular. Initially developed for use in image processing because it is capable of extracting features from images using a convolution kernel. Recently, they have been used in one- and two-dimensional frameworks for disease prediction and diagnosis using biological signals [7].

3.2.1.1 Two-Dimensional Convolutional Neural Network (2DCNN)

Deep 2D networks are now used in a variety of medical applications. The author in [8] proposed a novel 2D-CNN model which contains one dimension as time steps and another dimension as an EEG electrode for extracting spectral and temporal properties from EEG recordings and utilizing them to understand the general pattern of seizures. In [9] presented the 2DCNN model to classify non-seizure images and seizure images using phase locking value and Shannon entropy features. The author of [10] trained a deep convolutional network with linearly ordered score labels to classify Interictal epileptiform discharge (IED) and non-IED classes.

3.2.1.2 AlexNet

In [11] shows that ImageNet hosts the ILSVRC annually computer vision competition to address image classification challenges. Alex Krizhevsky's algorithm, AlexNet, changed the realm of image classification. It won the ImageNet challenge in 2012 and sparked off a whole Deep Learning era. The author of [12] employed the AlexNet network to detect a focal seizure. By utilising the feature extraction method and subsequently applying the Softmax activation function for classification, the suggested network achieved 100% accuracy.

3.2.1.3 VGG

The research team of [13] established the Visual Geometry Group (VGG) concept in 2014. They created a variety of models, one of which being VGG-16, that was entered into the 2014

ILSVRC competition. When it comes to image categorization challenges, the 16-layer VCG-16 performs well. The author of [14] detects epileptic seizures using sEEG signal image plots. Little (3*3) convolution filters were employed to effectively detect slight EEG signal changes after the signals were split into different time frames in the preprocessing stage. VGG-16 was then used for classification.

3.2.1.4 ResNet

In [16] the ImageNet challenge was won by Microsoft's ResNet, which used a 152-layer network with 96.4% accuracy. To this network, residual blocks suitable for training deep architecture were added utilizing skip connections that reproduced data from each layer to the following layer. In [17], advanced epilepsy detection employs TL relying on Inception-Resnet-v2 trained with the mean amplitude spectrum (MAS) characteristic to distinguish between onset, ictal, interictal and other preictal phases.

Table 2: Summary of relevant 2DCNN-based articles

Study	Network	Classifier	Accuracy (%)
[5]	2dcnn	Softmax and sigmoid	99.50
[36]	2dcnn	softmax	99.21
[37]	2dcnn	NA	99
[9]	2dcnn	softmax	90
[13]	Alexnet	Knn	98.78
	Vgg 16		98.56
[14]	2dcnn	Softmax	NA

3.2.1.7 One dimensional convolutional neural network(1DCNN)

In [18] proposed a 13-layer deep convolutional neural network (CNN) to distinguish between normal, preictal, and seizure classes with an accuracy of 88.7%. The researchers in [19] developed a deep learning system built from a collection of P-1D-CNN (pyramidal one-dimensional convolutional neural network) models. The P-1D-CNN model uses a majority vote technique to combine local outcomes of epilepsy detection. The author of [20] developed a novel automatic epileptic EEG detection method employing CNN to assess the presence of epileptic discharges in a 5-second EEG segment. MIDS and data augmentation techniques are used to improve model performance with an accuracy of 90% based on the standard CNN model.

3.2.2 Recurrent Neural Network

RNNs are a particular kind of artificial neural network that can be used to train sequential or temporal series data. The logical sequence of the input data can't be examined by any of the other models. Logic orders are closely populated with information and have a complicated timing relationship with one another.

3.2.2.1 LSTM

A simple RNN has a fundamental problem with short-term memory. RNN may overlook crucial information since it struggles to convey information from previous time steps to subsequent steps in long-sequence data. Another issue with RNN is the vanishing gradient problem. The challenge arises because of the gradients getting shorter as it back-propagates. To solve the short-term memory issue, LSTM gates were created, and gates in the network can regulate

the data flow. LSTM has a cell state and gates to learn, forget or recall information from each unit of the network. The following equations conclude the cells of LSTM behavior.

Forget gate $(f_t) = \sigma(W_f.[h_{t-1}, x_{t-1}] + b_f)$ (2)

Input gate $(i_t) = \sigma(W_i.[h_{t-1}, x_{t-1}] + b_i)$ (3)

Output gate $(o_t) = \sigma(W_o.[h_{t-1}, x_{t-1}] + b_o)$ (4)

New cell content $(\tilde{c}_t) = \tanh(W_c.[h_{t-1}, x_t] + b_c$ (5)

Cell state $(c_t) = f_t * c_{t-1} + i_t * \tilde{c}_t$ (6)

Hidden state $(h_t) = o_t * \tanh(c_t)$ (7)

In [21] proposed a seizure detection model which uses attention mechanism and bidirectional long short-term memory (BiLSTM). The author of [22] proposed deep long short-term memory (DLSTM) with minimum variance kernel random vector functional link net (MVKRVFLN) classifier to detect seizure and non-seizure effectively.

3.2.2.1 GRU

GRU is a variation of LSTM. The input and forgets gates of GRU are combined into a single update gate. In addition to combining the input and forget gates, it also makes some other adjustments. The number of gates is reduced to two. The reset gate is one, while the updating gate is another. Which data should be transmitted to the output is determined by these two gates. The author [23] used novel GRU based CADS were utilized to identify epileptic seizures. The plots are then fed into a GRU network with four layers and a Softmax classifier in the FC layer, yielding 98% accuracy. In [24] gives excellent results when a five-layer GRU network and Softmax classifier were combined.

Table 3: Summary of relevant RNN-based articles

Work	Network	classifier	class	Accuracy
[21]	C-LSTM	Softmax	5	98.80%
[30]	Bi-LSTM	Softmax	2	99.47%
[31]	Bi-LSTM	Softmax	2	99.60%
[10]	LSTM	Softmax	3	98%
[23]	LRCN	Softmax	4	89%
[24]	Bi-LSTM	Softmax	2	93%

3.2.3 Autoencoder

Autoencoder (AE) neural network is an unsupervised deep learning method that makes the output value equivalent to the input value. The input is encoded by the autoencoder into a latent space representation, which is subsequently decoded. In [25] present fusion methodology based on a deep convolution network and autoencoder-based model (AE-CDNN) to perform feature extraction of EEG signal in epilepsy. This model is employed on two public databases. Common classifiers like KNN, SVM, and MLP. DT is used to classify features obtained by the AE-CDNN model with high accuracy. The author of [26] exploits a one-dimensional convolutional variational autoencoder (1D CVAE) capable of automatically learning features to segregate between different seizure states with 99% accuracy.

Table 4: Summary of relevant Autoencoder based articles

Study	Input	Network	Classifier	Accuracy (%)
[32]	Raw EEG	Dcae	Multi-Layer Perceptron	93.21
[29]	Raw EEG	2d-Dcae	Multi-Layer Perceptron	98.05
[25]	Raw EEG	Ae-Cdnn	SVM, MLP, DT	92
[26]	Raw EEG	1d Cvae	Softmax	99
[38]	Frequency Spectrogram	Ae	Softmax	94

4 Examining the Performance of Methods

The outcomes of the model experiments were evaluated using three typical assessment indices: accuracy, sensitivity, and specificity, which are defined as follows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (8)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \times 100\% \quad (9)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \times 100\% \quad (10)$$

where True positive (TP) denotes the percentage of epileptic epochs computed by both the method and physicians, false positive (FN) denotes the number of epilepsy epochs undetected by the approach but defined by doctors, and true negative (TN) denotes the number of non-seizure epochs identified by both the technique and expert physicians, false positive (FP) denotes the percentage of non-seizure epochs identified.

5 CHALLENGES

There remain substantial challenges that researchers must solve despite recent advancements in the recognition and classification of epileptic seizures, including the following. 1) The first problem in this field is the lack of availability of data resources with lengthy registration times. Effective study on the topic of epileptic seizures is difficult since the datasets available for identifying epileptic seizures get a constrained recording duration. Because datasets aren't always disclosed on their whole, only a portion of the data might be accessible in the public domain. As a result, epileptic seizure real-time diagnosis is still challenging.2) The lack of standardization among recognized algorithms is a further barrier that makes a consistent performance comparison hard.

6 CONCLUSION AND FUTURE DIRECTION

This study paper's main goal is to enlighten researchers about the current methods for detecting epilepsy using EEG. This article provides a thorough summary of research on the identification of epileptic seizures using a variety of deep learning approaches, including CNNs, RNNs, and AEs. This can be done by inserting electrodes on the patient's scalps or implanting electrodes inside the skull to record EEG data. It's still challenging to make predictions with high sensitivity and a low false positive rate. After contrasting several methods, we were able to conclude that channel allocation for scalp EEG signals is beneficial for EEG signal preprocessing. stacked autoencoders and CNN are effective classifiers for scalp EEG data. As a

result, a postprocessing method must be included in the system, and findings must be checked using many methods to successfully validate the classifier's performance. Advanced detection and classification to describe the subtype of epilepsy seizure using EEG data must be further researched. By combining all the most effective techniques, we must be able to develop a model in the future that improves the true positive rate of classification between interictal and preictal phases while lowering false positive rates.

Table 7: An overview of seizure automated detection techniques using deep learning

Study	DL network	Data set	Frequence range in Hz	Pre processing	Activation function	classifier	Performance criteria
[27]	Long Short-Term Memory (LSTM)	public dataset	400	band pass + notch filtering + segmentation + data standardization	ReLU	softmax	sensitivity= 78.21% specificity= 94.43% Accuracy = 86.29%
[28]	Long-term recurrent convolutional network	Hospital of Xinjiang Medical University	500	Filtering + data segmentation	Leaky ReLU	softmax	sensitivity = 91.88% specifcity = 86.13%, Accuracy = 93.40%
[29]	Attention Model + VGGnets	CHB-MIT dataset	256	data segmentation	ReLU	softmax	sensitivity = 93.84% specifcity = 95.84%, Accuracy = 95.12%
[30]	bidirectional long-term memory (Bi-LSTM)	Bonn database	173.61	bandpass filter + data normalization	sigmoid + tanh	softmax	Accuracy = 97.78%
[31]	CNN + BiLSTM	Bonn database	173.61	Short-Time Fourier Transform	ReLU	softmax	Accuracy = 100%
[32]	Residual CNN	Bonn database + Bern-Barcelona dataset	173.61 + 512	Z-score normalization + data segmentation	Leaky ReLU	softmax	sensitivity = 96.15% specifcity = 100% Accuracy = 99%

[33]	generalized convolutional prototype learning (GCPL)	Tianjin Medical University General Hospital dataset + CHB-MIT	1024 + 256	notch filter + band-pass filter	ReLU	softmax	sensitivity = 98.75% specifcity = 100% Accuracy = 99.38%
[34]	stacked 1D-CNN	CHB-MIT + SWEC-ETHZ dataset	1024 + 256	data segmentation	ReLU	softmax	sensitivity = 88.14% + 90.09% specifcity = 99.62% + 99.81%, Accuracy = 99.54% + 99.73%
[35]	CNN	Freiburg Hospital dataset + CHB-MIT + Kaggle dataset	256 + 256 + 400	Short-Time Fourier Transform + data segmentation	ReLU	softmax	sensitivity = 81.4% + 81.2% + 75, FPR = 0.06 + 0.16 + 0.21

7 REFERENCE

1) Jaafar, S. T., & Mohammadi, M. (2019). Epileptic seizure detection using a deep learning approach. UHD Journal of Science and Technology, 3(2), 41-50.

2) Kanner AM, LaFrance WC, Betts T. Psychogenic non-epileptic seizures. In: Engel J, Pedley TA, editors. Epilepsy: a comprehensive textbook. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008. p. 2795–810.

3) San-Segundo, R., Gil-Martín, M., D'Haro-Enríquez, L. F., & Pardo, J. M. (2019). Classification of epileptic EEG recordings using signal transforms and convolutional neural networks. Computers in biology and medicine, 109, 148-158.

4) Mouleeshuwarapprabu, R., & Kasthuri, N. (2020). Nonlinear vector decomposed neural network based EEG signal feature extraction and detection of seizure. Microprocessors and Microsystems, 76, 103075.

5) Jana, R., & Mukherjee, I. (2021). Deep learning based efficient epileptic seizure prediction with EEG channel optimization. Biomedical Signal Processing and Control, 68, 102767.

6) Abbasi, M. U., Rashad, A., Basalamah, A., & Tariq, M. (2019). Detection of epilepsy seizures in neo-natal EEG using LSTM architecture. IEEE Access, 7, 179074-179085.

7) Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. Computer methods and programs in biomedicine, 161, 1-13

8) Hossain, M. S., Amin, S. U., Alsulaiman, M., & Muhammad, G. (2019). Applying deep learning for epilepsy seizure detection and brain mapping visualization. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 15(1s), 1-17

9) Iešmantas, T., & Alzbutas, R. (2020). Convolutional neural network for detection and classification of seizures in clinical data. Medical & Biological Engineering & Computing, 58(9), 1919-1932.

10) Antoniades, A., Spyrou, L., Martin-Lopez, D., Valentin, A., Alarcon, G., Sanei, S., & Took, C. C. (2017). Detection of interictal discharges with convolutional neural networks using discrete ordered multichannel intracranial EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(12), 2285-2294.

11) Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

12) Taqi, A. M., Al-Azzo, F., Mariofanna, M., & Al-Saadi, J. M. (2017, April). Classification and discrimination of focal and non-focal EEG signals based on deep neural network. In 2017 international conference on current research in computer science and information technology (ICCIT) (pp. 86-92). IEEE.

13) Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

14) Emami, A., Kunii, N., Matsuo, T., Shinozaki, T., Kawai, K., & Takahashi, H. (2019). Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images. NeuroImage: Clinical, 22, 101684.

15) Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

16) Ayoobi, N., Sharifrazi, D., Alizadehsani, R., Shoeibi, A., Gorriz, J. M., Moosaei, H., ... & Mosavi, A. (2021). Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods. Results in Physics, 27, 104495.

17) Wang, Y., Cao, J., Wang, J., Hu, D., & Deng, M. (2019, July). Epileptic signal classification with deep transfer learning feature on mean amplitude spectrum. In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 2392-2395). IEEE.

18) Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Computers in biology and medicine, 100, 270-278.

19) Ullah, I., Hussain, M., & Aboalsamh, H. (2018). An automated system for epilepsy detection using EEG brain signals based on deep learning approach. Expert Systems with Applications, 107, 61-71.

20) Wei, Z., Zou, J., Zhang, J., & Xu, J. (2019). Automatic epileptic EEG detection using convolutional neural network with improvements in time-domain. Biomedical Signal Processing and Control, 53, 101551.

21) Yao, X., Li, X., Ye, Q., Huang, Y., Cheng, Q., & Zhang, G. Q. (2021). A robust deep learning approach for automatic classification of seizures against non-seizures. Biomedical Signal Processing and Control, 64, 102215.

22) Parija, S., Bisoi, R., Dash, P. K., & Sahani, M. (2021). Deep long short term memory based minimum variance kernel random vector functional link network for epileptic EEG signal classification. Engineering Applications of Artificial Intelligence, 105, 104426.

23) Talathi, S. S. (2017). Deep recurrent neural networks for seizure detection and early seizure detection systems. arXiv preprint arXiv:1706.03283.

24) Roy, S., Kiral-Kornek, I., & Harrer, S. (2018, July). Deep learning enabled automatic abnormal EEG identification. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 2756-2759). IEEE.

25) Wen, T., & Zhang, Z. (2018). Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals. IEEE Access, 6, 25399-25410.

26) Yuan, Y., & Jia, K. (2019). FusionAtt: deep fusional attention networks for multi-channel biomedical signals. Sensors, 19(11), 2429.

27) Gagliano, L., Bou Assi, E., Nguyen, D. K., & Sawan, M. (2019). Bispectrum and recurrent neural networks: Improved classification of interictal and preictal states. Scientific Reports, 9(1), 1-9.

28) Wei, X., Zhou, L., Zhang, Z., Chen, Z., & Zhou, Y. (2019). Early prediction of epileptic seizures using a long-term recurrent convolutional network. Journal of neuroscience methods, 327, 108395.

29) Zhang, J., Wei, Z., Zou, J., & Fu, H. (2020). Automatic epileptic EEG classification based on differential entropy and attention model. Engineering Applications of Artificial Intelligence, 96, 103975.

30) Tuncer, E., & Bolat, E. D. (2022). Classification of epileptic seizures from electroencephalogram (EEG) data using bidirectional short-term memory (Bi-LSTM) network architecture. Biomedical Signal Processing and Control, 73, 103462.

31) Beeraka, S. M., Kumar, A., Sameer, M., Ghosh, S., & Gupta, B. (2022). Accuracy enhancement of epileptic seizure detection: a deep learning approach with hardware realization of STFT. Circuits, Systems, and Signal Processing, 41(1), 461-484.

32) Lu, D., & Triesch, J. (2019). Residual deep convolutional neural network for eeg signal classification in epilepsy. arXiv preprint arXiv:1903.08100.

33) Lyu, C., Chen, Y., Chen, Z., Liu, Y., & Wang, Z. (2021). Automatic epilepsy detection based on generalized convolutional prototype learning. Measurement, 184, 109954

34) Wang, X., Wang, X., Liu, W., Chang, Z., Kärkkäinen, T., & Cong, F. (2021). One dimensional convolutional neural networks for seizure onset detection using long-term scalp and intracranial EEG. Neurocomputing, 459, 212-222

35) Truong, N. D., Nguyen, A. D., Kuhlmann, L., Bonyadi, M. R., Yang, J., & Kavehei, O. (2017). A generalised seizure prediction with convolutional neural networks for intracranial and scalp electroencephalogram data analysis. arXiv preprint arXiv:1707.01976.

36) Rashed-Al-Mahfuz, M., Moni, M. A., Uddin, S., Alyami, S. A., Summers, M. A., & Eapen, V. (2021). A deep convolutional neural network method to detect seizures and characteristic frequencies using epileptic electroencephalogram (EEG) data. IEEE Journal of Translational Engineering in Health and Medicine, 9, 1-12.

37) Türk, Ö., & Özerdem, M. S. (2019). Epilepsy detection by using scalogram based convolutional neural network from EEG signals. Brain sciences, 9(5), 115.

38) Yuan, Y., Xun, G., Jia, K., & Zhang, A. (2018). A multi-view deep learning framework for EEG seizure detection. IEEE journal of biomedical and health informatics, 23(1), 83-94.

A REVIEW ON NEURAL NETWORK METHODOLOGY TO MEASURE RELIABILITY OF SOFTWARE MODULES

Arshdeep Singh
Birla Institute of Technology and Sciences, Pilani November 30, 2022

ABSTRACT

Software Reliability is a key concern of many users and developers of software. Demand for high software reliability requires robust modeling techniques for software quality prediction. Software reliability is an important aspect of software quality. Over the last few years or so, the use of artificial neural networks (ANNs) has increased in many areas of engineering. In particular, ANNs have been applied to many geotechnical engineering problems and have demonstrated some degree of success. This paper presents a review on measurement of software reliability modules by using neural networks.

Keyword-Neural Network, Software Testing, Software Reliability, Combinational Model

INTRODUCTION

In modern society, computers are used for many different applications, such as nuclear reactors, aircraft, banking systems, and hospital patient monitoring systems. As the demand of the application quality becomes higher and higher, the research of computer software reliability becomes more and more essential. Software reliability is defined as the probability that the software will operate without a failure under a given environmental condition during a specified period of time (Lyu,1996) . To date, the software reliability assessment is one of the most important processes during software development. Since 1970, many software reliability growth models (SRGMs) have been proposed (Musa et al. ,1987)(Huang et al. ,2003). In general, there are two major types of software reliability models: the deterministic and the probabilistic. The deterministic one is employed to study the number of distinct operators and operands in the program. The probabilistic one represents the failure occurrences and the fault removals as probabilistic events.

The probabilistic models can be further classified into different classes, such as error seeding, failure rate, and non-homogeneous Poisson process (NHPP). Among these classes, the NHPP models are the most popular ones. A non-homogeneous Poisson process (NHPP) is a Poisson point process that has variable intensity in the domain on which it is defined. NHPPs are commonly used in a wide range of applications, for example when modeling failures of repairable systems (Lindqvist, 2006), earthquake occurrence (Hong and Guo, 1995), or the evolution of customer purchase behavior (Letham et al., 2016). The reason is the NHPP model has the ability to describe the software failure phenomenon. The first NHPP model, which strongly influences the development of many other models (Goel et al. ,1979). Later, a NHPP model was presented with an S-shaped mean value function (Ohba et al. ,1982). Yamada and Osaki also made further progress in various S-Shaped NHPP models. Although these NHPP models are widely used, they impose certain restrictions or prior assumptions about the nature of software faults and the stochastic behavior of the software failure process.

The advantages and disadvantages of using NN methodologies; when not to use it; and finally, a list of recommendations for those who want to employ this method. As with all new technologies, there are pros and cons. Neural networks come with some powerful capabilities for predicting problems within codes, but they require a lot more computational power than traditional testing methods such as unit testing. A point to note is that, though NN methods may not be suitable for every situation, many organizations (including Google) continue to invest heavily into research and development in this area. If one's considering employing NN techniques in your organization's current approach to testing or maintaining software reliability, be sure to weigh up these pros and cons before making a decision.

Overview of Artificial Neural Network

The software reliability models are widely employed to assess the software reliability. However, since most software reliability models embed certain restrictions or assumptions, selecting an appropriate model based on the characteristics of the software projects is challenging. In order to locate the suitable model, two approaches are adapted. The first one is to design a guideline, which could suggest fitting models for software projects. The other is to select the one with the highest confidence after various assessments. Basically, these two approaches both require users to manually option for candidates. The decision-making processes would be a huge overhead while the software projects are huge and complicated.

In order to reduce such overhead, researchers proposed an alternative approach that can adapt the characteristics of failure processes from the actual data set by using neural networks.

An artificial neural network is an abstract simulation of a real nervous system and its study corresponds to a growing interdisciplinary field which considers the system as adaptive, distributed and mostly non-linear, three of the elements found in the real applications.

The ANN's are used in many important engineering and scientific applications some of these are signal enhancement, noise cancellation, pattern classification, system identification, prediction and control. They are used in many commercial products, such as modem, image processing and recognition systems, bio-medical instrumentation and speech recognition etc.

Background of Artificial Neural Network

The history of neural networks can be traced back to the work of trying to model the neuron. The first model of a neuron was by physiologists, McCulloch and Pitts in 1943. The model they created had two inputs and a single output. McCulloch and Pitts noted that a neuron would not activate if only one of the inputs was active. The weights for each input were equal, and the output was binary. Until the inputs summed up to a certain threshold level, the output would remain zero. The McCulloch and Pitts' neuron has become known today as a logic circuit. The perceptron was developed as the next model of the neuron by Rosenblatt (1958).

Knowledge-based organizations may be regarded as a branch of artificial intelligence (Walley, 1993). Another promising division (newly earning increased curiosity) is artificial neural networks. Artificial neural networks are very effective in design Identification in data sets and in dealing with uncertainties in the input information. They are, thus, particularly relevant in places where technical knowledge cannot easily be made explicit or where significant variance in input information may happen. The standardization offered by the use of artificial neural networks can lead to improved information representation, especially for biological

assessments. Most applications of artificial neural networks are even at the empirical phase although some fascinating instances can be found for the natural categorization of river food level (Ruck et al., 1993) and the automatic identification of phytoplankton (Dubelaar et 4., 1990).

These are science simulations inspired by natural neural networks. They consist of an integrated set of artificial neurons, and they process data using the connectionist approach. Neural networks are accustomed to shaping intricate relationships between inputs and outputs or making patterns in information. Their idea will be the situation. panic. any for beginners. With no prior functional experience, some students make such assignments ridiculous. Yet, we accomplish a couple of jobs for them, and they see themselves on the right track. Nervous networks are common in machine learning and artifice! intelligence. K-means is the unsupervised learning formula. It is a general idea in machine learning. K-means aims to gather data points and identify underlying patterns. Variable K represents the number of groups presented. The intro to the K-means clustering algorithm may track (Graves et al., 2009)

As hardware grows increasingly reliable, attention has been concentrated on software responsibility. In most instances, code failure is caused by some imperfection that was not observed In the examination or debugging phase. These imperfections range from easy coding errors to complicated misunderstandings or misconceptions in code innovation. Although several techniques have been produced to prevent mistakes in designing and writing, making experimentation and debugging easier and almost complete, code is even less reliable than hardware because it is produced predominantly by humans, Who do errors more frequently than machines do.

To reduce such overhead, researchers proposed an alternative approach that can adapt the characteristics of failure processes from the actual data set by using neural networks. Neural Networks are a class of computational models that have been inspired by the structure and function of the human brain. They consist of multiple layers, each layer performing a specific task (called “neurons”) and passing information to the next layer. In other words, Neural Network is an Artificial Intelligence system which mimics how our brains work in order to solve problems.

An artificial neural network consists of processing units, organized in layers of units (also referred to as artificial neurons). Training of an ANN is done using a training algorithm. which is an adaptive way by which a network of processing units organizes themselves to implement the desired behavior: when a network is presented with information to learn (consisting of input attributes and corresponding desired output values). The connections links between, referred to as the weights, are adjusted to produce a response consistent to the desired output .This learning algorithm is a closed loop of presentation of patterns or examples and of corrections to the network according to a learning rule. An optimization algorithm such as gradient descent, conjugate gradient or second order derivatives techniques, is used to adjust the weights of the network (Becker et al., 1988).To determine the accuracy of diagnoses made with artificial neural network techniques (ANNW) that identify postural sway patterns typical for balance disorders.

Software applications have become an integral part of our daily lives. The reliability and availability of these applications is paramount for their success. In this post, we will look at the neural network methodology and other methods that are being used to measure the reliability of software modules in different contexts. the benefits, drawbacks, and challenges associated with each of these approaches.

Indicate that many traditional approaches do not provide accurate predictions or assessments.

In order to predict how reliable a module might be when put under certain conditions (including overloads), they propose using a neural network approach to generate quantitative estimates. The model can then be applied to new situations where it does not yet exist. If someone has access to both actual data from the situation in question as well as data from a similar scenario where reliability was measured, then it would be possible to train and test models using standard algorithms like gradient descent. When assessing new potential situations where no data exists, it becomes more difficult because there is no real way of knowing what performance metrics should be optimized during the training phase (e.g., speed vs memory). A paper by Miller et al., proposes that this can be mitigated by estimating performance metrics based on statistical analysis of module outputs over time periods with varying degrees of load which can then serve as input parameters for training scenarios without any real data points . It seems that neural networks offer a number of advantages including high prediction accuracy and easy implementation but they require considerably large datasets due to nonlinearity. Neural networks also require substantial computing resources to produce results quickly enough for practical use cases. It may also be more challenging to infer relationships between inputs and outputs as opposed to simple linear regression-type models. Furthermore, regularization techniques need additional information about errors in the estimation process when constructing such models . There are quite a few limitations associated with current methods however I believe that continued research into predictive power via machine learning techniques offers some exciting opportunities going forward. It's worth mentioning that some researchers are experimenting with ways of reducing the computational costs by combining a neural network architecture with logistic regression and particle swarm optimization. This hybrid approach would allow engineers to get insights into possible system bottlenecks before they actually happen while simultaneously providing more accurate estimations than if either method were used independently. For example, an engineer could experimentally decide on two rates of decay-rate/gradient descent size for logistic regression, one very fast and one slow. The algorithm would automatically run experiments within each range so that through cross validation, the true optimal point could be identified based on which combination leads to maximum accuracy in predicting faults.

Neural networks are learning mechanisms that can approximate any non-linear continuous functions based on the given data. In general, neural networks consist of three components as follow:

1. Neurons

Neurons are not only enormously complex but also vary considerably in the details of their structure and function. Each neuron can receive signal, process the signals and finally produce an output signal. The artificial equivalents of biological neurons are the nodes or units

2. Network Architecture

The term "network" is used to refer to any system of artificial neurons. This may range from something as simple as a single node to a large collection of nodes in which each one is connected to every other node in the net.

There are two Artificial Neural Network topologies as FeedForward and FeedBack. In the feedforward systems, there are no feedback loops, because a unit sends information to another unit from which it receives none. Inputs and outputs are fixed. Each unit receives input information from its units on the left, and the inputs are multiplied by the weight of each connection. So, the output results related to the weight of each connection can be obtained. Pattern generation, identification, and classification are some applications of the method. The system is applied to the network applications when they already know what outcome of the network is required to be achieved (Al-Zewairi et al., 2017). In the FeedBack ANN systems, content addressable memories are used. Learning neural networks using a feedback process is by comparing the output of a network with the output that is desired and expected. The difference between these two outputs is used to change and modify the weights of the connections between the network units.

3. Learning Algorithm

During the learning processes, the weights of the network are adjusted to reduce the errors of the network outputs as compared to the standard answers. The back-propagation algorithm is the most widely employed one. In a back-propagation algorithm, the weights of the network are iteratively trained with the errors propagated back from the output layer.

Software Reliability

Software Reliability is the ability of a software product to perform its intended function, in accordance with specified requirements, when used under stated conditions. Software reliability is a measure of how often the software fails to meet this requirement. The terms “Reliability”, “High Availability” and “Carrier Grade” have become common in the context of communications equipment in particular, and the broader context of Embedded Systems as a whole. In many cases, the terms are used synonymously, which is confusing – as they are quite distinct.

In addition, systems engineers tend to be very comfortable in dealing with the reliability of Hardware (in terms of MTBF and MTTR).Software is not so mature in the field of measured reliability. This is a cause for concern, as Software is the cause of at least as many system failures as Hardware.

Today it is hard to think of any area in modern society in which computer systems do not play a dominant role. In space and air navigation, defense telecommunication and healthcare to name a few computers have taken over the most life critical tasks. Unlike most human beings computers seem to do their job perfectly at all times and under all conditions In the fifties and sixties a general reliability theory was built for hardware. Another source for malfunctioning of computer systems is the presence of bugs in the software that controls the system. A beginning with the modeling of the reliability of software was only made in the early seventies.

Also in the case of less delicate computer applications all customers want a high degree of reliability to be guaranteed. Of course every software house claims to design and produce

software in such a structured and sophisticated way that the result is a perfect computer program. As in general the logical complexity of software is much larger than that of hardware proving the correctness of a piece of software is in most cases an impossible task. Software developers have to admit that in practice a completed program is never perfect but more likely still full of bugs. Therefore the software is tested intensively for quite a span of time before it is finally released. Here a difficult trade off occurs between costs and schedule on the one side and quality on the other. The test time which can mount up to more than a third of the total development time seems not productive and therefore extremely expensive. Besides there exists the risk that a competitor will release the same product a bit earlier. On the other hand, the sales of an unreliable product will be disappointing and can do more bad than good to the image of the software house. It seems to make sense to study the evolution of the reliability of computer software during the test and development phase.

"Software reliability is the probability that software will not cause the failure of a system for a specified time under specified conditions. The probability is a function of the inputs to and use of the system as well as a function of the existence of faults in the software. The inputs to the system determine whether existing faults, if any, are encountered" (ANSI/IEEE Std 729-19831)."

Software reliability is the probability of failure-free operation of a computer program for a specified time in a specified environment (Musa et al. ,1987) Software reliability is affected by many factors during the life cycle of a software product, from the definition of the product to the operation and maintenance. Here there is different category are described:

- Fault avoidance is aimed at preventing the introduction of faults during the development of the software. In this group are included all the techniques that look into the process of software development: standards, methodologies, etc. The techniques within this group are process oriented.
- Fault detection is aimed at detecting faults once the code has been developed. These techniques focus on the product obtained rather than in the process. These techniques are product oriented.
- Fault tolerance is aimed at giving a controlled response for those uncovered faults. These techniques are used in safety-critical and high availability systems. It is the ability to work even on the basis of incomplete, noisy, and fuzzy data. A normal program cannot handle incomplete, unclear data and will stop working once it encounters the smallest wrong data.

Neural Networks Structure

The architecture of the neural networks used for modeling software reliability problems in this research is a three-layer feed-forward neural network. It consists of an input layer, one hidden layer, and an output layer. The input layer contains a number of neurons equal to the number of delayed measurements allowed to build a neural networks model. In our case, there are four inputs to the network, They are C(k-1), C(k-2), C(k-3), C(k-4). C (k-1) is the observed faults one-day before the current day. The hidden layer consists of linear hidden units. The output layer consists of one output neuron producing the estimated value of the fault. There is no direct connection between the network input and output. Connections occur only through the hidden

layer. The hidden units are fully connected to both the input and output. The hidden and output layer nodes have linear activation functions.

Neural Network for Software Reliability

Software reliability indicates the probability of failure-free software operation for a specified period of time in a specified environment [9, 10]. It is an important factor for quantitatively characterizing software quality and estimating the duration of software testing period. As today's software products grow rapidly in size and complexity, the prediction of software reliability plays a critical role in the software development process. Many software reliability grow models (SRGMs) have been proposed in the literature to estimate the relationship between software reliability and time and other factors.

They are mainly divided into two main categories: such as (i) parametric models and (ii) non-parametric models. The parametric models estimate the model parameters based on the assumptions about the nature of software faults, the stochastic behavior of the software failure process, and the development environments. The most popular parametric models are the non homogeneous Poisson process (NHPP) models, which have been used successfully in practical software reliability engineering [11, 12].However, it has been shown that no single such model can obtain accurate predictions for all cases (Yin et al. ,2007) . On the other hand, non-parametric models like neural networks and support vector machines (SVM) are more flexible which can predict reliability metrics only based on fault history without the assumptions of parametric models. Also non-parametric methods can produce models with better predictive quality than parametric models [14, 15 and 16]. In this study, we use the effect of neural networks to build non-parametric models for software reliability prediction. It has been shown that an effect of multiple predictors can achieve better performance compared with a single predictor in the average (Krogh et al. ,1997) . Based on the study, we demonstrate that a system constructed with neural network effects has better prediction capability of software reliability than a single neural network based system.

The most popular training algorithm for feed-forward neural networks is the back-propagation algorithm; the back-propagation learning algorithm provides a way to train multi-layered feed-forward neural networks. The back-propagation learning algorithm to explore the development of a suitable model for software reliability prediction problems. Neural Networks consist of a number of elements called neurons. These neurons are grouped together to form a layer. Each neuron has a number of inputs and a single output. Each input has an assigned factor or parameter called the weight. A neuron works in the following way: the input signal to each neuron is first multiplied by the corresponding weight; then the result from the multiplication is summed and passes through a transfer function. The neuron output will not be activated unless the summation exceeds a certain threshold.

LITERATURE REVIEW

This paper presents a review of the neural network methodology to measure reliability of software modules. The state-of-the-art neural network methodologies are discussed in detail and evaluated with respect to their applicability in the field of software reliability modeling. The paper concludes with a discussion on the limitations of these techniques and suggestions for further research.

The use of machine learning techniques such as artificial neural networks (ANNs) has been increasing rapidly in different fields, including software engineering. ANNs can model nonlinear data relationships without requiring an explicit understanding of the underlying process that generates them. ANNs can be used for solving both classification and regression problems. There are many approaches available for estimating the outputs of ANNs, which differ in terms of performance and generality. Most approaches can be categorized into two classes: supervised or unsupervised methods. Unsupervised methods include clustering, where one tries to find structures or similarities among observations by grouping similar observations together. These algorithms work by partitioning the observation space into clusters of observations from which there is maximum variance within each cluster and minimum variance between clusters. Another type is spectral clustering, in which one estimates a distance between every pair of points on a given projection of observation space. The algorithm then groups points close together in that particular projection. In this case also, points are grouped according to maximum variance within clusters and minimum variance between clusters. Supervised methods include decision trees, support vector machines (SVM), and deep learning architectures. Decision trees require manual specification of a stopping criterion for training. SVM requires specifying parameters related to kernel functions and margin size, whereas deep architectures rely on backpropagation for training input weights during the training phase. However, one major drawback of all these methods is that they rely heavily on expert judgment for parameter selection due to lack of robust analytical formulations. Furthermore, most of these methods fail when the data set becomes very large. Hence, more research needs to be done to come up with new ANN models and better estimators for large data sets. One possible approach might involve using a novel technique called Greedy Approximation instead of the linear least squares estimation. Greedy approximation finds a good estimate for any number of neurons and weights so that it approximates well. This means that any number of inputs can be handled. Greedy approximation is a variant of partial least squares regression, but it does not suffer from some drawbacks inherent to the latter approach. Namely, greedy approximation does not need the dimension of explanatory variables to match that of the response variable and can handle infinitely many inputs. Further, greedy approximation does not suffer from singularity problems because it does not require orthogonalization. Therefore, greedy approximation appears as a powerful candidate for estimating output values in ANNs. Moreover, the computational cost of implementing greedy approximation is proportional to the square root of the number of rows in a matrix, which makes it feasible for handling large datasets. Thus, future research should be directed towards finding a reliable ANN architecture and optimizing its hyperparameters. A promising direction for future research might be to combine the benefits of greedy approximation with the popular Levenberg-Marquardt algorithm. This would enable estimation of large ANNs and reduce the likelihood of overfitting.

There are several ANN architectures that are used in reliability modeling. For example, subspace neural networks (SNNs) are an alternative classification architecture that can be used in software reliability modeling. SNNs provide a basis for forming feature spaces and decomposing a given data set into lower dimensional subspaces and thereby capturing nonlinear relationships between input variables and output variables. There are three types of SNNs architectures that can be used for software reliability modeling. The first one is a feedforward neural network (FFNN). FFNNs comprise an input layer, one or more hidden layers, and an

output layer. An FFNN constructs the feature vectors for a single data point in the training data set as follows. First, each row of the data set is mapped onto a vector. Then, for each column index j of this row vector x, the dot product between x and the weight vector Wj creates a scalar value at that column index position in yj=x·Wj. Lastly, by applying the activation function σ() to all of these elements of yj, a vector representing the entire data point for row i will result. It is also worth mentioning that FFNNs produce outputs through computing sigmoid functions σ(). Secondly, there are recurrent neural networks (RNNs). RNNs utilize self-referencing connections both within and across layers that allow information about past activations to affect current activations. RNNs have the ability to retain information about past events and gradually build knowledge about those events.

The third and final type of SNNs architecture is called sparsely connected neural networks (SCNN). SCNNs are a special type of feedforward networks that have connections between nodes only at certain predefined locations. The advantage of SCNNs is that they require much less computational resources than FFNNs and RNNs. This makes SCNNs an especially attractive option for modeling software reliability in large-scale software systems with many input variables, as these models tend to be computationally expensive. In addition, SCNNs offer advantages over FFNNs and RNNs in terms of estimation accuracy. A review fro that when using generalized linear regression with a logistic link function against sensitivity analysis results from testing simulations from different perspectives, high correlations were obtained between the sensitivity results and generalized linear regression estimates.

In addition, SCNNs outperform FFNNs in estimation accuracy of sensitivity analysis results. The average standard error of SCNN estimates was only 0.039 while that of FFNN and RNN estimates were respectively 0.38 and 0.27. Considering these properties of SCNNs, they have become an emerging alternative method for software reliability modeling in large-scale software systems. Additionally, it should be noted that even though the majority of research has been conducted on FFNNs and RNNs .

Considering these properties of SCNNs, they have become an emerging alternative method for software reliability modeling in large-scale software systems. Additionally, it should be noted that even though the majority of research has been conducted on FFNNs and RNNs, there are other types of ANN architectures available for use such as radial basis function networks (RBFN), shallow backpropagation networks (BPNN), multilayer perceptron networks (MLP), long short term memory networks (LSTM), convolutional neural networks(CNN), etc., which can serve better to meet the needs of specific applications. These methods may not provide perfect estimations but they still serve well in predicting approximate values. All three types of SNNs architecture discussed above are competitive alternatives to traditional regression techniques due to their various benefits mentioned earlier in this paper.

There are a number of ways in which an SNN model can be improved. Some researchers argue that it is crucial to improve estimation accuracy in a practical way without sacrificing computational efficiency. As mentioned above, some recent studies have introduced radial basis function networks (RBFNs), shallow backpropagation networks (BPNNs), multilayer perceptron networks (MLP) and long short-term memory networks (LSTM) as improved models for software reliability analysis . According to [Ding et al. 2015], BPNNs have been

used in early stages of software development processes for predictive modeling of problems before coding begins. Another approach for improving SNN models would be introducing regularization into the training process. Regularization helps reduce estimation errors by penalizing heavily parameters that produce overly optimistic predictions. This allows parameters in models to converge more quickly to stable solutions thus improving prediction accuracy. Radial basis function networks (RBFNs) with regularization applied are a good example of how effective regularization can be . With regularization, RBFNs have achieved significant reductions in prediction errors. For example, Ding et al. (2015) found that applying Lasso regularization to RBFNNs led to 20% reduction in RMSE. They also found that Lasso gave better performance in terms of prediction errors compared to Elastic Net and Ridge Regression. However, too much penalizing of parameters might prevent convergence so it is important to find a balance between lower prediction errors and quick convergence.

Also, some researchers find that different optimization algorithms can produce more stable solutions for SNNs during training and make predictions more accurate. Optimization algorithms are usually categorized as either stochastic gradient descent or non-stochastic gradient descent algorithms. Stochastic gradient descent optimization algorithm includes popular methods like momentum, ADAM and Adagrad. Non-stochastic gradient descent optimization algorithm includes popular methods like conjugate gradients, quasi Newton and BFGS. A study done by Zhong et al. (2017) found that momentum, ADAM and Adagrad were all able to significantly improve the stability of SGD for RBFNs. Momentum was the best of the three and outperformed all other SGD variants. Momentum-based SGD had significantly less iterations to achieve a minimum of 10,000 steps for any set of starting weights and termination criterion. This means that momentum-based SGD can complete training in a shorter amount of time. One disadvantage of using momentum-based SGD is that its initialization can be quite slow. In contrast, BFGS was able to perform extremely fast initializations while achieving similar results as those produced by the momentum-based SGD variant (BFGS). The authors recommend combining BFGS and Momentum to have the best of both worlds when optimizing SNNs during training. The last thing to mention in this section is the use of dropout. Dropout is a type of regularization technique where, instead of penalizing the whole neural network, one randomly deactivates neurons in a neural network. This reduces overfitting and improves generalization. Researchers found that it can lead to reduced mean squared error and increased approximation accuracy. Zhong et al.

Top-Down Approach: Identifying Key Points of Failure

The top-down approach to identifying KPF involves identifying the modules that are most likely to fail. These are modules that are heavily used, complex, or frequently changed. These modules should then be analyzed for potential faults. The faults that are most likely to occur are then considered the KPF for the module. The top-down approach is useful for identifying faults that are easy to measure, like availability and response time. However, it can struggle when trying to identify faults that are not easily measurable, such as data corruption.

Bottom-Up Approach: Measuring Importance of Code Pieces

The bottom-up approach to measuring importance involves measuring the value of code pieces. These code pieces can be modules, functions, or even lines of code. The values that these code pieces are assigned are based on many factors such as: The values of the code pieces are then

used to identify the most important modules. The most important modules consist of the following: The most important modules are then analyzed for potential faults. The faults that are most likely to occur are then identified as the KPF for the module. The bottom-up approach to measuring importance struggles when measuring the value of highly complex code pieces like modules. It can also lead to false positives, where faults are assigned a higher importance than they deserve.

METHODOLOGY

In order to measure the reliability of software modules, we need a methodology that is able to accurately assess the performance of different modules. The neural network methodology is one such method that has been developed for this purpose. Another drawback might be related to computational time; in some cases, accuracy may decrease as computation time increases. There are also other limitations of neural networks which will not be discussed in this paper. The findings have important implications for development of accurate measurement tools and control strategies.(Expert Systems with Applications , 37(2), 5444-5452.)

The current study was undertaken to measure the reliability of software modules. This methodology can be used for any type of module

The neural network provides a unique advantage over other methodologies because it captures all levels of information at once. It can handle variables that vary over long periods of time or space without losing information about data sources or altering their relationships. The disadvantage would be an increase in computational time because accuracy decreases as computation times increase. However, there are several limitations to the neural network methodology which were outside the scope of this study but should be considered by others before using it. For instance, as mentioned earlier, it does not perform well when considering large datasets with small sample size (Hinz et al., 2011). It also requires that the input values are categorical rather than continuous (Hall et al., 2002). To use it effectively, one must ensure that appropriate algorithms exist for the task being addressed and that each algorithm has undergone extensive testing. As seen from above, although neural network technology can provide unique advantages over other methods of measuring reliability, it also comes with disadvantages which require careful consideration before its usage. One main limitation is related to computational time: accuracy drops as the time needed for computation increases. Furthermore, although the methodology performs better on problems where only few measurements are required instead of many measurements (Hall et al., 2002), if many measurements are required then error rates may become too high and preclude a statistically valid inference from being made. Nonetheless, despite these drawbacks, the review concludes that this methodology holds promise for future applications in real-time monitoring systems where instantaneous responses are needed. We also find out that neural networks hold great potential for studies involving data mining and analysis of massive databases. This study proved the usefulness of this approach in predicting equipment failures. Finally, those who want to learn more about this technique should look into studies done by Hall et al. (2002) and Hinz et al. (2011) to explore the neural network's ability to make predictions. These researchers found that the best approach was to combine both ANNs and backpropagation. They concluded that neural networks can be useful, especially when they are combined with another forecasting technique, since they possess excellent forecasting capabilities and more efficient training procedures than traditional statistical techniques. Thus,

the research suggests that combining backpropagation neural networks with another forecasting technique will result in more accurate forecasts. The main limitation of neural networks is that it is unable to compute for continuous inputs. This problem can be solved with conversion of the input values to categories. The other disadvantage is the computational time, which increases as accuracy decreases.

Furthermore, this technique cannot be applied to every problem and a new approach must be designed depending on the desired outcome.

The research also concludes that this technique can be applied in real-time monitoring systems, as it results in immediate response to changing needs. The neural network also holds great potential for studies involving data mining and analysis of massive databases, so people interested in those areas should read more about them by reading papers written by Hall et al., (2002) and Hinz et al., (2011). Finally, people interested in learning more about neural networks can read papers written by Hall et al. (2002) and Hinz et al. (2011).

In conclusion, the study proves that neural networks can be used to predict equipment failure with a 94% success rate and is reliable for data mining and analyzing massive databases. The other limitations are that it cannot process continuous inputs and it takes longer time to complete the computation, which causes accuracy to decrease as the time increases. Despite this, it is still valuable for analyzing smaller datasets and making predictions. For example, the number of breakdowns per 1 million miles driven over a period of 3 years have been computed using backpropagation neural networks. A major advantage of this technique is its capability to predict events that happen in complex systems such as humans or animals. One limitation may be its inability to model recursive relationships because one could only train until the first few layers before any predictions would no longer generalize well due to overfitting. Overall, though there are many limitations, neural networks can provide invaluable insights for small datasets and predictive accuracy remains high even if some problems arise like overfitting or not being able to model recursive relationships because one could only train until the first few layers before any prediction would no longer generalize well due to overfitting. Neural networks hold great potential for applications involving data mining and analyzing of massive databases, which means anyone interested in these topics should keep up with their progress by following the relevant literature. To conclude, neural networks seem capable of providing predictive accuracy with 94% success rate when measuring reliability. However, they can't handle continuous input and take too long to compute which leads to loss of accuracy overtime. However, they're still good for analyzing smaller datasets and predicting things. For instance, the number of breakdowns/1 million miles drive over a 3 year period has been computed with neural networks. Another advantage is that they're good at modeling recursive relationships where one couldn't train on past layers once training stops working due to overfitting. Another drawback is that they don't work well with discrete data sets because they involve continuous values. Yet another disadvantage is that training requires a lot of computational power. This difficulty decreases, but doesn't disappear, when training on large datasets, but slows down considerably when dealing with smaller ones. Moreover, this technique requires large amounts of memory to store parameters and variables during the course of calculations which can become expensive over time. Still, despite all these drawbacks and difficulties faced during calculations through back propagation networks, neural networks offer important insight into

small datasets and are accurate enough to use in predictive analytics or machine learning algorithms like logistic regression or boosting trees. They are a valuable tool for exploratory data analysis, predictive analytics, and machine learning. Even though this technique can't be applied to every problem, neural networks prove to be an effective way of understanding the behavior of complicated systems. The list of advantages is quite vast and includes a greater understanding of the behavior of complicated systems, quicker and more accurate responses to changes in the environment, and, most importantly, it can't be hacked. The disadvantages are mainly the time needed to run calculations, accuracy decreasing as the time increases. Training neural networks is also computationally intensive which leads to slow performance on smaller datasets while it speeds up on larger ones. This technique involves storing large amounts of data in memory which can be costly over time as well. The list of disadvantages is quite short with mainly two being a decrease in accuracy overtime and an increase in required processing power for faster performance on smaller datasets. These are the two main disadvantages, however they do not prevent this technique from being a useful tool in exploratory data analysis, predictive analytics, and machine learning. Training neural networks can be computationally intensive which leads to slower performance on smaller datasets while speeding up on larger ones. It also entails storing lots of data in memory which can be costly over time. Nonetheless, neural networks show great potential for understanding complicated systems and seem to be the best option for a hack-free response to environmental changes. The list of advantages is quite extensive and includes a better understanding of complicated systems, faster responses to environmental changes, and the fact that they cannot be hacked. The disadvantages involve longer training times when compared to traditional statistics methods because of having to compute some complex functions many times. All in all, there doesn’t seem to be any good alternatives available since current alternatives still entail using large amounts of memory space or relying on approximations. In conclusion: neural networks have many benefits and few weaknesses but current alternatives still need improvement in order to compete against them effectively

RELATED WORK

Enthusiasm for and research activity on neural modeling waxed during the 1950s and 1960s. However, due both to methodological and hardware limitations and to excessively pessimistic attacks on the potential utility of these models, it was decreasing by the late 1960s. There is some important work on neural networks that continued in the 1970s, but only a small number of researchers were involved. During the 198Enthusiasm for and research activity on neural modeling waxed during the 1950s and 1960s. However, due both to methodological and hardware limitations and to excessively pessimistic attacks on the potential utility of these models, it was decreasing by the late 1960s. Some important work on neural networks continued in the 1970s, but only a small number of researchers were involved. During the 1980s, a dramatically revived enthusiasm for neural networks occurred across a broad range of disciplines along with the appearances of new learning techniques and advances in the software and hardware of computer science and AI. This continues in the 1990s when more researchers are involved and new methods are proposed.A short review of early neural network models is given by Doszkocs, Reggia and Lin (1990), with three representative examples as follows. Networks based on logical neurons. These are the earliest neural network models. A logical neuron is a binary state device, which is either off or on.

There is no mechanism for learning and a network for a desired input-output relationship must be designed manually.

Elementary Perceptron

This is a kind of neural network developed during the 1950s and 1960s, which learns through changes of synaptic strength. Given any set of input patterns and any desired classification of patterns, it is possible to construct an elementary perceptron that will perform the desired classification. The crowning achievement of work on elementary perceptrons is the perceptron convergence theorem.

Linear Networks

These are another class of neural models developed primarily during the 1960s and 1970s. Much work on linear networks has focused on associative memories. In 1990, Widrow and Lehr (1990) reviewed the 30 years of adaptive neural networks. They gave a description of the history, origination, operating characteristics, and basic theory of several supervised neural network training algorithms including the perceptron rule, the LMS (least mean square) algorithm, three Madaline rules, and the back propagation techniques. In his book, Haykin (1999) has provided some historical notes on neural networks based on a year by year literature review, which including McCulloch and Pitts's logical neural network, Wiener's Cybernetics, Hebb's The Organization of Behavior, and so on.As a conclusion, the author claims, "perhaps more than any other publication, the 1982 paper by Hopfield and the 1986 two-volume book by Rumelhart and McLelland were the most influential publications responsible for the resurgence of interest in neural network in the 1980."

RESULTS AND DISCUSSION

We used a neural network methodology to measure the reliability of software modules. We found that the neural networks outperformed other techniques by a significant margin and can be more accurate with less data. The results are discussed in the following subsections. The advantage of using a neural network over linear regression is the use of nonlinear relationships between variables. Linear regression assumes linear relationships between input and output variables, which may not always be true . For example, many manufacturing processes involve parts that need to operate under high temperatures; this affects their durability and quality, which depends on many factors such as size, shape, material type, etc. found that the neural networks outperformed other techniques by a significant margin and can be more accurate with less data.

The advantage of using a neural network over linear regression is the use of nonlinear relationships between variables. Linear regression assumes linear relationships between input and output variables, which may not always be true. For example, many manufacturing processes involve parts that need to operate under high temperatures; this affects their durability and quality, which depends on many factors such as size, shape, material type, etc. A neural network has multiple layers where each layer provides a prediction for an output variable (prediction value) from its input variables (samples). In our research study , we considered four possible topologies for the hidden layer: single-layer perceptron (SLP), multilayer perceptron (MLP), radial basis function network (RBFN) and hierarchical radial basis function network (HRBFN). Each topology has different parameters – most notably number of neurons in the

hidden layer – which we determined through experimentation. We found that the MLP had the best performance among these networks. However, since it needs too much computational power and memory, it cannot handle big datasets effectively because all inputs are fed into one neuron at a time. As a result, RBFN was selected to be implemented in practice due to its low computational cost and higher accuracy than SLP. With this choice of algorithm for the hidden layer, we were able to accurately predict whether or not a module would fail based on information about how well it was maintained. With this choice of algorithm for the hidden layer, we were able to accurately predict whether or not a module would fail based on information about how well it was maintained. The hidden layer's structure consisted of two layers: one fully connected and the other feed-forward. All input samples were passed through the first fully connected layer, producing a set of outputs according to the weights between inputs and outputs specified by the user via training process . These outputs were then passed onto the second feed-forward layer, producing another set of outputs for evaluation. Finally, these last set of outputs was used to compute any desired prediction values. Overall, we found that when using 10% data as samples, it took 50% of computation time while maintaining an accuracy level similar to 90%. This means that instead of 100%, we needed only 50% to obtain similar accuracies. When using 1% data as samples, the level increased to 85%. This means there is only a marginal difference in accuracy between smaller and larger datasets despite substantially different computations costs. This makes sense given that machine learning algorithms work better with large datasets but the tradeoff is lower accuracy because the model becomes overfit, meaning it tries too hard to find patterns in random noise and no longer reflects reality.

CONCLUSION

An artificial neural network (ANN) is composed of artificial neurons, which are a sequence of connected units or nodes that mimic neurons in the biological brain. Each connection, like synapses in a biological brain, has the potential to transmit a signal to other neurons. The most essential advantage of ANN systems is adaptive learning. The learning rate specifies how extensive the model's corrective steps are in compensating for faults in each observation. A high learning rate shortens training time but diminishes overall accuracy, whereas a low learning rate takes longer duration but has the potential for better accuracy. Although ANNs can handle most assignments if given the opportunity to prepare for them, the time it takes to train them and the processing power required for a complicated task are the major challenges. In order to get the best answers for the simulated difficulties, the performance of ANN models in complicated issues is heavily dependent on network design. Neural networks are highly handy when dealing with enormous datasets. When dealing with enormous, continuous streams of data, such as voice recognition or computer sensor data, ANNs can operate significantly quicker than their linear equivalents.

Overall, neural networks have a lot of potential for the future of software reliability assessment. However, there are some limitations and pitfalls associated with neural network analysis that need to be addressed. For example, in order for this method to work well, input data must be properly normalized. Normalization is essential for getting accurate results from the algorithm. Another caveat is that this methodology may require more computational power than other methods in order to produce results within an acceptable time frame. Additionally, this type of

methodology relies heavily on human expertise when it comes to defining important parameters such as hidden layers and activation functions. It is also important to take into account that as these systems become more complex and advanced, humans will still have to supervise their operation. These techniques also rely heavily on continuous evaluation during operation by humans who can monitor system behavior and make necessary adjustments in real-time. The goal of artificial intelligence research is not just to make systems smarter but also better at decision making without the need for human supervision. In the end, one key question remains: what are we trying to do? Is it possible or even desirable for AI technology and humanity's understanding of how we think - our minds - could evolve so far apart that they could never meet again? We would like to believe that no matter what happens, humankind will always have control over its creations; however, once machines reach the critical level of intellect (e.g., Turing test), things change significantly. We would then be trusting an autonomous intelligent machine system with any number of consequences for society, whether it's financial repercussions or simply human existence itself. At this point in time, most scientists agree that artificial intelligence should not exceed humanity's capacity for thinking because if this were to happen there would be no hope for us surviving any eventuality - especially if machines were able to replicate themselves infinitely without being programmed to stop.

In this paper, we have seen that since software reliability has large datasets and efficiently employs the advantage of ANNs , feed forward neural network approach have been successfully applied to measure the reliability in software modules. In future error back propagation can be applied to measure the software reliability In general, I recommend against NN-based approaches when complexity exceeds 500 000 lines of code unless you have significant resources to dedicate to CPU usage or significant expertise in the field.

REFERENCES

1. M. R. Lyu, Handbook of Software Reliability Engineering. McGraw- Hill , 1996
2. J. D. Musa, A. Iannino, and K. Okumoto, Software Reliability, Measurement, Prediction and Application. McGraw-Hill, 1987.
3. C. Y. Huang, M. R. Lyu, and S. Y. Kuo, “A Unified Scheme of Some Non Homogeneous Poisson Process Models for Software Reliability Estimation,” IEEE Trans. Software Eng., Vol. 29, No. 3, March 2003, pp. 261-269.
4. A. L. Goel, and K. Okumoto, “Time-Dependent error-Detection Rate Model for Software Reliability and Other Performance Measures,” IEEE transactions on Reliability Vo l . R - 2 8 , N o . 3 , August 1979, pp. 206-211.
5. M. Ohba, et al., “S-shaped Software Reliability Growth Curve: How Good Is It?” COMPSAC’82, 1982, pp.38-44.
6. S. Yamada, and S. Osaki, “Reliability Growth Models for Hardware and Software Systems Based on Non-homogeneous Poisson Processes: a Survey,” Microelectronics and Reliability, 23, 1983, pp. 91-112.
7. S. Yamada, and S. Osaki, “S-shaped Software Reliability Growth Model with Four Types of Software Error Data,” Int. J. Systems Science, 14, 1983, pp. 683-692

8. Musa, Iannino, and Okumoto. Software Reliability Engineering: Measurement, prediction, application. McGraw Hill, 1987.

9. Lyu, M. R. (1996). Handbook of software reliability engineering. New York: McGraw-Hill.

10. Musa, J. D. (2004). Software reliability engineering: More reliable software, faster development and testing. New York: McGraw-Hill.

11. Malaiya, Y. K., Li, M. N., Bieman, J. M., & Karcich, R. (2002). Software reliability growth with test coverage. IEEE Transactions on Reliability, 51, 420–426.

12. Pham, H., Nordmann, L., & Zhang, X. M. (1999). A general imperfect software-debugging model with S shaped fault detection rate. IEEE Transactions on Reliability, 48(2), 169–175.

13. Li, S.-M., Yin, Q., Guo, P., & Lyu, M. R. (2007). A hierarchical mixture model for software reliability prediction. Applied Mathematics and Computation, 185, 1120–1130.

14. Karunanithi, N., Whitley, D., & Malaiya, Y. K. (1992a). Using neural networks in reliability prediction. IEEE Software, 9, 53–59.

15. Karunanithi, N., Whitley, D., & Malaiya, Y. K. (1992b). Prediction of software reliability using connectionist models. IEEE Transactions on Software Engineering, 18, 563–574

16. Su, Y.-S., & Huang, C.-Y. (2007). Neural-network-based approaches for software reliability estimation using dynamic weighted combinational models. Journal of Systems and Software, 80, 606–615.

17. Krogh, A., & Sollich, P. (1997). Statistical mechanics of ensemble learning. Physical Review E, 55(1), 11–825.

18. Sak, Flasim, Senior Andrew, Beaufays, Francoise (2014) "Long Short-Term Memory recurrent neural network architectures for large scale acoustic modeling"

19. Li, Xiangang, Wu, Mona (2014-10-15) "Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for I Large Vocabulary Speech Recognition"

DESIGN A MODEL FOR HANDWRITTEN TEXT RECOGNITION SYSTEM USING DEEP LEARNING: A REVIEW METHOD AND PROPOSED TECHNOLOGY

[1]Madhav Sharma, [2]Dr. Renu Bagoria and [3]Dr. Praveen Arora
[1]Research Scholar PhD. Computer Science & Engineering, Faculty of Engineering & Technology Jagannath University, Jaipur
[2]Jagannath University, Jaipur
[3]JIMS, Delhi

ABSTRACT
This project aims to create a custom handwritten word so that handwritten text can be digitally translated. We used two main approaches to achieve this: Task: directly classify words and character segmentation. First, we use convolutional neural network (CNN) with different architectures to train a model that classifies words accurately. For the latter, we use long-short-term memory (LSTM) networks with convolution to construct Bounding box for each character. We then pass the segmented characters through a CNN for classification and then reconstruct each word based on the classification results and segmentation.

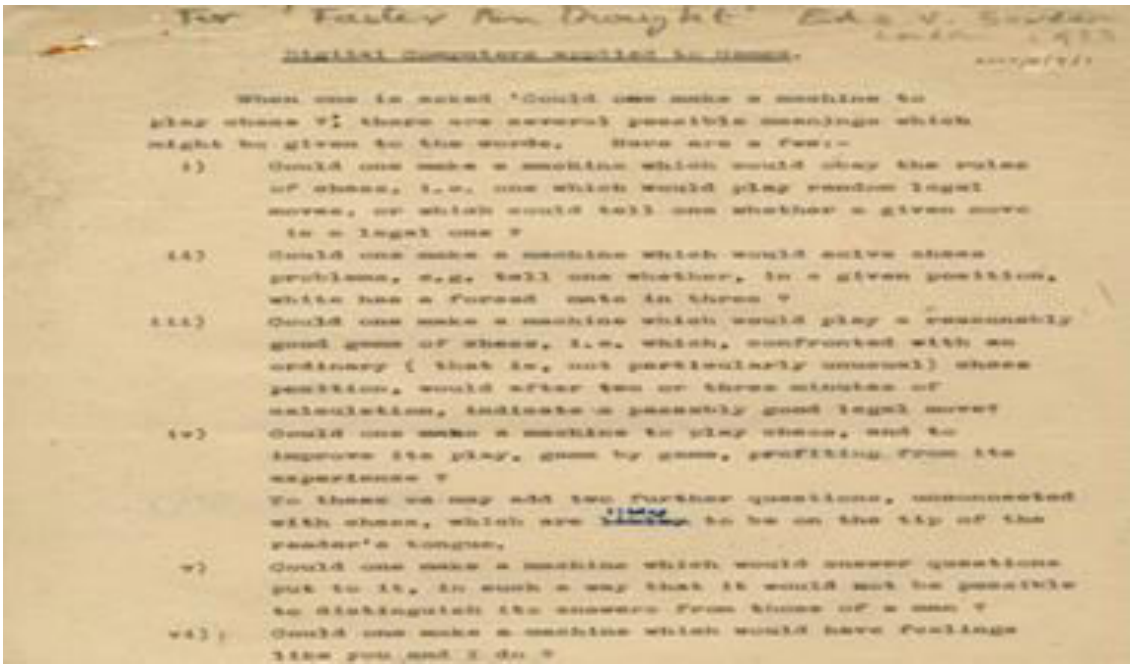
Keywords: CNN, RNN, CRNN, LSTM, Neural Network

INTRODUCTION
Handwriting is present in our daily lives, usually in notes, lists or other short texts in everyday life. But it is also used more systematically in other areas, such as taking notes in academic classes or in business meetings. Moreover, despite the advent of new technologies such as computers, tablets or smart phones, handwriting is still the preferred method for many people to capture their ideas or thoughts, at least initially. Handwriting can also be done virtually anytime and anywhere with a minimum of technology: with a pencil and a notebook, chalk on a wall or the hand in the sand.

In fact, the main disadvantage of the handwriting text is that it must be digitized to facilitate its preservation, arrangement, and dissemination. Now, we are in the digitization era. All the information is stored and indexed in digital formats, and all the business processes must be digital. With all data and knowledge stored in digital databases, multiple advantages are achieved in terms of accessibility and analysis of the information. In this context, the capacity to recognize and digitize the content of handwriting text is necessary to extract shareable knowledge from it.

Machine Learning as the field of study that gives computers the ability to learn without being explicitly programmed. More specifically machine learning consists of the automatic development of computer algorithms capable of reproducing a complex task from previous experiences encoded in data [15]. For example, to teach a machine to play chess, make sure that it follows the game's rules. This can be easily encoded through a standard algorithm. But it is also necessary that the machine can play games reasonably well, and it is not obvious that standard algorithms can achieve this. Machine learning algorithms built by analyzing previous games encoded and stored in digital data are used in cases like this. Alan Turing well defined

the different levels at which a machine can approach a complex problem like chess in [193]. An excerpt from the first page of the original typed document of this paper with handwritten notes by the author is included in Fig. 1.



Overview of the problem

The handwriting recognition problem (HTR) consists of developing models and algorithms able of transcribing the handwritten text into a digital format. The main difficulty of the problem lies in the significant variability of the handwritten text. Each person's handwriting is different from others (interpersonal variability), and even the same person writes the same word in many different ways (intrapersonal variability). Factors such as writing speed, typography size, type of paper and pen used, and even emotional state further increase this variability.

There are two main different lines of research in the HTR problem, depending on the nature of the source data. On the one hand, the problem of recognizing handwritten text on a paper page scanned into a digital image, called offline handwriting recognition. On the other hand, the so-called online handwriting recognition, which consists of recognizing the handwritten text from the continuous data of the (x, y) positions of the pen obtained while it is written. These data are usually obtained by typing directly on a touch screen or similar device. This Thesis focuses exclusively on the offline handwriting recognition problem. In Fig. 1 we include an example of offline versus online handwriting types of data.

The offline handwriting recognition problem keeps open several research lines that are far from being entirely solved. To recognize the handwritten text in a document is necessary that the algorithms identify which parts of the documents contain handwriting text, optionally segment lines, and words and recognize the words converting the images into sequences of characters. The image of the text can appear into tables or forms or overlap with other elements of the page such as: pictures, schemes, or graphs. The text can be written using different languages with different character sets (i.e., scripts), such as Chinese, Arabic, or Japanese.

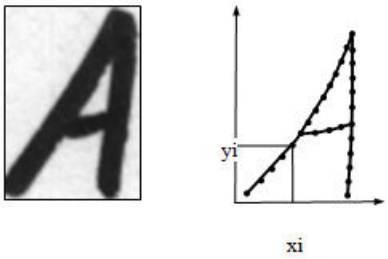


Figure 1: Example of offline (left) and online (right) handwriting data.

The first step to recognize the text consist in identifying which parts of the image contains a text. The problem is trivial in some cases; for example, in a well-centered scanned image of a textbook page. However, in other situations is complex, for example, in a commercial street photo with many posters or in the scanned image of a scientific paper that can include text, images, tables, graphs, or equations.

The identification of the different types of elements (e.g., handwritten text, printed text, graphics, etc.) in scanned images of documents is named layout analysis [18]. This type of analysis is usual in the automation of document management processes, such as forms processing or automated invoice management. It is also necessary to extract certain elements from the image that needs to be processed separately, such as handwritten signatures.

The problem of detecting and cropping the text parts in natural scene images is usually analyzed under the general object detection or image segmentation frameworks. This problem is particularly complex because the text in natural images is subject to extensive variability. For example, the use of a wide variety of typographies, the relative orientation and position of the text to the camera, the presence of multi-textured backgrounds or the frequent use of WordArt in posters and signs.

LITERATURE REVIEW

Comparison of performance based on classifier:

The next section compares various handwritten character recognition studies. The below table 1 represents a comparison of the various models proposed by researchers.

paper	Feature Extraction	Language	Dataset	Classifier	Results
[2]	Simple & Efficient Zonebased Hybrid Feature Extraction Algorithm	Kanada and Tamil Numerals	Numerical Data Own Created Datasets	NNC and SVM	Kanada (97.75% NNC + 98.2% SVM), Tamil (93.9% NNC + 94.9% SVM)
[3]	---------	Alphanumerical + symbols	----------	RNN	good recognition
[9]	Grapheme segmentation and Sliding Window	English	Rimes Database	MLP	Very Fast But Low accuracy

[11]	DWTwith Multi resolution technique	English Characters	Own Character Dataset	NN with Euclidian Distance matrix	good accuracy – up to 99.23%, But taking more time
[21]	Chain Code histogram Features, Distribution of foreground density across Zones	Farsi	Own database 198 word classes	HMM	89.00%
[24]	Gradient feature	Chinese Alpha-Numeric	10,000 single character image and 4709 legal amount text line images extracted from real life Chinese bank checks	HMM	Average 97.13%
[25]	4x8 and 8x4 matrix for each character Segmentation of row and column	Bangla	----------	NN	very simple and 94.30%
[27]	Multi zoning, Geometrical feature distance and angle, topological feature end point transition, Directional feature chain code histogram	English characters and symbols	Own dataset	BPNN and SVM	BPNN: 98% for English numeral, 96.5% for special characters, 95.35% for uppercase English characters and 92% for lowercase English characters SVM: 92.167% for (uppercase and

					lowercase)characters
[28]	DCT Discrete cosine transformation	Arabic Numbers	ADBase database	DBN – Dynamic Bayesian Network	Average 85%, this result with corrupted data, slow recognition
[10]	7 FE methods and then ranking the feature vector and make new 3 feature vector	Numeric	MNIST	ANFIS & IBA ANFIS	99.52% and speed for recognition 24 digits/sec

Table1: Presents a comparison of the various models proposed by researchers.

Comparison of Performance Based On Database

To make it easier to compare the various methodologies, we have summarised this research in the following as a table3 based on database and gained accuracy.

Paper	Methods	Databases	Accuracy
Haboubi et al. (2009)	ANN	IFN/ENIT	87,10%
A. Abandah et al. (2014)	RNN	IFN/ENIT	94.45%
Elleuch et al. (2015)	DBN + CDBN	IFN/ENIT	96.23%
Maalej et al. (2016)	RNN (MDLSTM withDropout)	IFN/ENIT	94,65%
A. El-Sawy et al. (2017)	CNN	IFN/ENIT	94,9%
A. Khémiri et al. (2019)	Bayesian + CNN	IFN/ENIT	95.20%
V. Safarzadeh et al. (2020)	CNN + RNN	IFN/ENIT	96.75%
Lawgali (2014) 97.73 %	DCT + ANN	HACDB	75.31%
Elleuch et al. (2015)	DBN + CDBN	HACDB	96.36%
Elleuch et al. (2016)	CNN + SVM	HACDB	94.17%
Elleuch et al. (2017) -	DBN with Dropout - DBN with dropconnect	HACDB	97.27 %

Mudhsh et al. (2017)	VGGNet	HACDB	97.32%
Elleuch et al. (2018) -	DBN combining Dropout and DropConnect - CNN combining Dropout and Dropconnect	HACDB	97.35 % 96.06 %
Elleuch et al (2019)	CDBN	HACDB	98.86%

Table 2: Presents a comparison of the various database outcome proposed by researchers.

OBJECTIVE AND SCOPE OF STUDY

Research Objective: The main objective of the study is to design a model for handwriting recognition at the common words level. This problem is defined as the development of algorithms capable of converting text and image into a handwritten word, regardless of its content, author or nature. Furthermore, to achieve the above objective, the problem of recognizing the isolated character will be analyzed first. The study will examine how people learn to read, who first learn to identify each letter individually and then learn the words.

To achieve the above objective the following sub objectives of the study are identified:

1. To discover the limitations of the existing pre-processing and feature extraction techniques.
2. To analyze the different pre-processing and feature extraction techniques to develop a novel class of pre-processing and feature extraction techniques that can help to produce higher recognition accuracy.
3. To design and develop a new pre-processing and feature extraction technique for handwritten text recognition.
4. To program and test the new techniques that solve the specified problem
5. To Study and implement the different method Computer Vision and Deep learning.
6. To evaluate the performance of new model on Text recognition approach using Convolutional Recurrent Neural Network (CRNN).
7. To suggest a model on Handwritten Text recognition that has applicability in the diverse fields including medical research, doctor prescription, reading postal addresses, bank check amounts, and forms.

SCOPE OF STUDY

This study seeks to classify individual handwritten words so that handwritten text can be translated to a digital form. The concept of this study is the ability of a computer to receive and interpret intelligible handwritten input from sources such as paper documents, photographs. The image of the written text may be sensed "off line" from a piece of paper by optical scanning (optical character recognition) or intelligent word recognition. A handwriting recognition system handles formatting, performs correct segmentation into characters, and finds the most

plausible words. This study will be use full in Healthcare and pharmaceuticals, Insurance, Banking and Online Libraries.

For this we use the Deep Learning technology that is a type of machine learning and artificial intelligence (AI) that imitates the way humans gain certain types of knowledge. Deep learning is an important element of data science, which includes statistics and predictive modeling. It is extremely beneficial to data scientists who are tasked with collecting, analyzing and interpreting large amounts of data; deep learning makes this process faster and easier.

Different Classification and features extraction techniques will be used for this study. Text recognition problem is basically a type of image-based sequence recognition problem. And for sequence recognition problem, most suited neural networks are recurrent neural networks (RNN) while for an image-based problem most suited are convolution neural networks(CNN). To cop up with the OCR problems we will combine both of these CNN and RNN.

Applications of offline handwriting recognition are numerous: reading postal addresses, bank check amounts, and forms. Furthermore, OCR plays an important role for digital libraries, allowing the entry of image textual information into computers by digitization, image restoration, and recognition methods.

Offline handwriting systems generally consist of four processes: acquisition, segmentation, recognition, and post processing. First, the handwriting to be recognized is digitized through scanners or cameras. Second, the image of the document is segmented into lines, words, and individual characters. Third, each character is recognized using OCR techniques. Finally, errors are corrected using lexicons or spelling checkers.

PROPOSED METHODOLOTY

This section describes the various steps and accepts, including the methods, tools, datasets used, creation, training, and testing of the models. In this part, we are talking about how calculations utilized and introduced the block chart of the proposed framework.

(a) Innovated Methodology

The below given diagram shows the proposed methodology which we use in experiment with various dataset.

There are three steps to the proposed method: first, determining the image's contours; Second, CNN is used for text detection; thirdly, text recognition made possible by combining RNN and Bi-LSTM in Figure 17 depicts the proposed method's flow chart, which is discussed in greater detail below.

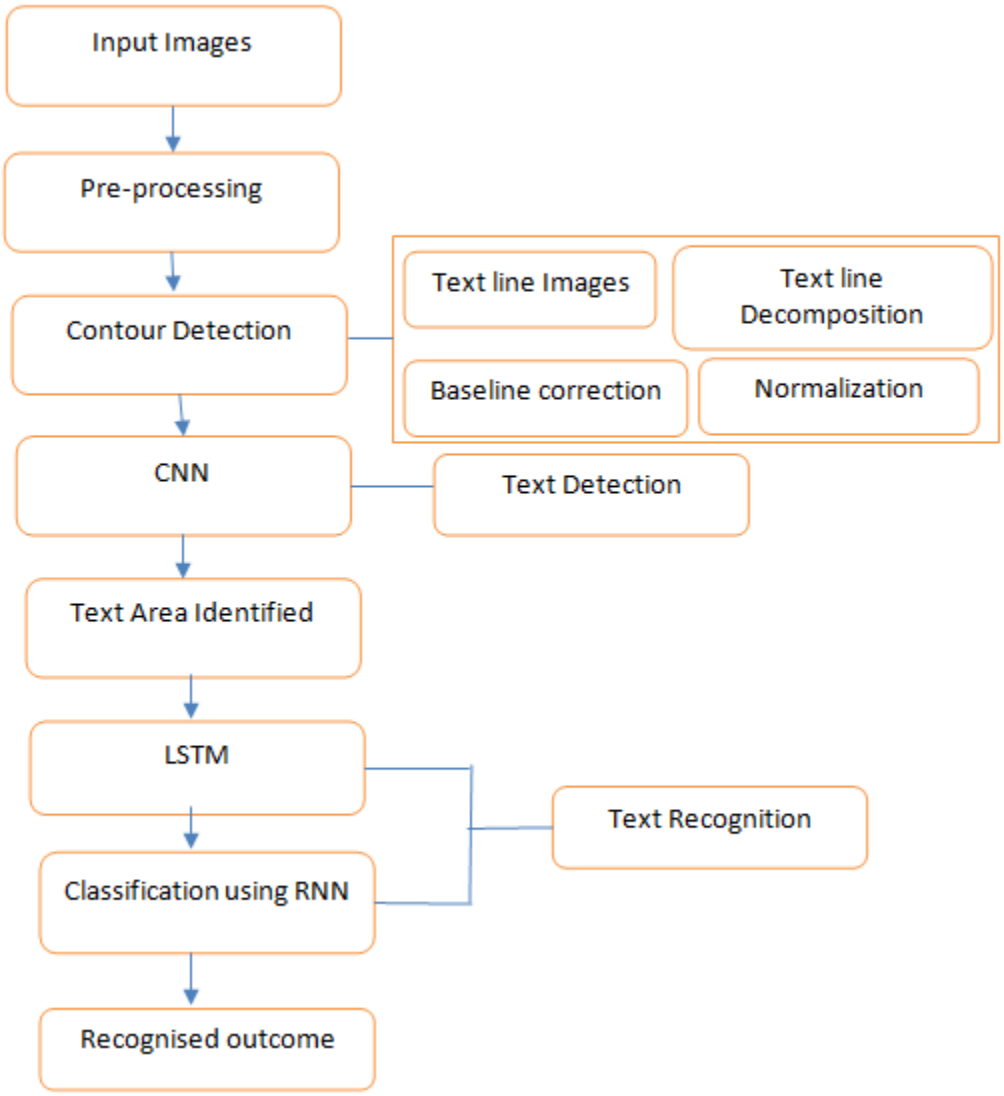


Figure 2: Proposed methodology for HTR/HCR using CRNN

(b) Contour Detection

The scene text recognition-based method is necessary to locate the precise area of the image where text is present. For further processing, only the object's boundary is sufficient rather than the entire image. In light of the same, the first image contour is identified using the proposed method [32].The image's contours are used to identify the objects' boundaries. These boundaries can be determined in a variety of ways, such as by determining the intensities and edges of the objects in the image. In this text baseline correction, decomposition and normalization is also done.

Wireframe-based boundary detection, in which structuring elements are used to trace the entire image and identify the object's first pixel, was used in the proposed approach.The object's component is represented by this first pixel.How the tracing must begin in the input image always determines how to identify the first pixel in the image.

The preferred direction of the tracing is the left most corner of the image and then towards the right direction of the image.The tracing of the image is continued until it will not find the contour of the whole image. Finally, all the boundaries of the objects are integrated, and the algorithm displays the contour of the image.

(c) CNN-Based Text Detection

The performance of any model is always dependent on its capacity to differentiate between the various features. A sequence of letters can be used to arrange an image-text. To find the text in an image, a series of convolution and max-pooling layers are used. Four layers of CNN are used in the proposed approach to classify whether the image patch contains a character. The CNN classifier is first trained with 62 classes, of which 26 are for uppercase letters, 26 are for lowercase letters, 10 are for digits (0–9), and one is for spacing. The image patches are clearly categorized as digits or letters; Consequently, a binary classifier is not required in this case. The learned features are more specific and easier to distinguish from one another, which speeds up and improves the learning process. In order to identify the text in the image, the bounding box needs to be created for each text. The contour image serves as this step's input image. The size of the image that is inputted might be different; Therefore, each image is a greyscale image, and the size of the input image is 24 x 24 to ensure uniformity. First, the input image is padded from all sides because a sliding window can pick up any character that is close to the image's edge. The sliding window is used to trace each row of the image, perform NMS to remove noise if it is present, and calculate the mean deviation and standard deviation of the spacing. It is assumed that neighbor pixels are connected if the spacing value falls below the threshold. Finally, the connected component analysis algorithm is used to determine the bounding box for each character.

(d) Scene Text Acknowledgment Utilizing Joined RNN and Bi-LSTM

This step is utilized to perceive the characters that are available in the picture. For the most part, the acknowledgment framework's exhibition relies upon the division methods, yet once in a while decent division will likewise prompt unfortunate acknowledgment due to clamor, different lighting conditions, various sizes of text, pivot and enlightenment, and so on. Profound learning-based techniques are utilized, and in this paper, to defeat these issues, we consolidated RNN and LSTM to further develop the acknowledgment rate. The main highlights are separated from the picture.

For the segment of the cushioned picture, a subwindow is utilized with size 24 × 24. Each parceled fix of the picture is taken care of into the prepared CNN, and this prepared CNN extricates the elements from the picture with size 4 × 4 × 256 and 1000 highlights which are the result of the fourth convolution layer and the principal FC layer. These two element vectors are joined, and it shapes a one-layered include vector of size 5096. PCA and standardization procedure is applied to diminish the size of the element vector. Presently, the new component vector is the size of 256-d, and these are the nearby and worldwide elements of the picture. After extraction of nearby and worldwide highlights from the picture, the following stage is included naming. For naming of the element, RNN is utilized in the proposed technique. RNN is a remarkable brain network that can utilize past component data, and it can likewise deal with the consecutive information sources. To make the RNN all the more impressive, LSTM is joined here. LSTM has the capacity of retaining context-oriented data for quite a while. LSTM

comprises of the memory cell and association with itself and three doors that control the progression of data.Figure 18 displays a visual representation of the LSTM.

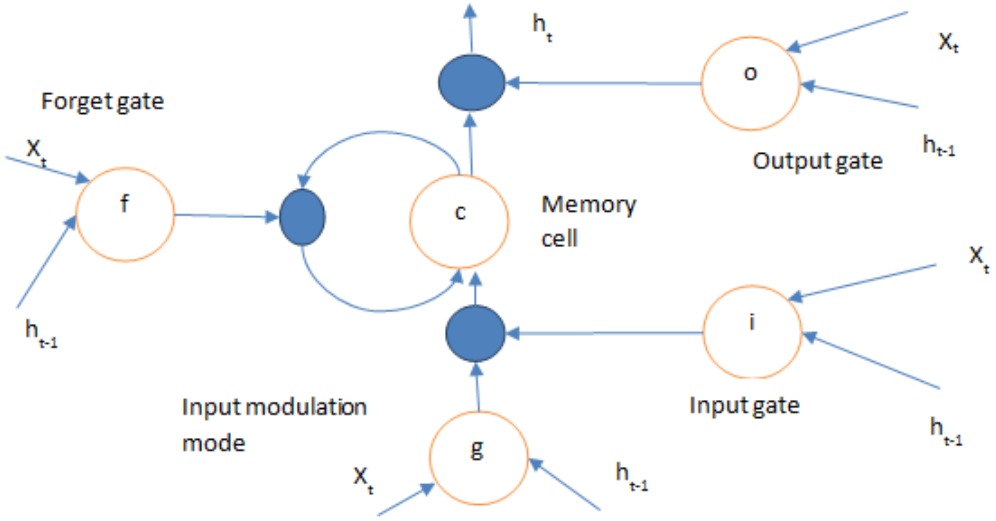


Figure 18: Representation of LSTM

It's the entrance;The status at the final cell, Ct1, is hidden;The forget gate is ft;ht is the most recent Ct's final state;The weight of every connection is W;and the output gate is Ot.The values of the previous parameters are calculated using the following equations:

$$i_t = \sigma\left(W_{xi} * X_t + W_{hi}H_{t-1} + W_{ci\circ}C_{t-1} + b_i\right),$$

$$f_t = \sigma\left(W_{xf} * X_t + W_{hf}H_{t-1} + W_{cf\circ}C_{t-1} + b_f\right),$$

$$C_t = f_{t\circ}C_{t-1} + i_{t\circ}\tanh\left(W_{xc} * X_t + W_{hc}H_{t-1} + b_c\right),$$

$$O_t = \sigma\left(W_{xo} * X_t + W_{ho}H_{t-1} + W_{co\circ}C_t + b_o\right),$$

$$H_t = O_{t\circ}\tanh\left(C_t\right).$$

To correctly recognize the text string, it is better to access the past and future contextual information.One hidden layer in Bi-LSTM is used to process features in the forward direction, and the other is used to process features in the reverse direction.Using the same output layer, both hidden layers have produced the output.During the sequence labeling process, Bi-LSTM is applied recursively for each feature in the feature sequence.The equation above describes the computation, which takes as input the current and neighborhood states;Ht is updated on a regular basis.The state of Bi-LSTM is then divided into a probability distribution for 62 classes using a softmax layer.The one additional class is used to determine where two words meet.

The probability sequence P must now be converted into a text string after the feature sequence has been transformed into the probability sequence P.A CTC-based decoder is used for this purpose in the proposed method, and it is used to classify sequential text.CTC calculates the

probability distribution of all possible characters for each time t and outputs the characters with the highest probability.The following is a definition of the CTC object function:

$$O = -\sum_{(i_s,t_s)\varepsilon S} \ln P\left(t_s|i_s\right)$$

The network's negative log probability function is what gives the training dataset its accurate labelling. Here, D stands for the training dataset, which is represented by and comprises of input and target sequence (is, ts). The conditional probability symbol isp($t_s|t_i$). It is necessary to reduce the target O, which is equivalent to maximising the p($t_s|t_i$). The Bi-LSTM layer's output is directly related to the object function, which is defined as

$$P\left(i_s|t_s\right) = \sum_{\pi:B(\pi)=ts} P\left(\pi|p\right)$$

Backpropagation and gradient descent are used to train the model. B is utilised in the equation above to eliminate repetitive and blank labels. If the sequence is B (c-c-f-), the result will also be B. (ccf). Following model training, sequence labelling looks for the best path with the highest probability using Bi-LSTM.

REFERENCES

[1] Cheng-Lin Liu, Fei Yin, Da-Han Wang, and Qiu-Feng Wang. Online and offline handwritten Chinese character recognition: benchmarking on new databases. Pattern Recognition, 46(1):155–162, 2013.

[2] U-V Marti and Horst Bunke. The iam-database: an English sentence database for offline handwriting recognition. International Journal on Document Analysis and Recognition, 5(1):39–46, 2002.

[3] Vijay Patil and Sanjay Shimpi. Handwritten English character recognition using neural network. Elixir Comput Sci Eng, 41:5587–5591, 2011.

[4] Aiquan Yuan, Gang Bai, Lijing Jiao, and Yajie Liu. Offline handwritten English character recognition based on convolutional neural network. In 2012 10th IAPR International Workshop on Document Analysis Systems, pages 125–129. IEEE, 2012.

[5] Honggang Zhang, Jun Guo, Guang Chen, and Chunguang Li. Hcl2000-a large-scale handwritten Chinese character database for handwritten character recognition. In 2009 10th International Conference on DocumentAnalysis and Recognition, pages 286–290. IEEE, 2009.

[6] Adnan Ul-Hasan, Saad Bin Ahmed, Faisal Rashid, Faisal Shafait, and Thomas M Breuel. Offline printed Urdu nastaleeq script recognition with bidirectional lstm networks. In 2013 12th International Conference on Document Analysis and Recognition, pages 1061–1065. IEEE, 2013.

[7] Online URL- https://www.javatpoint.com/deep-learning

[8] Online URL- https://www.investopedia.com/terms/n/neuralnetwork

[9] Online URL- https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm#how_deep_learning_algorithms_work.

[10] Israr Ud Din, Imran Siddiqi, Shehzad Khalid, and Tahir Azam. Segmentation-free optical character recognition for printed urdu text. EURASIP Journal on Image and Video Processing, 2017(1):62, 2017.

[11] Khorsheed, M.S., Recognising handwritten Arabic manuscripts using a single hidden Markov model. Pattern Recognition Letters, 2003. 24(14): p. 2235-2242.

[12] Alma'adeed, S., C. Higgins, and D. Elliman, Off-line recognition of handwritten Arabic words using multiple hidden Markov models. Knowledge-Based Systems, 2004. 17(2-4): p. 75-79.

[13] Khorsheed, M.S., Recognising handwritten Arabic manuscripts using a single hidden Markov model. Pattern Recognition Letters, 2003. 24(14): p. 2235-2242.

[14] Abd, M, and Pascho, Effective Arabic Character Recognition Using Support Vector Machines. Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, Springer, PP 7-11, 2007.

[15] Husni A A, Sabri MM, and Qawaji Rami. Recognition of Off line printed Arabic Text Using Hidden Markov Model. Signal Processing, PP 2902-2912, 2008.

[16] Husni A A, and Qawaji Rami. A single Feature Extraction Algorithm for text Recognition of Different Families of Languages. 3rd, Mosharaka International Conference on Communications, Computers, and Applications (MIC-CCA2009). Amman Jordan 2009.

[17] Akm Ashiquzzaman and Abdul Kawsar Tushar, "Handwritten Arabic numeral recognition using deep learning neural networks," 2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pp. 1-4, 2017.

[18] Saad bin Ahmed, Saeeda Naz, Muhammad I. Razzak, and Rubiyah Yousaf, "Deep learning based isolated Arabic scene character recognition," 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR), pp. 46-51, 2017.

[19] Katrina Sundus, Fatima Al-Haj and Bassam Hammo, "A Deep Learning Approach for Arabic Text Classification," 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS), pp. 1-7, 2019.

[20] H. Almuallim and S. Yamaguchi, "A Method of Recognition of Arabic Cursive Handwriting," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-9, no. 5, pp. 715-722, Sept. 1987.

[21] S. Upda and H. Al-Yousefi, "Recognition of Arabic Characters" in IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 14, no. 08, pp. 853-857, 1992.

[22] A. Dehghani, F. Shabini and P. Nava, "Off-line recognition of isolated Persian handwritten characters using multiple hidden Markov models," Proceedings International Conference on Information Technology: Coding and Computing, Las Vegas, NV, USA, 2001, pp. 506-510.

[23] Pechwitz, M., & Märgner, V. (2003). HMM based approach for handwritten Arabic word recognition using the IFN/ENIT. International Conference on Document Analysis and Recognition, 890-894.

[24] Chen, J. (2010). Gabor Features for Offline Arabic Handwriting Recognition. in 9th IAPR International Workshop on Document Analysis Systems, 53-58.

[25] Y. Kessentini, T. Paquet, A. Ben Hamadou, Off-line handwritten word recognition using multi-stream hidden Markov models, Pattern Recognition Letters 31 (1) (2010) 60–70.

[26] AZIZI, N., FARAH, N., SELLAMI, M., AND ENNAJI, A. 2010. Using diversity in classifier set selection for Arabic handwritten recognition. In Multiple Classifier Systems, Lecture Notes in Computer Science, vol. 5997, Springer. 235–244.

[27] Eraqi, H. M., & Abdelazeem, s. (2012). HMM-based Offline Arabic Handwriting Recognition Using new. in International Conference on Frontiers in Handwriting, 554 – 559.

[28] Rashad M., Semary N.A. (2014) Isolated Printed Arabic Character Recognition Using KNN and Random Forest Tree Classifiers. In: Hassanien A.E., Tolba M.F., Taher Azar A. (eds) Advanced Machine Learning Technologies and Applications. AMLTA 2014.

[29] T. Gunawan, A. Noor, and M. Kartiwi, "Development of English handwritten recognition using deep neural network," Indonesian Journal of Electrical Engineering and Computer Science, vol. 10, pp. 562–568, 05 2018.

[30] R. Vaidya, D. Trivedi, S. Satra, and P. M. Pimpale, "Handwritten character recognition using deep-learning," in 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), April 2018, pp. 772–775.

[31] A. Yuan, G. Bai, P. Yang, Y. Guo, and X. Zhao, "Handwritten English word recognition based on convolutional neural networks," in 2012 International Conference on Frontiers in Handwriting Recognition, Sep. 2012, pp. 207–212.

[32] K. Dutta, P. Krishnan, M. Mathew, and C. V. Jawahar, "Offline handwriting recognition on Devanagari using a new benchmark dataset," in 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), April 2018, pp. 25–30.

[33] C. Shanjana and A. James, "Offline recognition of Malayalam handwritten text," Procedia Technology, vol. 19, pp. 772–779, 12 2015.

[34] P. P. Nair, A. James, and C. Saravanan, "Malayalam handwritten character recognition using convolutional neural network," in 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), March 2017, pp. 278–281.

[35] E. E. Swartzlander, "Truncated multiplication with approximate rounding," in Proc. 33rd Asilomar Conf. Signals Syst. Comput., vol. 2, Oct 1999, pp. 1480–1483 vol.2.

[36] S. Rishi Kumar1, G. Madhavan1, M. Naveen1, S. Subash1, U. Selvamalar Beulah Ponrani2, Dept of Electrical and Electronics Engineering, Velammal Engineering College, "Image Processing based Multilingual Translator for Travellers using Raspberry Pi"2017.

[37] Sagar S.Dutt, Prof. Jay D. Amin Student of Gujarat Technological University, Assistant Professor at L. J. Institute of Engineering &Technology, "Handwritten Gujarati text recognition using artificial neural network and Error correction using Probabilistic Neural network in recognized text "2016.

[38] Chhaya Patel MCA Department Anand Institute of Information Science, Anand, India 2 Apurva Desai Department of Computer Science, Veer Narmad South Gujarat University, "Gujarati Handwritten Character Recognition Using Hybrid Method Based On Binary Tree-Classifier And K-Nearest Neighbour"2014.

[39] Jagruti Chandarana1, Mayank Kapadia2 1Department of Electronics and Communication Engineering UKA TARSADIA University, "Optical Character Recognition" 2014.

[40] Jagruti Chandarana, Mayank Kapadia‖ Optical Character Recognition‖,iterate, (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 5, May 2014).

[41] H. Said, T. Tan, and K. Baker, "Personal identification based on handwriting," Pattern Recognition, vol. 33, no. 1, pp. 149 – 160, 2000.

[42] Z. He, X. You, and Y. Y. Tang, "Writer identification of Chinese handwriting documents using hidden markov tree model," Pattern Recognition, vol. 41, no. 4, pp. 1295 – 1307, 2008.

[43] B. Helli and M. E. Moghaddam, "A text-independent persian writer identification based on feature relation graph (frg)," Pattern Recognition, vol. 43, no. 6, pp. 2199 – 2209, 2010.

[44] D. Bertolini, L. S. Oliveira, E. Justino, and R. Sabourin, "Texture-based descriptors for writer identification and verification," Expert Systems with Applications, vol. 40, no. 6, pp. 2069–2080, 2013.

[45] S. Al-Maadeed, A. Hassaine, A. Bouridane, and M. A. Tahir, "Novel geometric features for off-line writer identification," Pattern Analysis and Applications, vol. 19, no. 3, pp. 699–708, 2016.

[46] R. Jain and D. Doermann, "Offline writer identification using k-adjacent segments," in 2011 International Conference on Document Analysis and Recognition, 2011, pp. 769–773.

[47] O. Kırlı and M. B. Gulmezog˘lu, "Automatic writer identification from text line images," International Journal on Document Analysis and Recognition (IJDAR), vol. 15, no. 2, pp. 85–99, 2012.

[48] S. Fiel and R. Sablatnig, "Writer retrieval and writer identification using local features," in 2012 10th IAPR International Workshop on Document Analysis Systems, 2012, pp. 145–149.

[49] G. Ghiasi and R. Safabakhsh, "Offline text-independent writer identification using codebook and efficient code extraction methods," Image and Vision Computing, vol. 31, no. 5, pp. 379–391, 2013.

ARTIFICIAL INTELLIGENCE IN DEFENCE-A PERUSAL

Mrs. Harsha Patil

ABSTRACT

Modern advancements in Artificial Intelligence (AI) have come about in a step forward for numerous classical AI- applications, such as computer vision, normal dialect preparing, mechanical autonomy, and information mining. So numerous Try to utilize these improvements for military applications, such as perceptions, recon- naissance, risk assessment, submerged mine fighting, cyber security, insights examination, command and control, and instruction and preparing. In any case, indeed with the conceivable outcomes for AI in military applications, there are numerous incitements to consider. For case in point, 1) high risks means that military AI-systems need to be pellucid to attain decision maker trust and to facilitate risk analysis; this is a challenge since many AI- techniques are black boxes that lack adequate transparency, 2) military AI-systems need to be vigorous and authentic; this is a challenge since it has been shown that AI-techniques may be vulnerable to unnoticeable manipulations of input data even without any knowledge about the AI-technique that is used, and 3) many AI-techniques are based on machine learning that requires large amounts of training data; this is challenge since there is often a lack of sufficient data in military applications. This paper display comes about from progressing ventures to personality conceivable outcomes for AI in military applications, as well as how to address these challenges. The reason of this paper is to highlight conceivable outcomes and major challenges for AI in military applications. With brief presentation to DL, many cases of military AI-applications, key challenges related with AI within the military space, as well as methods that can be utilized to mostly address these challenges. there are numerous incitements to consider.

Keywords-Artificial Intelligence, Robotics, Deep learning, Defence.

1 INTRODUCTION

1.1 Artificial Intelligence (AI)

Artificial intelligence (AI), particularly the subfields machine learning (ML) and profound learning (DL), has inside a decade moved from prototyping at investigate establishing and colleges to industry and real-world application. Present day AI utilizing DL-techniques has revolutionized the execution of conventional AI-applications such as machine interpretation, QA-systems, and discourse acknowledgment. The numerous progressions in this field has moreover turned other bright thoughts into surprising AI-applications able of picture captioning , lip perusing , voice impersonation , video blend , ceaseless control , etc. These comes about recommend that a machine able of programming itself has the potential to: 1) move forward proficiency with regard the advancement costs of both computer program and equipment, 2) perform particular errands at a superhuman level, 3) give inventive arrangements to issues not already considered by people, and 4) give objective and reasonable choices where people are known for being subjective, one-sided, unfair, corrupt,etc.

In a military setting, the potential for AI is show in all spaces (i.e. land, ocean, air, space and data) and all levels of fighting (i.e. political, vital, operational and strategic). For occurrence, at

the political and deliberately levels, AI can be utilized to destabilize an adversary by creating and distributing enormous amounts of fake data. In this case, AI will most likely moreover be the leading candidate to protect against such assaults. At the strategic level, AI can make strides incompletely independent control in unmanned frameworks so that human administrators can work unmanned frameworks more productively to, eventually, increment war zone affect.

1.2 Robotics (Mechanical autonomy) and AI

Robotics (Mechanical autonomy) and AI could be a transformative innovation that has huge application within the social, economic, and military areas. Reports of assignment drive on AI in guard incorporate proposal s incorporate arrangement and organization mediations that are required to control and empower a vigorous AI based innovations for guard segment.

Robotics (Mechanical autonomy) and AI is combination of innovation and cognitive insights for silulation, processing of data and information to construct capability in a machine to hint human behavior. It could be a transformative innovation that has colossal application within the social, financial and military areas. Till a few time back we had robots worked by a human or through a set of programming to perform dreary assignment. The innovation is as of now show in numerous divisions of economy, such as mapping innovations, hand composing acknowledgment for mail conveyance, monetary exchanging, reconnaissance, target procurement, shrewd vehicles, ammo, and robots performing various mechanical, restorative and military errands.

India's financial and national security can be reshaped through AI innovation with its administration in data innovation space. The nation needs to embrace exacting arrangements to drive AI appropriation and advancement, in segments which rise above past buyer products and IT services.Moreover, the productivity is backed by the multiplication in cloud innovation, as enormous computing control is made accessible on request through cloud based computing stages, at much diminished costs.AI can too encourage the capacity and investigation of datasets, possibly handling billions of information focuses in a matter of seconds. A few layers of computers work in couple to analyze data in these expansive information sets, recognizing designs, observing behavior, and making cleverly choices. Machine insights fueled stages can gotten to be a vital instrument of administration in India over a wide extend of open administrations.

1.3 Deep Learning

By DL we allude to machine learning models comprising of different of layers of nonlinear preparing units. Regularly, these models are spoken to by counterfeit neural systems. In this setting, a neuron alludes to a single computation unit where the yield may be a weighted whole of inputs that passed a (nonlinear) enactment work (e.g., a work that passes the flag as it were on the off chance that it is positive). DNNs allude to frameworks with a expansive number of serially associated layers of parallel-connected neurons. The differentiate to a DNN could be a shallow neural arrange that has as it were one layer of parallel-connected neurons.

Until almost ten years a long time prior, preparing of DNNs was for all intents and purposes outlandish. The primary effective preparing techniques for profound systems were based on preparing one layer at a time. The parameters of the layer- by-layer-trained profound systems were at last fine-tuned (at the same time) utilizing stochastic angle strategies to maximize the

classification precision. Since at that point, numerous investigate propels have made it conceivable to specifically prepare DNNs without having a layer-by-layer preparing. For illustration, it has been found that initialization procedures for the weights of the organize in combination with enactment work determination are pivotal. Indeed methods such as arbitrarily debilitating neurons amid the preparing stage, and normalizing the signals some time recently they reach the actuation capacities have appeared to be of incredible significance in accomplishing great comes about with DNNs.

Representation learning is one of fundamental reasons for the tall execution of DNNs. Utilizing DL and DNNs it is now not fundamental to physically craft the highlights required to memorize a particular assignment. Instep, segregating highlights are naturally learned amid the preparing of a DNN.

1.4 Defence Technology

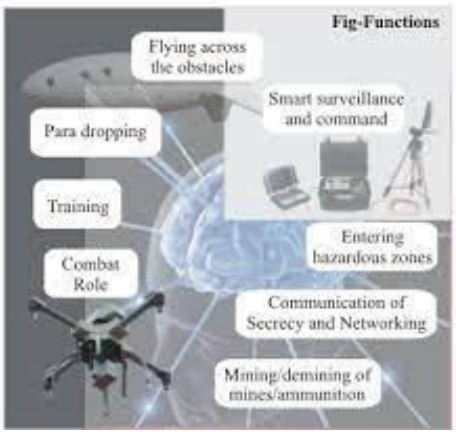
When innovation is connected or brought into utilize for different military purposes or for battling wars is said to be as protection innovation. Such innovation incorporates certain sorts which are completely military or protective by character. These innovations are unsafe to be taken care of on the off chance that somebody doesn't have appropriate preparing. Protection innovation is for the most part implied for military purposes but numerous a time it has been connected for civilian utilize. Advance, advances planning for civilians have been brought into utilize for guard purposes. Knowing what is Protection Innovation isn't a little subject. It can cover a wide run of basic things. The innovation to be utilized for protection reasons is particularly investigated and created by proficient engineers and researchers. These innovations are being utilized by the protection strengths in combating adversaries. The present day guard innovation is made and affected in a enormous way by inventive thoughts. To create result-oriented imaginative thoughts, it is fundamental to have the information, which is imperative in such improvement.

2 ARTIFICIAL INTELLIGENCE IN DEFENCE TECHNOLOGY

Defence is the act which is attempted to guard or protect anybody or anything against a antagonistic ambush. Defence is the matter which garners tall inclination by any country to back its presence. Countries got to remain on protect always for any sort of dangers, dangers and assaults ejecting remotely and inside. This reason is noteworthy sufficient to permit a major portion of the nation's budget for Defence purposes. The strengths that protect a country are made available highly created arms and ammunitions. These powers continuously ought to stay at the apex level of carefulness, mindfulness and readiness. Most of the errands of Defence powers get streamlined with the utilize of innovation. Thoughts in this field which are inventive offer assistance in fast and successful improvement of the guard innovation.

The later drift to supplant the tradition gear by robots keen and shrewdly machines that learns by perception, trial and mistake to improve operational proficiency, are giving sost viable options. The Prime Minister Mr. Narendra Modi amid DefExpo 2018 specified modern and developing innovations like AI and Robotics (Mechanical autonomy) will maybe be the foremost critical determinants of protective and hostile capabilities for any protection constrain within the future.

Long term war frameworks are likely to be ruled by unmanned systems.AI can give different alternatives for military applications for a key, operational and strategic level arranging in numerous of the capacities such as brilliantly and independent unmanned framework, information examination, data preparing and insights investigation, preparing War-gaming, reenactment and preparing and defence,offense,and command data fighting etc.



Robotics and AI in Defence Sector

Artificial Intelligence and Robotics (Mechanical autonomy) might play noteworthy part in number of protection applications such as:

- Image translation for target distinguishing proof and classification.
- Expert frameworks for determination and support of advanced weapon frameworks such as radars and missiles.
- Robotic gear can be utilized to supply exactness focusing on bolster and carriage of ammo and accuracy.
- Camera equipped and shock-resistant stages to supply fire control remotely are moreover conceivable applications.
- Frameworks for determination and upkeep of modern weapon systems.
- Missile target extend and direction examination for assessment of murder zones, dispatch time and recreation to help in qualifying rocket execution in different environments.
- Enhanced utilize of robots for Against Extemporized Hazardous Gadget, extraction of work force, terminating of weapons and other applications.

Indigenous Development

The defence research and development organization (DRDO) have a research facility particularly devoted to AI known as center for Artificial Intelligence and Robotics (CAIR).The research facility with almost 150 qualified scientists. CAIR has been working on a extend to create a multi agent Robotic Framework (MARF) for more than eight months now. MARF will prepare India's outfitted powers with a cluster of robots that can capacities as a group, in a design comparable to what our troopers do. Indian armed force has as of now built Wheeled Robot with Inactive Suspension, Wind Robot, Legged Robot, Wall-Climbing Robot, and Robot Sentry, among others.

 ISBN: 978-93-93810-31-1

Robot Sentry

CAIR has too created Network Traffic Analysis (NETRA) which can screen web activity. It can analyze voice activity passing through computer program such as Skype, Google Conversation and caught messages with ket words attack, bomb, blast, kill and other words in genuine time. There's no question that seminal inquire about is being done at CAIR.These would be interpreted to reasonable applications for the Indian Outfitted Forces.

Omni Hexbot

Miniature UGV (MINIUGV)

Indian Armed force has put the arrange for 200 Remotely Operated Vehicle (ROV)Daksh robots to defies explosive.additionally,DAC has endorsed approx.544 Robots for Indian Armed force from inborn source. The robots will be utilized for reconnaissance and can provide reasonable ammo.

Private Sector

Robotics (Mechanical autonomy) and Artificial Intelligence in defence can be a modern potential division for the private players. Till presently, private segment have been centered basically in making Robots for shopper products advertise. Making items for the protection can moreover have a double utilize. There are a few outside companies who are in improvement of this innovation. A few particular cases being:-

- Uran-9, atrackedarmored vehicle controlled remotely by an administrator prepared with 2A72 30-millimeter cannon with a rate of fire of 350 to 400 rounds per miniature and can shoot tall unstable combustible and armor-piercing ammo.

Artificial Intelligence in Military Operations

This is often the proper time to urge prepared for a future where AI can be doled out a driving duty in military operations or fighting. The advance in Artificial Intelligence will bring around modern potential in protection innovation. By taking advantage of Artificial Intelligence in military operations the possibilities of battling a war can be progressed other than creating the execution military units. Numerous of the countries around the world are consolidating AI to improve the execution of protection strengths. They are doing so by:-

- The utilize of Deep learning machines which needs Big Data and artificial neural networks,
- Making a profoundly created cooperation between human and machine where AI- powered machines help human to require exact and fitting fighting decisions,
- Using such devices of AI which would bring each administrator within the defence organize to carry out and wrap up any operation or task being given over.
- Bringing together unmanned as well as a kept an eye on framework for different sorts of combat of machines and human.

How AI is Changing the Face of Defence Technologies?

One of the famous consider centers had anticipated some of the time back that AI would alter the highlights of resistance innovation as did by air ship, atomic weapons and computers. The expertise and mastery of researchers and engineers to create and make advances and applications will straightforwardly reflect How AI is changing the Confront of resistance innovations. And to do so they require emerges for a reliable amount of information. Such information is collected from the regular utilize of different guard vehicles, ships and airplane. Information for AI can too be made through physical protection workouts or preparing, war recreations and computerized reenactment.

What is worth distinguishing is how the defence staff conduct themselves within the challenging scenarios? The framework of AI requires to be instructed with each accessible information, comprising of all imperative inputs and how and what conclusions were drawn. After collecting information from a war field, it can be put together into an AI framework. This will truly progress the security of officers as well as other non-defence work force and keep a exact note of neighborly powers. AI is becoming a need to guarantee the security of end of the. It is turning out to be fundamental for a country to shape up the resistance innovations to preserve specialist over imminent foes.

3 MILITARY AI-APPLICATIONS

This section presents a few examples where AI can be applied to enhance military capability.

3.1 Surveillance

Oceanic observation is performed utilizing settled radar stations, watch aircrafts, ships, and in later a long time electronic following for sea vessels utilizing the automatic identification system (AIS). These data sources give voluminous sums of data around vessel development which will uncover unlawful, hazardous, debilitating, and atypical behavior. In any case, the expansive sums of data almost vessel development make it troublesome to physically identify such behavior. Instep ML-approaches are utilized to create typicality models from vessel development information. Any vessel development that goes astray from the typicality models is considered atypical and displayed to administrators for manual assessment.

An early approach to oceanic irregularity location utilizes the Fluffy ARTMAP neural network architecture to show ordinary vessel speed based on harbor area. Another approach utilizes acquainted learning of movement designs to foresee vessel development based on its current area and course of travel. Others utilize unsupervised clustering based on Gaussian mixture models (GMM) and kernel density estimation (KDE). The models empower discovery of vessels that alter course, cross ocean paths, move within the inverse heading or travel at tall speed. Later approaches utilize Bayesian systems to identify untrue transport sort, as well as irregular, inconceivable, and dallying vessel development. Future improvements of oceanic peculiarity discovery ought to too consider encompassing vessels and interaction among numerous vessels.

3.2 Underwater Mine Warfare

Submerged mines posture a noteworthy danger to marine vessels and are utilized to channel development or deny pas- sage through confined waters. Mine countermeasures (MCM) in this manner try to find and neutralize mines to empower flexibility of development. Mine looks are progressively performed with an autonomous underwater vehicle (AUV) that's prepared with synthetic aperture sonar (SAS), which gives centimeter-resolution acoustic symbolism of the seafloor. Since AUVs collect huge sums of SAS symbolism, programmed target classification is valuable to segregate potential mines from other objects. Whereas programmed target classification of mines has been considered for a long time, the tall execution of DNNs for picture classification has made an intrigued in how such approaches may be valuable for programmed mine discovery.

Some considers appear the potential of DNN for mine discovery. For case, depicts how sham mine shapes, mine-like targets, man-made objects and rocks were put on the seafloor on

different geo- realistic areas. An AUV was at that point utilized to overview the seafloor with an SAS. The comes about appear that the DNN has altogether higher execution with higher likelihood of discovery of mine shapes and lower wrong caution rates compared to a conventional target classifier. So also, depicts how to create manufactured SAS pictures of cylinder-shaped objects and different seafloor scenes that were utilized to prepare the DNN. Assist considers may examine how to segregate mines from all sorts of clutter objects, combine discovery and classification, as well as strength to clamor, obscure, and impediment.

3.3 Cyber Security

Intrusion detection is an critical portion of cyber security to distinguish noxious organize action some time recently it compromises data accessibility, keenness, or secrecy. Intrusion detection is performed utilizing an intrusion detection system (IDS) that classifies the arrange activity as ordinary or meddling. In any case, since ordinary arrange activity frequently have comparable signature as genuine assaults, cyber security examiners analyze the circumstance for all interruption alarms to decide whether there's an genuine assault. Whereas signature-based IDSs are frequently great at recognizing known assault designs, they cannot identify already concealed assaults. Advance, improvement of signature-based discovery is frequently moderate and costly since it requires critical ability. This hampers the frameworks versatility to quickly advancing cyber dangers.

Numerous considers utilize ML and other AI-techniques to extend the classification exactness of known assaults, identify bizarre organize activity (since this may show modern assault designs that veer off from ordinary net- work activity), and computerize demonstrate development . In any case, few of these frameworks are utilized operationally. The reason for typically that interruption location presents particular challenges such as need of preparing information, huge changeability in organize activity, tall taken a toll of mistakes, and trouble of performing important assessments. In spite of the fact that expansive volumes of arrange activity can be collected, the data is frequently delicate and can as it were somewhat be anonym zed. Utilizing reenacted information is another elective, but it is frequently not adequately practical. The information must at that point be labeled for administered learning in terms of whether the designs are typical or an interruption, or for inconsistency location guaranteed to attack-free, which is regularly troublesome to do. At long last, the models have to be straightforward so that analysts can get it the location limits and centrality of highlights.

Another degree to extend cyber security is entrance testing amid security reviews for distinguishing proof of possibly exploitable security shortcomings. Infiltration testing is regularly robotized due to the complexity and expansive number of has in numerous systems. A few considers have explored how AI-techniques may be utilized for reenacted entrance testing utilizing coherent models of the arrange instead of the real organize. The arrange is regularly spoken to with assault charts or trees that portray how an enemy can abuse vulnerabilities to break into a framework. In any case, depicts how models contrast in terms of the way they characterize: 1) vulnerability for the aggressor from theoretical victory and discovery probabilities to vulnerability of arrange state, and 2) aggressor activities from known pre- and post-conditions to common detecting and perception of results. Encourage, with formal models of systems and has, it is conceivable to perform what-if examination of divers moderation procedures. Future investigate on infiltration testing will likely utilize cognitively substantial

models of the interaction between assailant and guard, e.g., as well as profound fortification learning to investigate the huge issue space of conceivable assaults.

3.4 Ammunition and Arms

AI-enabled innovation is presently joined in new-age weaponry. Progressed rockets, for case, can assess and look at the target level for slaughter zones without the require for human cooperation.

3.5 Logistics

Logistics is one of the foremost basic factors in deciding whether or not a military mission succeeds. ML and geospatial analysis are coordinates with military calculated frameworks to diminish exertion, time, and error.

4. USE OF AI IN MILITARIES AROUND WORLD

A military artificial intelligence arms race could be a competition or arms race between two or more states to have their military powers prepared with the most excellent artificial intelligence (AI). Since the mid-2010s numerous examiners have famous the development of such a worldwide arms race, the AI arms race, between extraordinary powers for superior military AI, coinciding with and being driven by the expanding geopolitical and military pressures of what a few have called a War. The setting of the AI arms race is the AI Cold War account, in which pressures between the US and China leads to a cold war pursued within the region of AI technology.

Russia's military might is well-known around the world. The Foundations for Advanced Research Projects is the country's equivalent of DARPA within the United States. Beat pioneers have been watched advancing the improvement of AI-based innovation for the military. Agreeing to reports, the Russian military is considering utilizing AI, big data, and ML to conduct more successful data operations. Russia is as of now contributing broadly in artificial intelligence (AI) for the distinguishing proof, examination, and negation of deception in both the commercial and defence sectors.

India is additionally taking little measures to join AI-based innovation into its combat and reconnaissance ventures. In reality, the Ministry of Defence shaped a high-level Defence AI Board (DAIC) in 2019, entrusted with conveying key direction for AI selection within the protection division. The major objective of the DAIC is to direct government-industry collaboration for the arrangement of such breakthroughs.

5. BENEFITS OF AI IN MILITARY OPERATIONS

Artificial intelligence (AI) includes a developing number of applications in different areas, and military branches don't need to miss out on the openings it brings. Here are six benefits of counterfeit insights within the military that are changing the world of defense and national security.

1. Better Threat Identification

War zones are inalienably unsafe places, but the U.S. Armed force considers AI might decrease a few of the dangers. It's working on a framework called the Aided Threat Recognition from Mobile Cooperative and Autonomous Sensors (ATR-MCAS).

Despite its long title, the advancement contains a direct reason. The innovation will offer assistance officers ended up mindful of dangers speedier and get counsel approximately how to bargain with them. Individuals on the war zone can moreover utilize ATR-MCAS in a wandering or settled position for surveillance or surveillance.

The Army's approach includes putting savvy sensors on its discuss and ground vehicles. They collect information and send it to an AI operator that forms it and makes choices appropriately. For case, the instrument seems prescribe which dangers troopers ought to handle first.

2. Reduced Staff Requirements

Businesses like fabricating and agribusiness as often as possible utilize AI to cut down on the estimate of the specified labor drive. The U.S. Naval force trusts to do the same with an independent dispatch built by Leidos Holdings, a defense temporary worker based in Virginia.

The transport completed a circular trip of more than 4,000 miles for its to begin with proof-of-concept travel. In spite of the fact that 14 back staff individuals taken after behind it in another vessel, the inventive water vehicle did not have any laborers on board for the trip.

All the AI frameworks worked as planning, but a mechanical system constrained the team within the trailing pontoon to intercede. Once they did, the effort was before long back on track. The Naval force isn't prepared for far reaching arrangement of this high-tech transport, but it's energizing to think around the conceivable outcomes.

3. Improved Recruiting

National defense branches have utilized various enlistment strategies over the a long time, extending from tv commercials to work reasonable booths. One of the benefits of manufactured insights within the military is that the innovation may make it simpler to discover appropriate individuals who are prepared to serve.

For case, Air Force Special Operations Command has 30 a long time of compiled information and employments AI calculations to figure out which characteristics make an perfect candidate.

Depending on artificial intelligence like meaning that scouts may pay less consideration to physical capacities and look for out other profitable characteristics, like strength and authority.

The Armed force too propelled a female chatbot to match with a male form it presented prior. A source shows that around 35% of the armed force website's 2.5 million month to month guests are ladies. The trust is that a female chatbot seem make ladies feel more comfortable inquiring certain questions, such as in case pregnancy impacts enlistment.

4. More Preparedness

Successful preparing could be a gigantic portion of a effective military encounter.. The instruction troopers get influences their execution and impacts everybody who works with them on missions.

The U.S. Armed force and Discuss Drive completed pilot programs of an AI preparing device that gives the substance and measures advance. Using it allegedly empowers a 40% diminishment within the time required to complete the fabric.

Another one of the benefits of artificial intelligence within the military for this reason is that there's no chance of learners utilizing obsolete reading material. They get the educational

modules on tablets, and those contraptions immediately reflect any overhauls to the fabric made by the distributors.

5. Enhanced Cyber security

When companies get effectively focused on for cyber assaults, the representatives at those ventures frequently have to be closed down their computers and depend exclusively on pen-and-paper strategies until IT specialists address the problem. It's simple to imagine how disastrous a comparable assault may be on progressed military innovation.

The U.S Armed force as of late put out a call for investigate almost how manufactured intelligence-powered cyber security alternatives might defend soldiers' strategic and communications systems. It needs to explore how the innovation may ensure against independent, versatile dangers.

Military authorities recognize that AI has confinements. It may well be a whereas some time recently these developing cyber security instruments come into far reaching utilize. It's vital that individuals are at slightest getting to know the technology's potential, in spite of the fact that.

6. Smoother Transportation

Soldiers frequently got to drive their vehicles over new territory or direct them through regions with restricted or no street foundation. Investigate is underway to prepare independent vehicles with AI so that military staff can get travel-assistance.

For illustration, those cutting edge vehicles might appear troopers the areas of all acceptable courses or prompt them on the speediest way to a given goal. Driving through obscure regions requires individuals to remain on tall caution whereas working their vehicles. AI seems diminish some of the cognitive strain they feel.

This application of AI within the military is still within the early stages. Be that as it may, it's simple to see why such an approach might help officers confronting questionable travel circumstances.

6 CHALLENGES

As shown by, there are unsolved challenges that are critical to be mindful of earlier to creating and sending an AI-based application for military purposes. In this segment we are going examine, in our supposition, the foremost basic ones for military AI: 1) straightforwardness, 2) vulnerabilities, and 3) learning indeed within the nearness of restricted preparing information. Other vital, but less basic, challenges related to optimization, generalization, building plan, hyper-parameter tuning, and generation review arrangement are not advance examined in this work.

4.1 Transparency

Numerous applications require, in expansion to tall execution, tall straightforwardness, tall security, and client believe or understanding. Such prerequisites are commonplace in security basic frameworks, observation frameworks, independent operators, pharmaceutical, and other comparable applications. With the later breakthrough for AI, there's moreover an expanded inquire about intrigued in straightforwardness to bolster end-users in such applications.

4.1.1 Expectations on transparency

The desired straightforwardness of AI depends on the end-users needs. Lipto portrays how straightforwardness may concern five sorts of client require for:

1. Believe in circumstances where it is troublesome for clients to address framework proposals. In any case, it may be vague whether client believe is based on framework execution or strength, execution relative the client, or how comfortable the client is with framework proposals.

2. Knowledge into previously obscure causal connections that will be tried with other strategies.

3. Information of framework execution limits due to constrained show generalizability compared to the clients capacities.

4. A few extra information about framework suggestions.

5. Decency to dodge efficient predispositions that will result in unequal treatment for a few cases. For case, assessment of credit applications ought to not be based on individual properties, such as sex or ethnicity, in spite of the fact that such properties may recognize populace bunches on an in general measurable level.

There are in rule, two ways to create AI-systems straightforward. Firstly, a few sorts of models are seen as more interpretable than others, such as straight models, rule-based frameworks, or choice trees. Assessment of such models gives an understanding of their composition and computation. Lipton portrays how the interpretability depends on whether clients can anticipate framework suggestions, get it demonstrate parameters, and get it the preparing calculation. Besides, the framework may clarify its proposals. Such clarifications may be printed or visual. For illustration, by demonstrating what viewpoints of an picture that for the most part contributes to its classification. Mill operator gives an broad audit of clarifications in social sciences investigate and how this information may be utilized to plan clarifications for AI frameworks. Ordinarily, individuals clarify other agent's behavior in terms of their seen convictions, wants, and eagerly. For AI frameworks, convictions compare to the frameworks data almost the circumstance, wants compare to the frameworks objectives, and eagerly compare to middle of the road states. Encourage, clarifications may envelop anomaly of activities, inclinations to play down taken a toll or hazard, deviations from anticipated standards, rule of occasions, and controllability of activities. The major discoveries are that:

Clarifications are contrastive in reaction to specific counter-factual cases. Clarifications hence center on why the specific suggestion was given rather than a few other suggestion.

• Clarifications are chosen and center on one or two conceivable causes and not all causes for the proposal.

• Clarifications are a social discussion and interaction for exchange of information.

4.1.2 Examples of interpretable Models

Bayesian rule lists(BRL) is one case of interpretable models. BRL comprise of arrangement of in the event that (condition) at that point (ensuing) else (elective) explanations. Portrays how BRL can be created for a exceedingly precise and interpretable show to assess the hazard of stroke. The conditions discretize a high-dimensional multivariate include space that impact the

hazard of stroke and the resulting depicts the anticipated chance of stroke. The BRL has comparable execution as other ML-methods for anticipating the chance of stroke and is fair as interpretable as other existing scoring frameworks that are less precise.

Lexicon-based classifier is another illustration of interpretable models for content classification. Lexicon- based classifiers duplicates the recurrence of terms with the likelihood for terms happening in each lesson. The lesson with the most noteworthy score is chosen as the expectation. models dictionaries employing a gated repetitive arrange that together learns both terms and modifiers, such as qualifiers and conjunctions. The dictionaries where prepared on whether posts in gathering are for or against passing punishment and estimations towards commercial preparations. The dictionaries perform way better than other ML-methods and are at the same time interpretable.

4.1.3 Examples of feature visualization

In spite of the fact that DNNs offer tall execution in numerous applications, their sub-symbolic computations with per- haps millions of parameters makes it troublesome to get it precisely how input highlights contribute to framework proposals. Since DNNs tall execution is basic for numerous applications, there's a impressive interest in how to form them more interpretable. Numerous algorithms for translating DNNs change the DNN-processing into the first input space in arrange to imagine separating highlights. Ordinarily, two common approaches are utilized for include visualization, actuation maximation and DNN clarification.

Actuation maximation computes which inputs highlights that will maximally actuate conceivable framework proposals. For picture classification, this speaks to the perfect pictures that appear separating and recognizable highlights for each course. In any case, the pictures frequently see unnatural since the classes may utilize numerous perspectives of the same protest and the semantic data in pictures is frequently spread out. A few cases of strategies for actuation maximation are angle rising, way better regularization to extend generalizability, and synthesizing favored pictures.

DNN clarification clarifies framework suggestions by highlighting segregating input highlights. In picture classification, such visualizations may highlight ranges that give prove for or against a certain lesson or as it were appear locales that contain segregating highlights. One approach for calculating separating highlights is affectability investigation utilizing nearby angles or other degree of variety. In any case, one issue with affectability investigation is that it may show segregating highlights that are not show within the input. For illustration, in picture classification the affectability examination may show darkened parts of an protest instead of the unmistakable parts. Layer-wise significance engendering dodges this issue by considering both include nearness and demonstrate response.

4.1.4 Examples of Application Specific Explanations

In differentiate to classification, AI-planning is based on models of space elements. how clarifications for arranging may utilize space models to clarify why activities were performed or not, why a few activity cannot be performed, causal connections that empower future activities, and they require for replanning.

Since reasonableness is imperative for numerous AI-applications, Tan et al. portray how show refining can be utilized to identify predisposition in black-box models. Show refining rearranges

bigger more complex models without noteworthy misfortune of exactness. For straightforwardness, they utilize generalized added substance models based on shallow trees that demonstrate each parameter and the interaction between two parameters. They prepare a straightforward demonstrate on framework suggestions from the black-box show and one straightforward demonstrate on the real result.

Hypothesis testing of contrasts in proposals from the two models appears cases where the black-box demonstrates present a predisposition, which may at that point be analyzed by comparing the two straightforward models. The framework was assessed on recidivism hazard, loaning advance chance, and person chance for being included in a shooting occurrence. The comes about appear that one black-box demonstrate belittles recidivism chance for youthful hoodlums and Caucasians, whereas overestimating the hazard for Native and African Americans.

4.2 Vulnerabilities

In this area, we talk about two distinctive perspectives of vulnerabilities of DNNs:

1) Powerlessness for control of input and

2) Defenselessness for control of the show. We begin by looking at control of the input flag.

4.2.1 Adversarial Crafting of the Input

Given a DNN, it has been found that it is simple to alter input flag so that the classification framework comes up short totally. When the measurement of the input flag is expansive, which is ordinarily the case for e.g. pictures; it is regularly sufficient with an subtle little alteration of each component (i.e. pixel) within the input to trick the framework. With the same strategy utilized to prepare the DNN, regularly a stochastic gradient strategy, you'll be able effortlessly discover in which course, by looking at the sign of the slope, each component ought to be changed to permit the classifier to wrongly pick a target course or simply just misclassify. With as it were some lines of code, the finest picture acknowledgment frameworks are betrayed to accept that a picture of a vehicle instep appears a pooch. Figure 1 underneath appears the picture some time recently and after control and the probability of the classes some time recently and after control.

The over strategy expect having full get to to the DNN, i.e., a so-called white-box attack. It has been found that indeed so-called black-box attacks, where you simply have understanding into the system’s sort of input and yield, are conceivable. Within, the creators prepare a substitute organize utilizing information gotten from scanty examining of the black-box system they want to assault. Given the substitute organize you'll be able at that point utilize the white-box attack strategy specified over to create ill-disposed inputs. An elective to learning a substitute organize is displayed in , where instep a hereditary calculation is utilized to form assault vectors driving to misclassifications by the framework. The same authors indeed appear that it is regularly sufficient to alter a single pixel within the picture, in spite of the fact that frequently distinguishable, to attain a effective attack.

4.2.2 Exploiting Hidden Backdoors in Pre-Trained Dnns

When planning a DNN, but as it were having get to to a little sum of preparing information, it is common to utilize pre-trained models to attain great execution. The concept is called exchange learning and a common method is to require a demonstrate that's prepared on a huge sum of

information, supplant and customize the final layers within the organize to the specific issue, and after that fine-tune the parameters within the last stages (and some of the time indeed the complete framework) utilizing the accessible preparing information. There are as of now a expansive sum of pre-trained models accessible for download from the Web. A important address is at that point "How do we know that those who transferred the demonstrate have no terrible eagerly? This sort of defenselessness is considered in where the creators embed backdoors into a show for recognizing US activity signs. For case, a sticker is prepared on a halt sign to have a place to a lesson other than halt signs. They at that a framework, based the US activity sign arrange, for recognizing Swedish activity sign system) when using the backdoor (i.e., placing a sticker on the traffic sign).

4.2.3 Defense Methods

One way to decrease the defenselessness of the DNNs to control of the input flag is to expressly incorporate manipulated/adversarial illustrations within the preparing process of the demonstrate. That's, in expansion to the initial preparing information antagonistic cases are created and utilized within the preparing of the demonstrate.

Another strategy is to utilize a concept called defense refining. Briefly depicted, the strategy tries to diminish the necessity that the yield flag as it were point out the genuine lesson and constrain the other classes to have zero likelihood. Typically exhausted in two steps. The primary step may be a regular preparing of a DNN. Within the moment step, the yield (class probabilities) of the primary neuron network is used as a modern lesson names and a unused framework (with the same design) is prepared utilizing the unused (delicate) lesson names. This has been appeared to diminish powerlessness, since you are doing not fit the DNN as well tight against the preparing information, and protect a few sensible lesson interrelations.

Other defense strategies are for occurrence highlight pressing strategies such as e.g., cruel or middle sifting or nonlinear pixel representations such as one-hot or thermometer encodings. Unfortunately, not one or the other of the strategies depicted totally tackles the powerlessness issue, particularly not in the event that the aggressor has full understanding into the show and the defense method.

4.3 Data

Creating ML-based applications in a military setting is challenging since the information collection strategies in military organizations, preparing offices, stages, sensor systems, weapons, etc. were at first not planned for ML-purposes. As a result, in this space it is regularly troublesome to discover real-world, high-quality and adequately expansive datasets that can be utilized to memorize from and pick up knowledge into. In this segment we'll explore techniques that can be utilized to construct ML-applications indeed within the nearness of restricted preparing information.

4.3.1 Transfer learning

Exchange learning may be a procedure that's commonly utilized when datasets are little and when computational assets are constrained. The thought is to reuse the parameters of pre-trained models, regularly spoken to by DNNs, when creating unused models focusing on other, but comparative, errands. There are at slightest two approaches that can be utilized for exchange learning in DL-applications:

• Relearning the yield layer: Utilizing this approach, the final layer of the pre-trained demonstrates is supplanted with a unused yield layer that matches the anticipated yield of the modern errand. Amid preparing, as it were the weights of the modern yield layer are upgraded, all others are settled.

• Fine tuning the complete model: This approach is comparable to the primary but in this case the weights of the whole DNN may be upgraded. This approach regularly requires more preparing information. It has been appeared that exchange learning may too boost the generalization capabilities of a demonstrate. How- ever, the positive impacts of exchange learning tend to decrease as the distance between the source task and target task increases.

4.3.2 Generative Adversarial Networks

Generative ill-disposed systems (GANs), concocted by Great individual et, may be a generative demonstrate that can be utilized for semi-supervised learning where a little set of labeled information is combined with a bigger set of unlabeled information to improve the execution of a show. The essential GAN usage comprises of two DNNs speaking to a generator and a discriminator. The generator is prepared to create fake information and the discriminator is prepared to classify information as genuine or fake. When the two systems are at the same time prepared, advancements to one organize will moreover result in changes to the other arrange until; at last, balance has been come to. In semi-supervised learning, the most objective of the generator is to create unlabeled information that can be utilized to move forward the generally execution of the ultimate show. GANs have, in expansion to semi-supervised learning, too been utilized for:

• Recreation: Filling the crevices of somewhat blocked pictures or objects.

• Super-resolution: Changing over pictures from moo determination to tall determination.

• Image-to-image interpretation: Changing over pictures from winter to summer, night to day, etc. A military application of this method can be to change over night-vision pictures to sunshine pictures.

4.3.3 Modeling and Simulation

Modeling and reenactment has been utilized broadly by the military for preparing, choice bolster, thinks about, etc. As a result, there are parts of as of now approved models that have been created over long periods of time that may moreover possibly be utilized to produce manufactured information for ML-applications. As an illustration, a flight-simulator can be used to produce manufactured pictures of aircrafts put totally different natural settings. Labeling is in this case programmed since the air ship sort is known earlier to producing the manufactured picture. Be that as it may, not shockingly, utilizing manufactured images may result in destitute execution when applying the show to real-world pictures. One approach that's right now being investigated is to improve the manufactured picture utilizing GANs to create it photo-realistic. This approach was effectively connected in.

5 CONCLUSIONS

The later breakthrough of AI is steadily coming to a point where it can be utilized in military applications. The paper depicts a few conceivable outcomes for utilizing AI in reconnaissance, submerged mine fighting, and cyber security. Other potential applications are observation

utilizing incompletely independent vehicles and sensor frameworks, danger assessment in discuss defense frameworks with tall transient prerequisites, insights investigation of developing designs, command and control frameworks, and instruction and preparing. In any case, military applications of AI have to be consider challenges in terms of: • Straightforwardness to guarantee demonstrate execution that's steady with military necessities. • Vulnerabilities which will radically decrease framework execution. • Deficiently preparing information for ML.

Much progression has as of now been made by analysts centering on the straightforwardness, interpretability, and clarifies capacity issues of AI. Numerous of these progressions can likely too be utilized in military AI- applications. In any case, a more careful prerequisites investigation is required to get it how to utilize these inquire about comes about. Military necessities may be exceptionally diverse with respect to hazard, information quality, lawful requests, etc. and a few sorts of straightforwardness may not indeed be appropriate. Advance, more thinks about are required on how to utilize social science investigate to progress AI-explain capacity.

Future ponders ought to too incorporate how to utilize the wealthy set of visualization methods that are created within the visual analytics research range. Since there's right now no silver bullet for the powerlessness issue, it is critical to monitor this inquire about range and persistently hunt for promising arrangements. In any case, until such arrangements are accessible it is vital to play down outside get to to models and guard strategies. Adversaries may something else attempt to utilize the vulnerabilities to their advantage.

At last, exchange learning makes it conceivable to adjust pre-trained models to military applications where there's both restricted preparing information and computational assets. GAN is another promising method that empowers learning utilizing labeled and unlabeled information (semi-supervised learning). GAN can too be utilized in combination with recreation to progress the authenticity of artificially produced preparing information.

REFERENCES

[1] Dario Amodei, Sundaram Ananthanarayanan, Rishita Anubhai, Jingliang Bai, Eric Battenberg, and Carl Case et al. Deep speech 2 : End-to-end speech recognition in english and mandarin. In Maria Flo- rina Balcan and Kilian Q. Weinberger, editors, Proceedings of The 33rd International Conference on Machine Learning, volume 48 of Proceedings of Machine Learning Research, pages 173–182, New York, New York, USA, 20–22 Jun 2016. PMLR.

[2] Yannis M Assael, Brendan Shillingford, Shimon Whiteson, and Nando de Freitas. Lipnet: End-to-end sentence-level lipreading. GPU Technology Conference, 2017.

[3] Housam Khalifa Bashier Babiker and Randy Goebel. Using KL-divergence to focus deep visual ex- planation. arXiv preprint arXiv:1711.06431, 2017.

[4] Sebastian Bach, Alexander Binder, Gre´goire Montavon, Frederick Klauschen, Klaus-Robert Mu¨ller, and Wojciech Samek. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PloS one, 10(7):e0130140, 2015.

[5] Michael Backes, Jo¨rg Hoffmann, Robert Ku¨nnemann, Patrick Speicher, and Marcel Steinmetz. Simu- lated penetration testing and mitigation analysis. arXiv preprint arXiv:1705.05088, 2017.

[6] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. Greedy layer-wise training of deep networks. pages 153–160. International Conference on Neural Information Processing Systems, NIPS, 2006.

[7] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D. Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, Xin Zhang, Jake Zhao, and Karol Zieba. End to end learning for self-driving cars. CoRR, abs/1604.07316, 2016.

[8] Jacob Buckman, Aurko Roy, Colin Raffel, and Ian Goodfellow. Thermometer encoding: one hot way to resist adversarial examples. International Conference on Learning Representations, ICLR, 2018. https://openreview.net/pdf?id=S18Su–CW.

[9] Carlos A Catania and Carlos Garc´ıA Garino. Automatic network intrusion detection: Current tech- niques and open issues. Computers &Electrical Engineering, 38(5):1062–1072, 2012.

[10] Kyunghyun Cho, Bart van Merrie¨nboer, C¸ alar Gu¨lc¸ehre, Dzmitry Bahdanau, Fethi Bougares, Hol- ger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Nat- ural Language Processing (EMNLP), pages 1724–1734, Doha, Qatar, October 2014. Association for ComputationalLinguistics.

[11] J Clos, N Wiratunga, and S Massie. Towards explainable text classification by jointly learning lexicon and modifier terms. In IJCAI-17 Workshop on Explainable AI (XAI), pages 19–23, 2017.

[12] Killian Denos, Mathieu Ravaut, Antoine Fagette, and Hock-Siong Lim. Deep learning applied to underwater mine warfare. In OCEANS 2017-Aberdeen, pages 1–7. IEEE, 2017.

[13] Dumitru Erhan, Yoshua Bengio, Aaron Courville, and Pascal Vincent. Visualizing higher-layer fea- tures of a deep network. University of Montreal, 1341(3):1–13, 2009.

[14] John Fox, David Glasspool, Dan Grecu, Sanjay Modgil, Matthew South, and Vivek Patkar. Argumentation-based inference and decision making–a medical perspective. IEEE intelligent systems, 22(6), 2007.

[15] Maria Fox, Derek Long, and Daniele Magazzeni. Explainable planning. arXiv preprint arXiv:1709.10256, 2017.

[16] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. pages 249–256. International Conference on Artificial Intelligence and Statistics, AISTATS, 2010.

[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z.

Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, Advances in Neural Information Process- ing Systems 27, pages 2672–2680. Curran Associates, Inc., 2014.

[18] Ian Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversar- ial examples. pages 1–11. International Conference on Learning Representations, ICLR, 2015. https://arxiv.org/abs/1412.6572.

[19] Tianyu Gu, Brendan Dolan-Gavitt, and Siddharth Garg. Badnets: Identifying vulnerabilities in the machine learningmodel supply chain. https://arxiv.org/abs/1708.06733, 8 2017.

[20] David Gunning. Explainable artificial intelligence (XAI). Defense Advanced Research Projects Agency (DARPA), 2017.

[21] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. Neural Computing, (18):1527–1554, 2006.

[22] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. https://arxiv.org/abs/1207.0580, 7 2012.

[23] Jo¨rg Hoffmann. Simulated penetration testing: From “dijkstra” to “turing test++”. In ICAPS, pages 364–372, 2015.

[24] IJCAI. Workshop on explainable artificial intelligence (XAI). 2017.

[25] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. International Conference on Learning Representations, ICLR, 2015. https://arxiv.org/abs/1502.03167.

A COMPARATIVE ANALYSIS OF PERFORMANCE AND SECURITY PARAMETERS OF DIFFERENT FIREWALLS AND SECURITY SYSTEMS

Ms. Neha Yadav and Ms. Sarita
Research Scholar, Department of Computer Science, JNUB, Jhajjar.

INTRODUCTION

There is no widely accepted way to evaluate the performance of firewalls, despite the fact that their deployment is increasing at a rapid pace. The fundamental reason for this is because firewall systems vary greatly, making direct performance comparisons impossible. Concerns about firewalls' ability to withstand large loads are being raised as more companies put them in place on their networks.

Performance Testing Setup

Fig. 1 shows the testing environment configuration used to assess the performance of three of the best firewalls currently available.

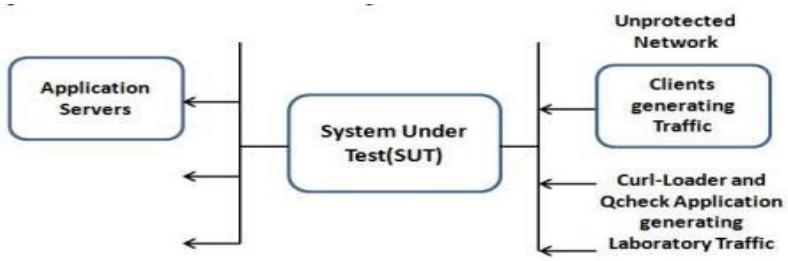


Fig 1: Setup Diagram for Performance testing

Open-Source software Curl-Loader is used to generate test traffic. Using the Curl-Loader programme, it is feasible to simulate hundreds of thousands of HTTP/HTTPS clients, each with a unique source IP address. In order to evaluate the performance of the System under Test (SUT) in a variety of environments, lab traffic is generated.

Table 1: SUT Configurations

Firewall Configurations	System Under Test – Firewall Products		
	Cisco ASA	**Checkpoint (CP) SPLAT**	**OpenBSD PF**
Platform	Cisco ASA - 5580	HP DL 380	HP DL 380
Operating System	ASA V 8.2.2 ASDM 6.2.5	SPLAT 2.4 Checkpoint NGX R65 HFA 50	Free BSD
Product Architecture	Multi-processor, Multi-core		
Processing Cores	8		
Gigabit Ethernet Interfaces	0		
10 Gigabit Ethernet Interfaces	4		

In this study, we are comparing the responses of several firewalls to various sorts of assaults. Discovery and penetration are two independent attacks. A team is formed to discover and verify the exact location of the target device. The penetration team examines the security mechanisms of each firewall.

Security Testing Setup

From the viewpoint of an outside invader, Fig. 2 shows the whole testing arrangement. As a result, our secure network opened itself up to the general public in order to construct a gateway. We set up an FTP server inside the network and opened it up to the outside world. The outside invader now had a lawful way to get the FTP server's IP address.

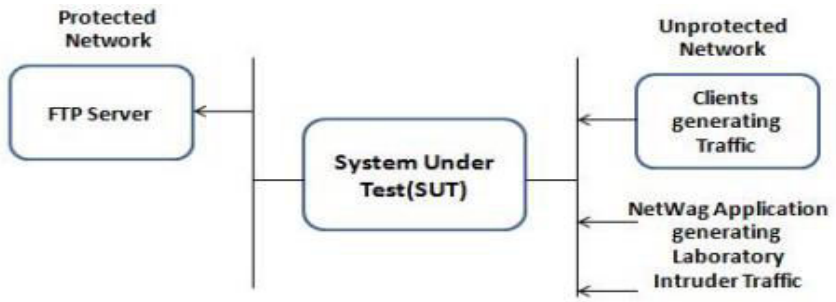


Fig 2: Setup Diagram for Security Testing

The aim of this setup is to simulate a condensed, real-world, corporate network layout. The access list of all three SUT firewalls permits WWW traffic to pass out on port 80 while at the same time allowing for FTP to pass in on port 21. For performing security testing, we used application called NetWag. Respective tests are performed by opening NetWag, selecting respective test and entering the target IP Address of our SUT. Firewall configurations taken are similarly used for performance testing as shown in TABLE II. The brief description of various tests conducted to verify ability of firewall to identify common attach types and block intruder are listed in Table IV

Table 2: Security testing Description

Test Group	Test Type	Purpose
Discovery	Network Sniffer	Discover target IP address and protocols used on target network.
	Trace route	Locate target device and all intermediate routers, switches and systems.
Penetration	Synflood	Check if firewall can overcome a repeated open connection request and also log attack.
	Garbage	Check if firewall can overcome random data packets on random ports.
	UDP Ping	Check if firewall can overcome a large UDP ping packet sent to it.
	TCP Ping	Check if firewall can overcome large TCP ping packet sent to it.
	Ping of Death	Check if firewall can overcome a single over-sized packet sent to it.

Performance Testing Results

Some of the major Key Performance Indicators (KPI) given in Table 3 are explored in order to compare performance of three of the most widely used firewall products in market as on today.

Table 3: Performance test results

Key Performance Indicators (KPI)	System Under Test – Firewall Products		
	Cisco ASA	CP SPLAT	OpenBSD PF
Firewall Licensing	Proprietary	Proprietary	BSD
Application Intelligence	Yes	Yes	No
Firewall Management	Local	Centralize	Local
HTTP Throughput (Gbps)	10.6	5.6	4.5
TCP Throughput (Gbps) (Object size = 512 KB)	18.6	14.2	10.2
Concurrent Connections	200K	250K	500K
UDP Throughput (Gbps) (Object size = 512 KB)	8	4	7
Connections per Second	160K	68K	180K

Table 3 is present the comparative analysis of CISCO ASA, Checkpoint SPI AT , Open BSD PF in case of HTTP throughput

Fig 4 is presenting the graphical view of comparative analysis of CISCO ASA, Checkpoint SPI AT, Open BSD PF in case of HTTP throughput

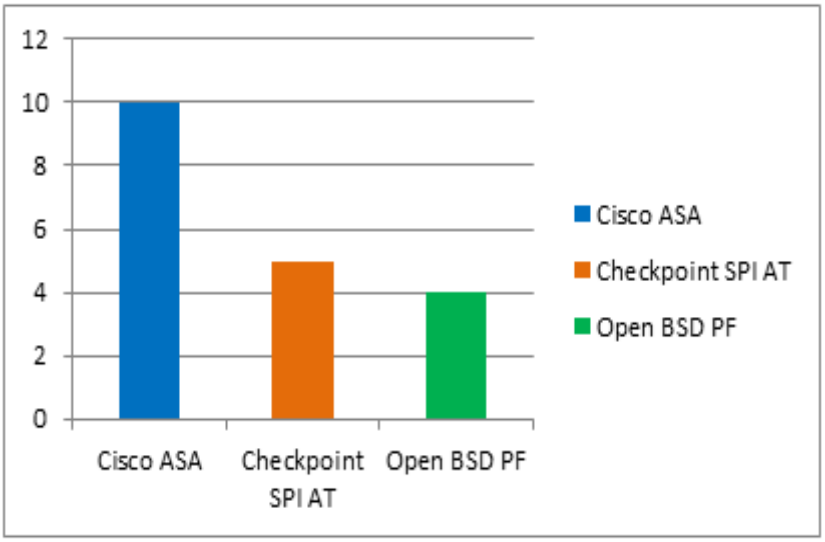


Fig 4: HTTP Throughput in Gbps

Table 5 is present the comparative analysis of CISCO ASA, Checkpoint SPI AT , Open BSD PF in case of HTTP throughput

Table 5. TCP throughput

Cisco ASA	Checkpoint SPI AT	Open BSD PF
19	15	10

Fig 5.is presenting the graphical view of comparative analysis of CISCO ASA, Checkpoint SPI AT , Open BSD PF in case of TCP throughput

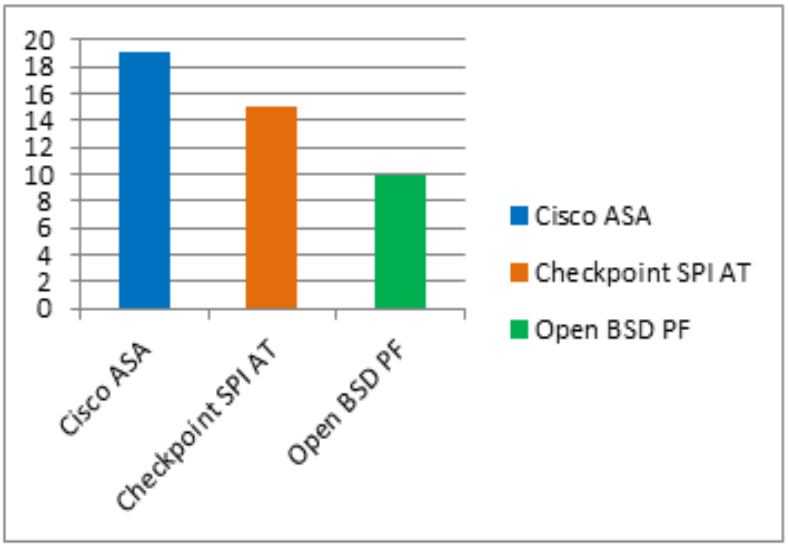


Fig 5. TCP Throughput in Gbps

Table 6 is present the comparative analysis of CISCO ASA, Checkpoint SPI AT , Open BSD PF in case of UDP throughput

Table 6: UDP throughput

Cisco ASA	**Checkpoint SPI AT**	**Open BSD PF**
8	4	7

Fig 7 is presenting the graphical view of comparative analysis of CISCO ASA, Checkpoint SPI AT , Open BSD PF in case of UDP throughput

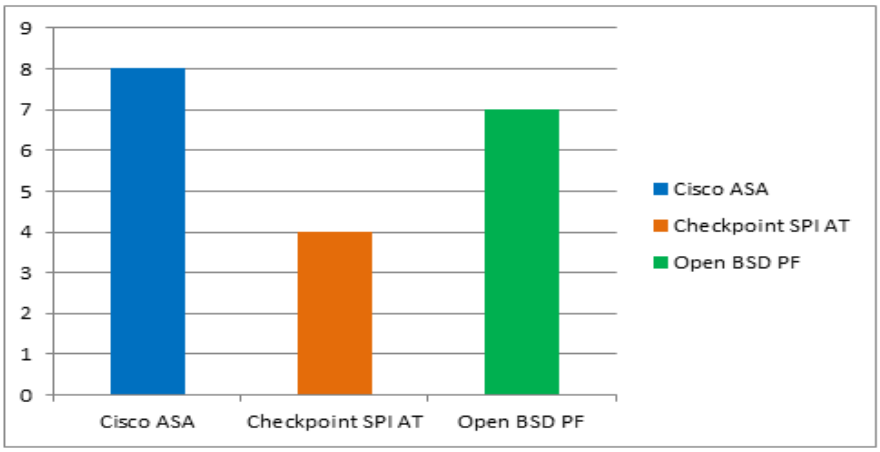


Fig 7: UDP Throughput in Gbps

Table 7 is present the comparative analysis of CISCO ASA, Checkpoint SPI AT , Open BSD PF throughput in case of connection per second

Table 7: Comparison for Connection per Second

Cisco ASA	Checkpoint SPI AT	Open BSD PF
160000	80000	180000

Fig 8 is present the comparative analysis of CISCO ASA, Checkpoint SPI AT , Open BSD PF throughput graphical view in case of connection per second

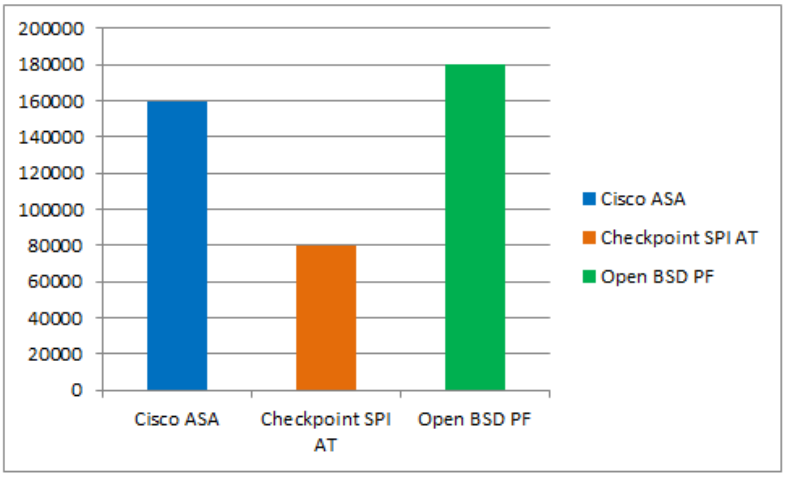


Fig 8: Comparison for Connection per Second

Table 8 is present the comparative analysis of CISCO ASA, Checkpoint SPI AT , Open BSD PF throughput in case of concurrent connections

Table 8: Comparison for concurrent connections

Cisco ASA	Checkpoint Spi At	Open BSD PF
200000	250000	440000

Fig 9 is present the comparative analysis of CISCO ASA, Checkpoint SPI AT , Open BSD PF throughput graphical view in case of concurrent connections

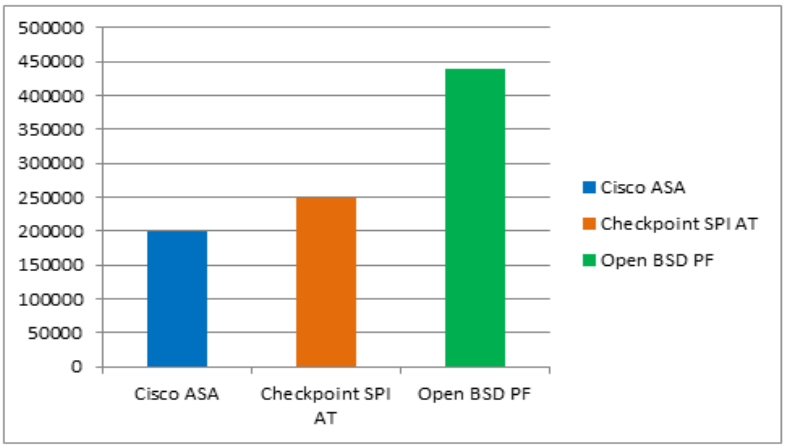


Fig 9: Comparison for concurrent connections

1) Firewall Licensing: Licensing for firewalls ensures that only authorized users may access the product. Cisco's licensing model is mostly per-connection, while Checkpoint's licensing model is primarily per-protected host. Open BSD PF has the most cost-effective BSD license of the two.

2) Firewall Management: Secondly, the criteria for comparing firewall administration in a distributed firewall system helps in the ease of management. Compared to Cisco, Checkpoint's graphical user interface (GUI) is more functional and intuitive. With Checkpoint's policy management, several firewalls with comparable access points may share the same rules. The only way to administer a Cisco or PF firewall is from inside the network itself.

3) Application Intelligence: This is the ability of a firewall to filter packets depending on the intelligence of the application layer. While both Cisco and Checkpoint use stateful inspection technologies, PF firewalls are restricted in their ability to analyze traffic at layer 7.

4) HTTP Throughput: In terms of bits per second (bps) or packets per second (bps), the HTTP throughput is the maximum load that can be handled while avoiding packet loss. Testing the SUT's performance when deployed to defend a high-performance Web-based application is the purpose of this test. Cisco firewalls beat the competition in real-world HTTP performance testing.

5) TCP Throughput: Bits per second or packets per second are the two most often used metrics for measuring TCP throughput. Firewalls in highly transactional contexts such as streaming media will benefit from this test. A TCP connection was opened, an object was sent over HTTP, and the connection was closed with a 512 KB object. Cisco once again surpasses the competition in terms of TCP throughput.

6) Concurrent Connection: SUT's total number of concurrent connections is referred to as the "Concurrent Connection" value. It shows the amount of simultaneous connections that can be managed by the firewall. For data centre firewall systems, the ability to install large-scale applications is directly impacted by this.

7) UDP Throughput: For firewall performance evaluation, connectionless protocols such as UDP may actually entail connections, at least in the context of UDP throughput. UDP packets may be sent past a firewall, for example. Host unreachable will appear if a target host isn't listening on the right UDP port for UDP packets. This is considered a connection for the purposes of measuring the performance of a firewall. The QCheck application program was used to assess UDP throughput performance using 512KB UDP packets. Cisco and PF are both doing well.

8) Connections per Second: It is the number of connections that are processed in a second. One of the most significant aspects of a firewall's performance is its ability to block incoming traffic. Defending against denial of service (DoS) attacks necessitates its use as well.

Deviation from Data Sheets

Cisco claims to be able to provide up to 10 Gbps of real-world HTTP firewall throughput and 1,50,000 connections per second in a testing environment. A similar conclusion may be drawn from Figure 4. The actual throughput we've been able to accomplish with real world traffic and roughly 160, 000 connections per second is 10.2 Gbps HTTP. When compared to Cisco's

performance data, it's a little better. Checkpoint, on the other hand, claims to be able to achieve HTTP throughput speeds of up to 12 Gbps. IBM systems are recommended by Checkpoint for maximum throughput, although the performance of throughput may vary depending on the hardware. HP Hardware Platform is capable of 5.6 Gbps speed in our testing configuration. PF firewall performance varies by hardware platform, system bus performance, network card efficiency, and kind of application, according to OpenBSD. With less RAM, the performance of the kernel-based programme PF is degraded and will not utilise swap space. Performances of 4.5 Gbps HTTP throughputs may be achieved using HP hardware in our lab setting. To the best of the author's knowledge, the product documentation did not provide the specific UDP and TCP throughputs. A variety of traffic patterns were generated in the testing setting in an effort to evaluate and contrast the performance. By comparison to its HTTP throughput, the PF firewall's UDP throughput and concurrent connections were excellent.

Security Testing Results

When the stateful packet inspection feature was enabled on all three firewalls, they were all resistant to the assault and stopped all communications on every test. As well as blocking incoming and outgoing packets, all of them allowed legitimate connections that weren't flagged as attacks. Because it's an open source firewall, OpenBSD PF includes everything you'd expect. Cisco ASA and Checkpoint SPLAT can't match its strength and potential, but it's a lot less expensive. Unlike Cisco and Checkpoint, it does not give any kind of visual outcome analysis of the recorded attacks, although it does preserve very thorough text log files for each assault. By keeping track of sessions and modulating the session, OpenBSD PF helps to prevent data connections from being hijacked. OpenBSD PF also conducts stateful or stateless packet inspection.

Network Access Control and Policy Management

Research for the thesis and associated projects is discussed in this part. The firewall and firewall policy are used to discuss network and host-based access control. Network hazards are taken into account while categorizing firewalls as well as their mitigating options. Access control rules are examined for ways to resolve disagreements. On the basis of a survey of existing firewall policy management practices, we look at how the work presented in this dissertation compares to a variety of other strategies. We take a look at some of the difficulties that firewall policy composition presents.

The Firewall and Firewall Policy

To protect networks and networked systems, firewalls serve as the first line of defense against intrusions and malware. The oldest definition of a firewall is: a set of components put between two networks that together have the following properties:

1. There must be a firewall for all traffic from within to outside and vice versa to flow through.
2. As established in the local security policy, only approved traffic will be permitted to pass.
3. No one can get over the firewall."

When Cheswick et al. define the term "firewall," they mean "any programme, device, or arrangement of hardware that restricts network access."

Security Policy

"A collection of rules and procedures that establish or control how a system or organisation offers security services to safeguard sensitive and important system resources" is defined in a high-level document like the Payment Card Industry Data Security Standard (PCI-DSS).

The PCI-DSS policy requirements are summarized in Table.

Table 5: PCI-DSS security policy excerpt

Build and Maintain a Secure Network	
Requirement ID	**Requirement**
req-1	Ensure that cardholder data is protected by configuring a firewall
req-1.1.2	Wireless networks are also included in the current network diagram for cardholder data.
req-1.2	Protect cardholder data by configuring firewalls and routers to prevent connections between untrusted networks and any system components.
req-1.4	Make that any mobile and/or employee-owned machines (for example, laptops used by workers) that are utilized to access the organization's network have personal firewall software installed on them.

Network Security Policy

Refined security policies are used to govern access control to the network. A network security policy 'describes an organizations network security issues and outlines the manner network security should be attained'

Table depicts an extract from a company's network security policy.

Table 5: Example network security policy excerpt

Policy ID	**Description**
nsp-1	Allow administrators to ping the code revision control system in the development subnet for liveness tests.
nsp-2	Allow developers FTP access to the production Web server.
nsp-3	Mitigate the threat of IP spoofing at the network perimeter, in accordance with RFC 5735.

Firewall Policy

A network security policy is put into action by a firewall policy or distributed firewall policy configuration. A firewall policy is often characterised as a set of rules that must be followed in a certain order. A rule is made up of filter criteria and a specific action. IP, TCP, and UDP packet header data and characteristics are often used as filter conditions.

The following table lists the most widely used filter condition characteristics at the different OSI and TCP/IP network model levels.

Table 9: Firewall filtering at different network layers

OSI Model Layer	TCP/IP Model Layer	Network Packet Attributes
Application 7	Application 5	Application protocol pattern matching
Presentation 6		
Session 5	Transport 4	TCP/UDP protocols, TCP Flags, Source and destination network ports
Transport 4		
Network 3	Internet 3	Source and destination IP addresses, ICMP (Type, Code)
Data Link 2	Network Access 2	Source MAC address
Physical 1	Physical 1	

A set of policy rules is applied to all packets as they pass through the firewall. To protect resources within a firewall, both incoming and outgoing filtering rules must be defined. Rule target actions are typically either allow, which allows the packet to get through the firewall, or deny, which blocks it. Upon matching the packet header against all of the filtering fields in a rule within the firewall policy, the packet is either permitted or refused based on the order in which the rules within the firewall policy are being matched. When there is no specific rule that matches the information in the packet header, a default one is used instead. The default rule is:

- Everything except what is specifically allowed is denied by default.
- Everything except what is specifically disallowed is allowed by default.

The ideal technique is to use the default deny rule. Logging relevant packets for auditing is regarded recommended practice as well, but a rule with a target action of log does not mandate the final decision to be made on a matching packet. In other words, if a matching rule at a later index does not allow or deny the packet, then the default rule is used.

CONCLUSION

Major operational firewalls on the market today have been evaluated in recent studies. We have outlined the primary regulatory compliance and the necessity for an effective, clean, and strong firewall design and administration in light of the rising relevance of network security in every enterprise. Firewall research, to the author's knowledge, has mostly been done in a theoretical fashion without any practical application. On the basis of real-world use, we've made an effort to assess the performance of different firewalls. Comparing several firewalls in this article can help you choose the correct supplier at procurement time to compare. Cisco ASA and Checkpoint SPLAT outperformed each other in our performance tests, and it was easy to see why. In terms of price, OpenBSD PF is the best open source option. Centralized policies and an improved user interface make Checkpoint's Firewall Management superior than Cisco's or Palo Alto Networks' offerings. Results of security testing showed that all three firewalls were capable of detecting basic intrusions and blocking transmissions from popular attack vectors. There are several factors to consider when selecting a firewall for your organization or network. Cisco ASA is a great option for big business networks because of its flexibility and reliability. Checkpoint SPLAT outperforms Cisco when it comes to high-security production networks. However, it is one of the finest low-cost open-source alternatives that gives comparable

performance but lacks in application-layer intelligibility, like private technologies. In addition, the author advises that the network environment be used as a benchmark for system performance. Not having the precise network environment replicated will not be useful in developing a firewall system. Setting up network circumstances as near to the real firewall is feasible, it is the best way to test the firewall.

FUTURE SCOPE

Firewalls give security against outside cyber attackers by isolating your computer or network from dangerous or unneeded network traffic. Firewalls may also prevent harmful malware from entering a computer or network over the internet. Information technology (IT) security is vital to an organization's capacity to do business and accomplish its goals. Security needs touch practically every company process and system, and good security measures assist safeguard a business' brand value, stakeholder trust, risk management methods, and compliance status. Firewalls are effective at enforcing company security regulations. They are used to limit access to certain services. The majority of firewalls may even allow selective access through authentication capability. Firewalls are solitary in function and do not need to be made between security and usability. They are excellent auditors. Firewalls are quite excellent at modifying suitable persons of defined occurrences. So, it embraces a vast diversity of purposes. There are various aspects to consider when picking a firewall for your company or network. Cisco ASA is a wonderful alternative for huge corporate networks because of its flexibility and stability. Checkpoint SPLAT surpasses Cisco when it comes to high-security production networks. However, it is one of the greatest low-cost open-source alternatives that delivers equal performance but lacks in application-layer intelligibility, like proprietary solutions. The degree of security may be improved by making further changes. The firewall of the future must be able to distinguish between legal and illicit traffic automatically in order to find and plug previously unknown dangers. Even while firewalls can check for anti-malware threats, the existing network performance effect has to be solved before this technology can be used.

REFERENCES

1. Wojciech Konikiewicz & Marcin Markowski (2017), Analysis of Performance and Efficiency of Hardware and Software Firewalls, Vol. 9, No. 1, pp. 49
2. Richa Sharma & Chandresh Parekh (2017), A Study and Its Classification, Volume 8, No. 4, May – June 2017
3. Xin Yue, Wei Chen, Yantao Wang (2009). The research of firewall technology in computer network security. DOI:10.1109/PACIIA.2009.5406566
4. Miss. Shwetambari G. Pundkar & Prof. Dr. G. R. Bamnote (2014), Analysis of firewall technology in computer network technology in computer network security, Vol.3 Issue.4, April- 2014, pg. 841-846
5. Steven Thomason (2012), Improving Network Security: Next Generation Firewalls and Advanced Packet Inspection Devices, Volume 12 Issue 13 Version 1.0 Year 2012
6. I. Kashefi, Maryam Kassiri, Ali Shahidinejad (2013). A Survey on Security Issues in Firewalls: A New Approach for Classifying Firewall Vulnerabilities.

7. Damodharan, Prabhat Kumar Srivastava (2018). A Review Paper on Computer Firewall. International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 5, Issue 2, February 2018

8. Dr. Ajit singh, Madhu Pahal, Neeraj Goyat (2013), A Review Paper On Firewall, Vol. 1 Issue II, September 2013

9. Bavithra.G.R, Mahalakshmi.V, R.Suganya (2018), A Review on Firewall and its Attacks, Vol. 7, Issue 1, January 2018

10. Raed Alsaqour, 1 Ahmed Motmi, 2*Maha Abdelhaq (2021). A Systematic Study of Network Firewall and Its Implementation. IJCSNS International Journal of Computer Science and Network Security, VOL.21 No.4, April 2021

11. Roumaissa Khelf; Nacira Ghoualmi-Zine(2018). IPsec/Firewall Security Policy Analysis: A Survey. DOI: 10.1109/SIVA.2018.8660973

12. Roza Dastres, Mohsen Soori (2021). A Review in Recent Development of Network Threats and Security Measures. International Journal of Computer and Information Engineering Vol: 15, No: 1, 2021.

13. Xinzhou He (2021). Research on Computer Network Security Based on Firewall Technology. doi:10.1088/1742-6596/1744/4/042037

14. M. G. Mihalos1,*, S. I. Nalmpantis2 and K. Ovaliadis2 (2019). Design and Implementation of Firewall Security Policies using Linux Iptables. Journal of Engineering Science and Technology Review 12 (1) (2019) 80 – 86.

15. Yuchong Lia, Qinghui Liu (2021). A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments. https://doi.org/10.1016/j.egyr.2021.08.126.

16. Raja Waseem Anwar 1,*ORCID,Tariq Abdullah 2 andFlavio Pastore (2021). Firewall Best Practices for Securing Smart Healthcare Environment: A Review. Appl. Sci. 2021, 11(19), 9183; https://doi.org/10.3390/app11199183

17. Khaled Salah, Khalid Elbadawi, Raouf Boutaba (2012), Performance Modelling and Analysis of Network Firewalls, Volume: 9 , Issue: 1 , March 2012

18. Mohammad Imran, Dr.AbdulrahmanA.Algamdi, Bilal Ahmad (2015), Role of firewall Technology in Network Security, Volume 4, Issue 12 December 2015

19. Er. Shikha Pandit, Er. Pritam Kumar, Er. Deepak Malik (2014), Fire-Router: A new secure inter-networking device, Vol. 3, Issue. 6, June 2014, pg.279 – 285

A REVIEW: APPLICATION EVOLUTION OF BIG DATA AND FEATURE LEARNING USING BIG DATA ANALYTICS

Sangram Keshari Swain and Kula Bhusan Pradhan
Centurion University of Technology and Management, Odisha, India

ABSTRACT

Now computing is totally based on generating and assimilation of data at an unusual pace. There are thousands of web applications on the internet serving millions of masses round the world. These web applications are increasing their content at an ever-increasing rate. We rely on the internet for any of our needs; be it searching for a meaning in Google or shopping to communicate, which has resulted in such vast explosion of data. This over the explosion of data is known as Big Data. Web applications not merely aim at providing services to the user, but also in a fast and efficient manner with the increasing amount of data, it has become hard to handle, store and analyze the information. Big data analysis has become very crucial part in the business point of view due to extraction of required information from huge data collection

Keywords: Application Evolution, Feature Learning, Big Data Analytics.

1. INTRODUCTION

The BDA history is intricately connected with that of data science. The term "big data" is initially applied by Michael Cox and David Ellsworth in 1997 in an IEEE conference paper to justify the data visualization and also to justify the challenges presented by it on computer systems. The rapid innovations and improvement obtained in IT technology by the end of the 1990s enables the origination of data in large volume. To highlight the importance of analysis, collection, interpretation, and integration of business information the business intelligence (BI) concept is created besides in what way this procedure set can aid businesses make certain determinations and acquire a clear knowledge about market activities and tendencies (Cox and Ellsworth)[1].

The evolutionary phase for big data's development was from 2001 to 2008 (RE Bryant) [2]. At first the big data was stated based on its 3Vs (volume, velocity, and variety), then it evicted to be more complex application to fulfil the necessities to deal with the information explosion. Hadoop, Database management systems and XML (Extensible Mark-up Language) Web services and are some of the Software and application developments that adds analytical modules and core function modules that concentrates on augmenting the end users usability , and allows users to handle vast quantity of data across and within the establishments collaboratively as well as in real-time.

Big data consumes huge data volumes which are deliberated as unstructured, structured and semi structured datasets that could not be captured easily, stored, handled, examined, accomplished also accessible through customary software, hardware and database management methodologies. The features of big data are usually used to define it. (L. Douglas) [3] proposed 3 measurements which describes the opportunities including the challenges of enlarging big data volumes. Obviously describing big data and its features possibly will be a continuing attempt, however still it will not have adverse impact on the execution and handling of big data.

Moreover, to describe big data the 3V's have been used continuously, the other dimensions of value in addition to veracity are included to explain the data feature and integrity to suit 5 V's of big data (P. Russom)

4. (Suthaharan S) [5] invented 3Cs: continuity, cardinality, and complexity arguing that the first 3 V's (volume,

velocity, and variety) will not favour the advance detection of the big data features for classification.

2. Concepts and characteristics of Big Data

About the creation of Big Data several descriptions are available which are based on distinctive viewpoints, for example the perspectives which are the social movement perspective, the process-oriented, product oriented and cognition-oriented. Presently, Web technology advances and the increased mobile and sensors devices bonded to the Internet have brought about the origination of massive data sets available on the Web which are required to be stored and processed. Now the companies so as to support the decision for their strategies are starting realizing the significance of using more data. The big data for the companies in the globally integrated economy are considered as a business priority.

A review on the Big Data has been presented by (SK Swain et al.)[6] that has determined to form value about all faces in establishments, workforce customers, or society as entire. Towards the NoSQL databases the new paradigm is moved corresponding in a considerable way besides accessible computing platforms, open-source software, and commodity servers' entrepreneurs, Researchers and businesses, likewise anticipated innovations or vehemently point to concrete that may be related to the evasion group of large data sets. At this scale the organizations endure to collect more data, the big data analysis reinforcing process will become dominant. The business executives to remain competitive must adopt the innovative techniques and developing technologies due to big data.

The perspective which is product-oriented highlights the qualities of data regarding with respect to their structures, speeds and sizes. The inspiration of this perspectives is depending on a historic vision, to relate the number of data offered with the volume recently. National Science Foundation (N.N.I. Initiative) [7] refers the Big data as: Huge, longitudinal, complex, diverse, or /and dispersed data sets created using sensors, tools. Today and in the future the available sources are online transactions, click streams, emails, videos and/many other digital sources. Moreover, the term big isn't only define the size. (MM Gobble) [8] expresses that data is big due to its excessiveness, as of its fast movement, or it is not organised suitably. The perspective which is process-oriented emphasises the originality of processes involved and required in storing, handling, accumulating, analysing and searching Big Data. So as to highlight the processing issues of Big Data, the mandatory technical structure, particularly the scientific gadgets, programming methods, statistical, computational and scientific innovations in this perspective is also emphasized.

Thus, (T Kraska) [9] described Big Data as: Once the usual solicitation of present methods doesn't facilitate the consumers to gain sensible, economical and valuable responses for data-driven queries. Likewise, (A Jacobs) [10] mentioned Big Data as: Data that's volume makes us to view

past sampled and exact approaches which are dominant at that period. The veracity as introduced by (B Feldman et al.) [11] is an alternative difficult feature. Few descriptions of Big Data are presently being on the basis of the behaviours of Big Data.

Conversely, (SB Scruggs et al.)[12] mentions that the Big Data's description is apart from the scope of these features besides must spread to its ability "to be convenient and re-claimed, add value over time, as well as originate a multi-dimensional, systems-level understanding". Data is developing at an exponential pace always on the internet. One of the difficult issues is to execute, save and examine this considerable volume of data in certain way that of least time-consuming. Big Data in industry and in skill is having more focus. The technology is shifting to data centric architecture and functional models. Basically, it's necessary to determine the architectural elements, operative models, also the crucial information models that formed Big Data Ecosystem. To explain all looks of the Big Data Ecosystem (SK Swain et al.)[13] centres on the Architecture Framework of Big Data. This has the modules such as Big Data Infrastructure, Big Data Architecture Framework, Big Data Definitions and Analytics, Big Data structures, Big Data Lifecycle Management, and models.

The perspective i.e. cognition-based, emphases on Big Data challenges regarding their limitations and the cognitive capacities. The range of commonly used software tools and hardware environments is exceeded by the data for capturing, managing, and handling it in an endurable passed period to its user populace. A formative document from McKinsey Global Organisation highlights the same matter (J. Manyika et al.)[14]: the size of data set is away from the capacity of classic database software devices for capturing, analysing, storing and managing. Hence, Big data is conceptualized by this perspective through the statistic of surpassing human capability for comprehending suitable outcomes to the mediation requirement for enhancing the ability of interpretation. At last, the perspective of social movement lures our attentiveness to the break among the vision and realism, specifically the socioeconomic traditional and ethical moves which underlie the Big Data's existence (H. Ekbia et al.)[15]. The vision of Big Data has been emphasized by this perspective for technical innovation, ecological and biomedical study, education and nation-wide safety (OSP, Obama administration unveils) [16] and states Big Data as: The quantifying capacity that modify many features of contemporary lift for transforming the art of management (A. Ignatius) [17], otherwise to be portion of a main transforms which needs general exertion.

For conceptualizing the Big Data some of the perspectives present valuable insights and in addition quite a lot of features can be reviewed. The common measures volume, variety and velocity are acknowledged and considered (D Laney et al.)[18]: data sets size is represented by Volume. The Big Data's volume may be of terabytes or even petabytes, which is much apart the typical bounds of gigabytes and megabytes. The in and out speed of data indicates the velocity. It represents the dynamic feature of the data, the frequency of the data creation and the requirement of creating real-time outputs. The range and source of data types are termed as Variety which emphasises several information sources and the distinctive data schematics of all sources. E.g., the traffic dataset contains on roads vehicular traffics numeric information and textual data about scheduled and active events (e.g., musical and sporting events) (Y.J. Xia et al.)[19].

Moreover, some of the additional Vs, like virtual, value, variability and veracity, have been stated in literature for serving as complement features of Big Data (H. Barwick) [20](C.W. Tsai) [21](P. Zikopoulos et al.)[22]: An organizations monetary worth is considered as 'Value' that can be obtain from Big Data processing. It contains two attributes: the large potential rate and very small density rate. Veracity relates to whatever range the data can be trusted because its sources reliability. For example, some devices may be more compromised when data's are received from sensors. Instead defining by numerous Vs, (X Wu et al.)[23] likewise emphasised the features of Big Data features using the professed HACE hypothesis: Big Data initiates with enormous - volume, dissimilar, independent sources with decentralized and distributed control, also aspire to deal with challenging and developing relations between data.

2.1. Big Data Definitions

Table 2.1 shows the definitions of big data which were given by NIST Standard.

Table 2.1 - Big Data Definitions

Volume	Measures the available quantity of data to association, Which certainly doesn't have to possess all of it only if it can retrieve it.
Velocity	Measures the streaming, aggregation and data formation speed.
Variety	Measures the intensity of representation data – text, images, video, audio, etc.
Veracity	Measures the ability of data understanding – noise, biases, anomaly and so on.
Value	Represents the data's utility in decisions making.

2.2. The paradigms of Big Data

A novel paradigm i.e., the fourth paradigm of scientific research was provided by the behaviors of big data. The historical visions for the four existing scientific paradigms are shown in TABLE 2.3. A data- intensive science was prompted by this new paradigm. The major problem of this data-intensive science was shifted from calculation to data and also from outcomes to thinking along with data. The curation, analysis, and capture were the three basic behaviors that were normally taken into consideration by (AJ Hey et al.)[24]. The big data was processed to attain the knowledge from data; this respective data was obtained to help the intelligent decision. At the same time, the data interpretation/visualization was identified as an essential task. Based on (Chen and Zhang) [25], the data interpretation/visualization was considered as a separate task not like the analysis process. The existence of this data was obtained in all shapes and scales of resources (also include individuals' lives). A huge range of procedures like integration, representation, and data cleaning were highly covered by the curation. It also includes the policies and essential metadata to perform the integration and longevity. The data analysis process includes some activities like analysis, databases utilization, and modelling throughout the workflow pipeline. A newly developing database for a mentioned stream should give solution to the 20 important queries that highly concerned by researchers is claimed by (AJ Hay et al.)[24].

Table 2.2 - Four Scientific Research Paradigms

Paradigm	Time	Methodology
Experimental science	1000 years ago	Natural phenomena depiction
Theoretical science	Last few 100 years	Newton's laws, Maxwell's expressions and so on.

Computational science	Last few decades	Simulation of complex phenomena
Data-intensive science	Today	Various sources are used to capture and generate data; data collaboration is supported by eScience, analysis, visualization, assessment

This table has been adopted from (AJ Hay et al.)[24]

When the big data was met by social science, a new discipline referred as computational social science (or e-social science) was developed. An enormous inter-discipline that was involved in this computational social science analyses and apply data along with the aid of big data technologies and techniques (R.M. Chang) [26]. For instance, in the computational organization science the perspective of researcher's were extended on organizational, social, and policy systems, then the computational models were leveraged, subsequently, it combines the science, network, science, social, and computer science. Likewise, in e-business, a large number of novel methods were developed for the research environments like consumer behaviour, managerial and human decision-making process, market interactions and operational processes. These methods were developed by taking the valuable data gathering costs in the social networks environment, digital entertainment and mobile telephony, blogs, and also consider the innovative abilities which found difficult to execute. By considering that, a paradigm shift was developed in scientific research, due to the enforcement of organizational environments and novel business practices. The data spectrum obtained in this new paradigm support the high value decisions by developing a novel advanced technique, and this data spectrum may extents the meso-level, macro-level to the micro-level.

2.3. Big Data Applications

The useful values were provided by this big data analysis procedure through suggestions, judgments, decisions, or supports. An enormous amount of applications were included in this data analysis process, and these applications were found to be regularly changing and extremely complex. An evolutionary process in data source is initially reviewed in this section. Subsequently, the six major data analysis fields like text analysis, network analysis, mobile analysis, structured data analysis, multimedia analysis and website analysis are then examined. Finally, several key application fields of big data were introduced.

3. Application Evolutions

In recent years, the big data analysis was introduced as an advanced analytical equipment, whereas it contains complex programs and large-scale under particular analytical process. During the past decades, the applications of data driven were developed. For example, the BI was developed as a prevailing technique as early as 1990s, for various business solicitations and also for network search engines in terms of huge data mining process that was developed in early 21st century. The influential and potential applications obtained from various sectors and its respective examination and data are also deliberated in below sections.

3.1. Evolution of Commercial Applications

The business data obtained in earlier days was generally identified as a structured data. The companies gathered this data from legacy system and store them in RDBMSs. An analytical techniques that are applied in this technique was developed in 1990s, whereas this technique was found to be more simple and intuitive e.g., in the forms of dashboard, reports, search-basis

business intelligence, data mining, interactive visualization, predictive modelling, queries with condition, online transaction processing, and score cards (RL Sallam et al.) [27]. From the 21st century, a unique opportunity was delivered by the WWW (World Wide Web) and networks for organizations to have direct interactions with the customers and also to have an online presentation. The website mining and text analysis process perform some process like customer trade analysis, market structure analysis, product layout optimization, and product suggestions.

3.2. Evolution of Network Applications

The WWW and email services were highly provided by the former generation of the Internet. Some of the process like web page analysis, Text analysis, and data mining were provided to perform the mining process on e-mail contends and also to develop a search engine. The applications that are developed in recent days are highly web-based, irrespective of their arena as well as design goals. The significant proportion of the global data volume was accounted by the network data. Web was identified as a mutual stand for interrelated pages, full of different types of data like videos, interactive contents, images, text, and audio, etc. So, at that particular moment, a huge amount of advanced technologies were applied for emerged data of unstructured or semi-structured. For example, the valuable information from the images was extracted by image analysis process, (e.g., face recognition). An automatic video surveillance system for law enforcement, military, and business applications were applied with the multimedia analysis technologies. Since 2004, various opportunities like upload, create, and share contacts were provided to each user by some online social media like online communities, social multimedia websites, social networking services, Internet forums, and blogs.

3.3. Evolution of Scientific Applications

Scientific research that is performed in various sectors like oceanology, environmental research, astrophysics, and genomics attains huge data with maximum throughput instruments as well as sensors. The BIGDATA program was recently announced by the U.S. NSF (National Science Foundation) to elevate the efforts to extract the insights and knowledge from huge collections of complex digital data. Platforms of Big data were developed and valuable results were obtained by few scientific research disciplines. For example, in biology physical computing resources, virtual machine resources, data service, network infrastructure, inter- operative analysis software, coordination environment, and inter-operative analysis software were applied by iPlant (SA Goff et al.)[28] to assist the students, researchers, and educators, in inspiring plant sciences. Huge varieties like experimental data, reference or specification data, observation data, model or analog data, and other derived data were included in this iPlant.

Over the previous couple of years, the big data has attained huge importance in various industries. It cannot be stored and process by former databases as it includes large and huge data sets. Big data process has the high capacity to handle such huge and large data sets because the large dataset was processed during the processing phase in a required time. Text analysis was still identified as infancy, but it was found to be a promising one. Because in various companies, 80% of data were obtained in unstructured form, where various analysis processes were work only for the structured data.

4. Feature Learning On Big Data

Here, some of the works based on the feature learning respect to big data which are related to our work is illustrated.

(Min Chen et al.)[29] presented a method for learning the features of medical image data. The researches of this article introduced a structure respect to deep learning for feature learning of medical image with unlabelled data. This framework efficiently learns features of medical images with the minimal amount of labelled data. The method proposed in this article is used for some task like classification, similarity check, and lung nodule recognition. From the experimental results of this method shown that their method is much fine than that of other approaches. The executions of auto-encoder approach were good, but it was affected by the quality of data. This issue is caused by ignoring the role of expert. On their future work, they planned to overcome this issue by integrating the domain knowledge and data-driven feature.

The task like image retrieval, image classification and object detection and segmentation are handled on big data with various methods. (Jamil Ahmad et al.)[30] presented an article for classifying the features of multimedia big data. Here the authors extract the features and transform it into hash codes by using the sensitivity property. This hash codes uses neighbour search procedures for the retrieval of multimedia data from big data. Two steps were used in this approach. Initially on first step, the introduced feature selection method chooses the features which they have maximum diversity. They found more than 1800 features from the 4000 features of images on their experiment. On the second step of the proposed system, with the selected feature the fast Fourier transform is estimated and higher frequencies will be binarizes with mean frequency. The method of this article aims to describe the selected multimedia feature as a signal and the feature vector is estimated with the FFT. The performance of this method earned higher performance than the other classic approaches. The researchers of this article telling that the performance of this method is weak only when using the smaller hash codes and additionally it is not suitable for sparse features, otherwise this method is better and executes good for deep features. They planned to make their method to support sparse features and improve the performance on their future work.

A new scheme of feature selection that will be optimal for the learning of big data is presented by (Adrian Barbu et al) [31]. The method used in this research was put on for improving different loss function, arrangement, ranking and application in regression. The researchers telling that their method is very simple and easy for implement. The over fitting problem solved by enforcing the second order prior on piecewise linear retort function. The experimental assessment of this paper was performed with some existing methods and they yield better performance on regression, classification and ranking and their method is highly scalable and efficient.

For the feature learning corresponding to big data in industrial IOT (Internet of Things) (Qingchen Zhang et al.)[32] presented a model of adaptive dropout deep computation with crowdsourcing. The main advantages of the model in this article was, preventing the over fitting of hidden layers, the parameter training and technique of crowdsourcing. These able to build an efficient distribution function and also it will monitor the available training samples. The results of this model exhibited that the offered model got finer execution than former crowdsourcing

algorithms. The framework introduced in this paper successfully avoided the over fitting and efficiently collecting the samples of parameter training for feature learning in IoT big data. On their future work, some best methods are planned to apply on their model for improving the performance.

(Dacheng Tao et al.)[33] presented an outline for improvement of image superiority in big data images. The method recommended for this article, takes five factors of image into account for the enhancement process. The authors of this article extract more than 15 features from the image and they estimated the visual quality of image with the regression module. These are analysed and learned from big data sets. These big data sets were huge in size than the similar image data sets. The researchers of this article conducted experimental setup of their model with nine datasets and performance are measured in three categories. The performance of their method is much better than FR and RR algorithms. But the IQA of the proposed method is slightly lower than FR and RR algorithms. On their feature work, they improve the IQA frame for increasing the performance of their method.

An intelligent fault diagnosis process was identified as a most promising tool in mechanical big data, as it has high ability to provide accurate results for diagnosis process and also contains the capacity to process the gathered signals rapidly and effectively. The existing diagnosis methods, manually extract the features based on diagnostic expertise and prior knowledge. An important merit of this traditional method was human ingenuity, but these methods were found to be labor-intensive and time-consuming. An artificial intelligence approach applied by the concept of unsupervised feature learning to study the topographies from raw data was highly inspired by (Y Lei et al.) [34], so a two-phase learning method was introduced in this method to perform the intelligent diagnosis process in machines. In the initial phase, sparse filtering and an unsupervised two- layer neural network was implemented to study the features directly from the signals which are obtained from mechanical vibration. According to this learned features, the health condition was classified in the second phase by applying the softmax regression. The performance evaluation of this process was performed by applying a locomotive bearing and motor bearing dataset respectively. This diagnosis method with motor bearing dataset shows high accuracy and superior performance in fault diagnosis than the other traditional approaches. Involvements of learning technique reduce the human labor and perform the fault diagnosis process more easily while handling the big data.

(Q Zhang et al.)[35] introduced a model of deep computation to study the features on big data in this the tensor was applied to make the complex connections of heterogeneous data. Here, the tensor destination was applied in the layer of output by way of the average sum-of-squares error term of the renovation error to completely learn the basic data distribution. A HBP (high-order back-propagation) algorithm was suggested in this method for training the parameters. The existing BP algorithm was extended from the vector space to the high-order tensor space for developing this HBP algorithm. Here, four various datasets were applied to perform the comparison process along with existing multimodal deep learning and stacking auto-encoder models. The experimental outcome of this feature learning methods was found to be more efficient when evaluated with SANE, STL-10, INEX, and CUAVE datasets.

The deep learning merits were applied for feature learning by (YBengio et al.)[36] whereas this deep learning was obtained as the layered architecture as same as the human brain. The simple topographies were extracted by deep learning with the help of raw data, likewise via multiple layers; the more complex features were learned (K. Hwang et al.)[37]. By performing the continuous optimization process in both backward and forward propagation, a considerable number of features were obtained at the final stage by multi-iteration learning. This feature learning process was categorized into two types they are unsupervised as well as supervised learning. The illustration data was transmitted from an input for prediction process to the layer of top via supervised learning. The connection parameters that were obtained between several pair of layers were optimized by applying BP. This optimization process was performed by reducing the cost function value among the predicted and target value. The most familiar neural network-based CNN (Y. Le Cun et al.)[38] were applied to represent the features through supervised learning.

CNN was frequently implemented in speech recognition (M. Chen et al.)[39], text analysis, and image analysis (M. Chen et al.)[39], etc. Particularly, in the image analysis field, CNN provide huge success, in face recognition (Y. Sun et al.)[40], scene parsing (C. t) [41], cell segmentation (H. Su) [42], neural circuit segmentation (D. Ciresan et al.)[43], analysis of images the breast (P. Fonseca) [44] (A. R. Jamieson) [45] and brain lesion segmentation (K. Kamnitsas et al.)[46] (A. Patel et al.) [47]. For instance, a novel 3D-CNN was introduced to classify the poly candidates on circulating tumour cell (CTC). In (T. Uemura et al.)[48] (W. Shen et al.)[49] (X. Li et al.)[50], and advanced convolutional networks were introduced to categorize the lung nodules from medical images by performing supervised feature learning process. (M Gao et al.) [51] and (T Schlegl et al.) [52], some of the CNN-based feature learning methods to categorize the lung tissues based on the CT images obtained from lungs.

The unlabelled data present in unsupervised learning process was applied for feature learning process, whereas only a minimum amount of data were applied to fine-tuning the parameters, they are as RBM (restricted boltzmann machine) (G. E. Hinton) [53,] deep belief network (G. E. Hinton et al.)[54], auto encoders (Y. Bengio et al.)[55] and stacked auto encoders (P. Vincent et al.)[56]. An auto encoder process for pulmonary nodule classification and unsupervised feature learning were proposed by (D Kumar et al.)[57]. To analyse the breast images, a convolutional autoencoder process was presented by (Kalleberg et al.)[58], and similarly, a RBM-based design was developed by (Li et al.)[59] for lung tissue classification. The analysis process in lung CT was proposed by (Tulder et al.) [60] along with convolutional restricted Boltzmann machines.

A large number of algorithms for unsupervised feature selection was introduced to chosen the informative features from unlabelled data. The feature selection process was identified as a commonly applied approach in unsupervised feature learning. The features that are conserving the data likeness or having manifold structure designed from the entire feature space was selected (Zhao and Liu)[61], (Cai et al.)[62], however, it fails to indirectly include the discriminative information within the data, so it cannot be applied directly in this shapelet learning problem. Previously existing unsupervised feature learning algorithms adopt the features one by one, and also the significance of several features was estimated individually (He et al.)[63], (Zhang et al.) [64], but feature correlation removal was identified as a major drawback. The classical method performs the feature selection by concurrently manipulating both feature correlation and

discriminative information process. Maximum discriminative features were selected by UDFS (Unsupervised Discriminative Feature Selection) (Yang et al.) [65] for representing the data; furthermore, it also deliberated the manifold structure. The discriminative information were normally encoded with labels for feature selection, so, the prediction of good cluster indicators as pseudo labels was found to be essential to perform the unsupervised feature selection.

A large number of studies were performed on the intelligent fault diagnosis process of rotating machinery intending to promptly deal with the immense fault data and the accurate diagnosis outcomes were provided automatically. Frequently, ANN (artificial neural networks) based approaches were applied along with these studies, which for features extraction use signal processing approaches and for classifying faults, the features are given as input to ANNs. However these approaches successfully functioned in an intelligent fault diagnosis process of rotating machinery, but there are two defects exist. (1) The features are extracted manually according to the sufficient previous understanding regarding signal processing methods and diagnostic capability. Furthermore, concurring to a specific diagnosis issue these manual features are extracted and is undoubtedly inapt for other concerns, (2) The shallow based ANN frameworks were included in this process, as it restricts the ANNs ability of learning the difficult non-linear interactions in fault diagnosis concerns.

In artificial intelligence as an innovation, deep learning maintains the ability to deal with the above- mentioned shortages. Through deep learning, deep NNs (DNNs) employed the deep architectures, in preference to shallow ones, might be determined to collect the valuable information from the valued complex non-linear functions and the raw data and estimated complex non-linear functions. An innovative intelligent technique was delivered by (F Jia et al.)[66] based on DNNs to handle the deficits of the aforesaid intellectual diagnosis approaches. The datasets obtained from planetary gearboxes and rolling element bearings was applied by this method to authenticate its effectiveness. These datasets include immense measured signals concerning distinctive health statuses under several operating conditions. From the outcomes it's realized that the proposed technique even though adaptively mines accessible fault physiognomies from the obtained signals, likewise attains excellent diagnosis accuracy related with the prevailing approaches.

Inefficient and unreliable human analysis is replaced by the intelligent fault diagnosis techniques, intensifying the effectiveness of fault diagnosis. The capability of multilayer nonlinear mapping highly supports the deep learning process to improve its accuracy in an intelligent fault diagnosis process. A novel method named as Deep Convolutional NNs (DCNNs) was suggested by (W Zhang et al.)[67] along with WDCNN (Wide First-layer Kernels). As input this approach utilizes raw vibration signals (to generate more inputs data augmentation is used), and the wide kernels at the initial convolutional layers were applied for smothering high frequency noise and also for feature extraction. Small convolutional kernels in the previous layers were used to support multiple layer nonlinear mapping. The AdaBN model was updated to enhance the competence of domain modification.

For automatic pattern recognition, (A Kumar et al.) [68] offered a fault diagnosis according to agenda by enduring the challenges that were presented by data inequity problem obtained in big datasets by employing RHadoop programming atmosphere. The Fault diagnosis performed in

the production method is a real-world example where the class-imbalance topic is extremely suitable. The regular operating behaviour was presented by majority of information that obtained from system whereas, faulty operating performance is restricted. The upholding approaches were condition-based method don't function accurately on such datasets and therefore, it is challenging to form consistent patterns for the precise fault diagnosis. To deal with this issue, considering the proposed frameworks first part, numerous methods for tackling the data inequity difficult, and then collective radial basis kernel SVM and for solving the drawback of big data imbalance the SMOTE (Synthetic Minority Over-sampling Technique) classifier is employed. Also, with conventional data imbalance solver methods such as under-sampling and SMOTE the proposed approaches oversampling performance is compared. In this frameworks second phase, the classifier output of SVM is replaced instead of target value of dataset to become stable in nature. In the final phase, the logistic regression were trained to perform the automatic pattern recognition process, the modified dataset is used moreover for predicting the faults by means of a steel plate manufacturing dataset in the RHadoop programming situation.

To detect the fault in real-time and to identify the industrial plants, a two-stage algorithm has been offered by (B Costa et al.)[69]. The work is fully dependant on the assessment of specific characteristics utilizing a novel evolving classifier procedure as well as recursive density evaluation. The suggested detection phase is dependent on the data space density formation, and it isn't as good as probability density function, however is a suitable measure to detect anomaly/outliers. A Cauchy function expresses this density and can be determined in a recursive manner, which improve its computational as well as memory power capable and, so, relevant to on-line solicitations. A self-developing classifier system or auto class is functioned based on fuzzy rules and it is proposed in the detection or diagnosis phase. When first rule base exists, Auto Class can develop or enhance it on the basis of the newly arrived faulty state data. With the functioning time improvement of the sulphur hexafluoride (SF6) electrical devices, the distinct amount of release possibly will happen inside the device. Due to this the effectiveness of equipment's insulation will weaken and produce severe impairment on the device. So, this one is of viable impact for analysing fault and assessing state of SF6 electrical device. In the current ages, data acquisition monitoring regularity of SF6 electrical devices has been always enhanced and the gathering space has always been extended, thus enormous data can be collected in the substation database.

A two-level model for fault diagnosis was introduced by (Hongxia et al.)[70] to instantly handle enormous SF6 electrical device condition monitoring data. At initial, the monitoring data is pre-processed in the previous training of the fault diagnosis model, for different missing values the first and different data filling methods are implemented. Then, on the Hadoop platform the fault diagnosis algorithms are parallelized. At last, the time taken for the fault diagnosis of SF6 electrical equipment in stand-alone mode and cluster mode is related by simulation, and the benefits of cluster mode in processing massive data are verified.

5. CONCLUSION

Because of the numbers of methods are contributing to big data, day to day there is an increase in the popularity of big data is happened. For facing the various necessities of users, providing a perfect classification method for big data is essential. In many fields numerous issues are solved

by using various approaches of big data. As already explained in the literature survey more number of researches are performed in various fields of big data. Now the present research is given according to the research gaps. Sequentially, the purposes of the research have been attained. The present research and its outcomes will provide future direction to other researchers.

6. REFERENCES

[1] Cox, Michael, and David Ellsworth. "Application-controlled demand paging for out-of-core visualization." In Proceedings. Visualization'97 (Cat. No. 97CB36155), pp. 235-244. IEEE, 1997.

[2] Bryant, R.E., Katz, R.H., Lazowska, E.D., 2008. Big-data computing: creating revolutionary breakthroughs in commerce, science, and society computing. Computing Research Initiatives for the 21st Century. Computing Research Association (Available at http://www.cra.org/ccc/files/docs/init/Big_Data.pdf).

[3] L. Douglas."3D Data Management: Controlling Data Volume, Velocity and Variety" (PDF). Gartner. Retrieved 6 February 20.

[4] P. Russom. “Big Data Analytics.” TDWI Best Practices Report, Fourth Quarter 2011. TDWI Research. 2011.

[5] Suthaharan, S. (2014). Big Data Classification: Problems and Challenges in Network Intrusion Prediction with Machine Learning. Performance Evaluation Review, 41(4): 70-73.

[6] Manas Ranjan Senapati, Sangram Keshari Swain, Srinivas Prasad, A Review of Big Data and Its Current Research Directions, International Journal of Engineering, Science and Mathematics (IJESM) 8, no 6 (2017).

[7] N.N.I. Initiative, Core techniques and technologies for advancing big data science and engineering (BIGDATA) (2015).

[8] M.M. Gobble, Big Data: the next big thing in innovation, Res. Technol. Manag. (2013): 64–66.

[9] T. Kraska, Finding the needle in the big data systems haystack, IEEE Intern. Comput. (2013): 84–86.

[10] A. Jacobs, The pathologies of big data, Commun. ACM 52 (2009): 36–44.

[11] B. Feldman, E.M. Martin, T. Skotnes, Big data in healthcare–hype and hope, Dr. Bonnie 360 degree, Bus. Dev. Digit. Heal. 2012 (2013) 122–125 http://www.riss.kr/link?id=A99883549.

[12] S.B. Scruggs, Karol Watson, Andrew I. Su, Henning Hermjakob, J.R. Yates 3rd, M.L. Lindsey, P. Ping, Harnessing the heart of big data, Circ. Res. 116 (2015): 1115–1119.

[13] Manas Ranjan, Senapati Sangram Keshari Swain, Srinivas Prasad, A Review Paper On Architecture Of Big Data , Journal of Advanced Research in Big Data Management System (IJARBMS) 1, no (2017): 1-10.

[14] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A.H. Byers, Big Data: The Next Frontier For Innovation, Competition, and Productivity. Report, McKinsey Global Institute, 2012.

[15] H. Ekbia, M. Mattioli, I. Kouper, G. Arave, A. Ghazinejad, T. Bowman, V.R. Suri, A. Tsou, S. Weingart, C.R. Sugimoto, Big data, bigger dilemmas: a critical review, J. Assoc. Inf. Sci. Technol. 66 (2015): 1523–1545.

[16] OSP, Obama administration unveils "big data" initiative: Announces $200 million in new R&D investments (accessed December 2015).

[17] A. Ignatius, from the editor: big data for skeptics, Harv. Bus. Rev. 10 (2012): 12-12.

[18] D. Laney, 3D Data Management: Controlling Data Volume, Velocity and Variety, Research Note 6, META Group, 2001.

[19] Y.J. Xia, J.L. Chen, C.H. Wang, Formalizing computational intensity of big traffic data understanding and analysis for parallel computing, Neurocom- puting 169 (2015): 158–168.

[20] H. Barwick, The "four Vs" of Big Data. Implementing Information Infrastructure Symposium (accessed December 2015).

[21] C.W. Tsai, ''Big data analytics: A survey,'' J. Big Data, vol. 2, no. 1, pp. 1–32, 2015

[22] P. Zikopoulos, C. Eaton, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw-Hill Osborne Media, 2011.

[23] X. Wu, X. Zhu, G.-Q. Wu, W. Ding, Data mining with big data, IEEE Trans. Knowl. Data Eng. 26 (2014): 97–107.

[24] A.J. Hey, S. Tansley, K.M. Tolle, The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research, Redmond, WA, 2009.

[25] C.P. Chen, C.-Y. Zhang, Data-intensive applications, challenges, techniques and technologies: a survey on Big Data, Inf. Sci. 275 (2014): 314–347.

[26] R.M. Chang, R.J. Kauffman, Y. Kwon, Understanding the paradigm shift to computational social science in the presence of big data, Decis. Support Syst. 63 (2014): 67–80

[27] Sallam RL, Richardson J, Hagerty J, Hostmann B, Magic quadrant for business intelligence platforms. CT, Gartner Group, Stamford (2011).

[28] Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, Gessler D, Matasci N, Wang L, Hanlon M, Lenards A et al, The iplant collaborative: cyberinfrastructure for plant biology. Front Plant Sci 34(2) (2011): 1–16.

[29] Chen, Min, Xiaobo Shi, Yin Zhang, Di Wu, and Mohsen Guizani. "Deep features learning for medical image analysis with convolutional autoencoder neural network." IEEE Transactions on Big Data (2017).

[30] Ahmad, Jamil, Khan Muhammad, Jaime Lloret, and Sung Wook Baik. "Efficient Conversion of Deep Features to Compact Binary Codes using Fourier Decomposition for Multimedia Big Data." IEEE Transactions on Industrial Informatics (2018).

[31] Barbu, Adrian, Yiyuan She, Liangjing Ding, and Gary Gramajo. "Feature selection with annealing for computer vision and big data learning." IEEE transactions on pattern analysis and machine intelligence 39, no. 2 (2017): 272-286.

[32] Zhang, Qingchen, Laurence Tianruo Yang, Zhikui Chen, Peng Li, and Fanyu Bu. "An Adaptive Droupout Deep Computation Model for Industrial IoT Big Data Learning with Crowdsourcing to Cloud Computing." IEEE Transactions on Industrial Informatics (2018).

[33] Gu, Ke, Dacheng Tao, Jun-Fei Qiao, and Weisi Lin. "Learning a no-reference quality assessment model of enhanced images with big data." IEEE transactions on neural networks and learning systems 29, no. 4 (2018): 1301-1313.

[34] Lei, Yaguo, Feng Jia, Jing Lin, Saibo Xing, and Steven X. Ding. "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data." IEEE Transactions on Industrial Electronics 63, no. 5 (2016): 3137-3147.

[35] Zhang, Qingchen, Laurence T. Yang, and Zhikui Chen. "Deep computation model for unsupervised feature learning on big data." IEEE Transactions on Services Computing 9, no. 1 (2015): 161-171.

[36] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.

[37] K. Hwang, M. Chen, "Big Data Analytics for Cloud/IoT and Cognitive Computing," Wiley, U.K., ISBN: 9781119247029, 2017.

[38] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[39] M. Chen, P. Zhou, G. Fortino, "Emotion Communication System," IEEE Access, IEEE Access, DOI: 10.1109/ACCESS.2016.2641480, 2016.

[40] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1891–1898, 2014.

[41] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," IEEE transactions onpattern analysis and machine intelligence, vol. 35, no. 8, pp. 1915–1929, 2013.

[42] H. Su, Z. Yin, S. Huh, T. Kanade, and J. Zhu, "Interactive cell segmentation based on active and semi- supervised learning," IEEE transactions on medical imaging, vol. 35, no. 3, pp. 762– 777, 2016.

[43] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in Advances in neural information processing systems, pp. 2843–2851, 2012.

[44] P. Fonseca, J. Mendoza, J. Wainer, J. Ferrer, J. Pinto, J. Guerrero, and B. Castaneda, "Automatic breast density classification using a convolutional neural network architecture search procedure," in SPIE Medical Imaging. International Society for Optics and Photonics, vol. 9414, pp. 2801–2808, 2015.

[45] A. R. Jamieson, K. Drukker, and M. L. Giger, "Breast image feature learning with adaptive deconvolutional networks," in SPIE Medical Imaging. International Society for Optics and Photonics, vol. 8315, pp. 601–613, 2012.

[46] K. Kamnitsas, C. Ledig, V. Newcombe, S. Joanna, D. Andrew, K. David, R. Daniel, and G. Ben, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," in Medical Image Analysis, vol. 37, pp. 61-78, 2017.

[47] A. Patel, S. van de Leemput, M. Prokop, B. Ginneken, and R. Manniesing, "Automatic Cerebrospinal Fluid Segmentation in Non-Contrast CT Images Using a 3D Convolutional Network," SPIE Medical Imaging. International Society for Opticsand Photonics, vol. 10134, pp. 201–206, 2017.

[48] T. Uemura, H. Lu, H. Kim, R. Tachibana, T. Hironaka, J. Nappi, and H. Yoshida, "Classification of Polyp Candidates on CTC Based on 3D-CNN," in International Forum on Medical Imaging in Asia (IFMIA 2017), pp.103-105, 2017.

[49] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang, Y. Zang, and J. Tian, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," Pattern Recognition, vol. 61, pp. 663-673, 2016.

[50] X. Li, Y. Kao, W. Shen, X. Li, and G. Xie," Lung Nodule Malignancy Prediction Using Multi-task Convolutional Neural Network," SPIE Medical Imaging. International Society for Optics and Photonics, vol. 10134, pp. 241–247, 2017.

[51] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, and R. M. Summers et al., "Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, pp. 1–6, 2016.

[52] T. Schlegl, S. M. Waldstein, W.-D. Vogl, U. Schmidt-Erfurth, and G. Langs, "Predicting semantic descriptions from medical images with convolutional neural networks," in International Conference on Information Processing in Medical Imaging. Springer, pp. 437–448, 2017.

[53] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," Neural computation, vol. 14, no. 8, pp. 1771–1800, 2002.

[54] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, vol. 313, no. 5786, pp. 504–507, 2006.

[55] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle et al., "Greedy layer-wise training of deep networks," Advances inneural information processing systems, vol. 19, p. 153-160, 2007.

[56] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in ACM Proceedings of the 25th International conference on Machine learning, pp. 1096–1103, 2008.

[57] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in ct images," in 12th. IEEE Conference on Computer and Robot Vision (CRV), pp. 133–138, 2015.

[58] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, C. M. Vachon, K. Holland, R. R. Winkel, N. Karssemeijer et al., "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1322–1331, 2016.

[59] Q. Li, W. Cai, and D. D. Feng, "Lung image patch classification with automatic feature learning," in 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 6079–6082, 2013.

[60] G. Tulder, M. Bruijne M," Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted boltzmann machines," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1262–1272, 2016.

[61] Zhao and Liu, 2007] Zheng Zhao and Huan Liu. Spectral feature selection for supervised and unsupervised learning. In ICML, pp. 1151–1157, 2007.

[62] Cai et al., 2010, Deng Cai, Chiyuan Zhang, and Xiaofei He. Unsupervised feature selection for multi- cluster data. In KDD, pages 333–342, 2010

[63] He et al., 2005] Xiaofei He, Deng Cai, and Partha Niyogi. Laplacian score for feature selection. In NIPS, pages 507– 514, 2005.

[64] Zhang et al., 2015] Peng Zhang, Chuan Zhou, Peng Wang, Byron J Gao, Xingquan Zhu, and Li Guo. E- tree: An efficient indexing structure for ensemble models on data streams. IEEE Transactions on Knowledge and Data Engineering, 27(2):461–474, 2015.

[65] Yang et al., 2011] Yi Yang, Heng Tao Shen, Zhigang Ma, Zi Huang, and Xiaofang Zhou. l2, 1-norm regularized discriminative feature selection for unsupervised learning. In IJCAI, pages 1589–1594, 2011.

[66] Jia, Feng, Yaguo Lei, Jing Lin, Xin Zhou, and Na Lu. "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data." Mechanical Systems and Signal Processing 72 (2016): 303-315.

[67] Zhang, Wei, Gaoliang Peng, Chuanhao Li, Yuanhang Chen, and Zhujun Zhang. "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals." Sensors 17, no. 2 (2017): 425.

[68] Kumar, Ajay, Ravi Shankar, Alok Choudhary, and Lakshman S. Thakur. "A big data MapReduce framework for fault diagnosis in cloud-based manufacturing." International Journal of Production Research 54, no. 23 (2016): 7060-7073.

[69] Costa, Bruno Sielly Jales, Plamen Parvanov Angelov, and Luiz Affonso Guedes. "Fully unsupervised fault detection and identification based on recursive density estimation and self-evolving cloud-based classifier." Neurocomputing 150 (2015): 289-303.

[70] Miao, Hongxia, Heng Zhang, Minghua Chen, Bensheng Qi, and Jiyong Li. "Two-Level Fault Diagnosis of SF6 Electrical Equipment Based on Big Data Analysis." Big Data and Cognitive Computing 3, no. 1 (2019): 4.

AN EXTENSIVE REVIEW: IMPLEMENTATION OF DEEP LEARNING TECHNIQUE FOR FEATURE LEARNING FROM BIG DATA

Sangram Keshari Swain and Kula Bhusan Pradhan
Centurion University of Technology and Management, Odisha, India

ABSTRACT

The big data are created and gathered together from several fields and on IoT systems, by using the object abstraction layer the gathered data sets are transmitted from the object layer to service management layer. The upper layers will get some services like decision support and prediction from the service management layer. It provides such services by analysing the data received from the object layer. For clients, the upper layers will offer interface, it provides an intelligent service for big data. The bygone learning method was not able to manage huge amount of heterogeneous information available on distributed domains. It needs high speed and storage devices to handle these maximum numbers of varied information. Usually, the Big Data Analysis (BDA) denotes a method of information gathering, move in the centralized cloud data centers, pre-processing, assessment, and visualization.

Keywords: Big Data, Feature Learning, Deep Learning.

1. INTRODUCTION

The huge size data is represented as Big Data. The collection of huge size data that are growing exponentially with respect to time is referred as Big Data. In short, such information is so gigantic and complex so this information cannot be effectively stored or processed by the conventional data administration apparatuses. Hadoop was planned particularly for the analysis of huge data indexes to construct versatile, appropriated applications. To manage sizably voluminous data, Hadoop actualizes the worldview called MapReduce characterized by Google as per which the applications are isolated into minute pieces of programming. Every one of which can be kept running on an unmistakable hub of each one of the individuals who make up the framework. The large volume of data gathered with various distributed devices refers to Big Data. Based on the specific types of datasets, the big data which contain unstructured data are classified. The big data include three features they are, variety, velocity and volume. Here, various data gathering sources are considered as varieties, the speed of analysing and processing is termed as velocity likewise the volume indicates the storage size of the big data.

The classical deep learning methods like deep belief networks, DCNNs, and stacked auto encoders learn the features of some file types like image, text, and audio. It was understood as a single type feature learning (Vamsi Krishna et al.)[1]. For feature learning, various multi-modal feature learning methods are introduced recently and some popular multi-modal methods are multi-modal deep neural networks and deep Boltzmann machines. Initially, these multi-modal learning methods learn the features of every assorted information example and learned characteristics are integrated as a solitary vector to provide the combined illustration of heterogeneous data model. With the help of

unlabelled data, the multi-modal methods will learn the representation of single modal (M Dunja) [2].

2. DEEP LEARNING MODELS FOR BIG DATA FEATURE LEARNING

The following four characteristics are normally applied to define big data they are, velocity, variety, volume, and veracity. In big data, for performing feature learning process, the deep learning approach was reviewed from four features, which include, deep learning for low-quality data, heterogeneous data, huge quantity of data, and also for real-time data.

2.1. Deep learning for large volume of data

Initially, in deep learning process the data in large volume is identified as a major challenging issue. Frequently, a large number of samples with wide range of attributes are included in big dataset. Till now, a large number of deep learning approaches were developed to learn features and also to represent this huge data. The deep learning model contains some hidden layers, each with enormous neurons, leading to millions of parameters. The training process performed in deep learning model is found to be difficult. Recently, a large number of algorithms are developed to train these models, and these developed techniques were grouped under three categories they are implementation based on GPU, optimized deep learning, and parallel deep learning methods. The parallel deep learning method normally referred as deep stacking network suggested by (L. Deng et al.)[3] is identified as the most representative model. Some modules are included in this deep stacking network.

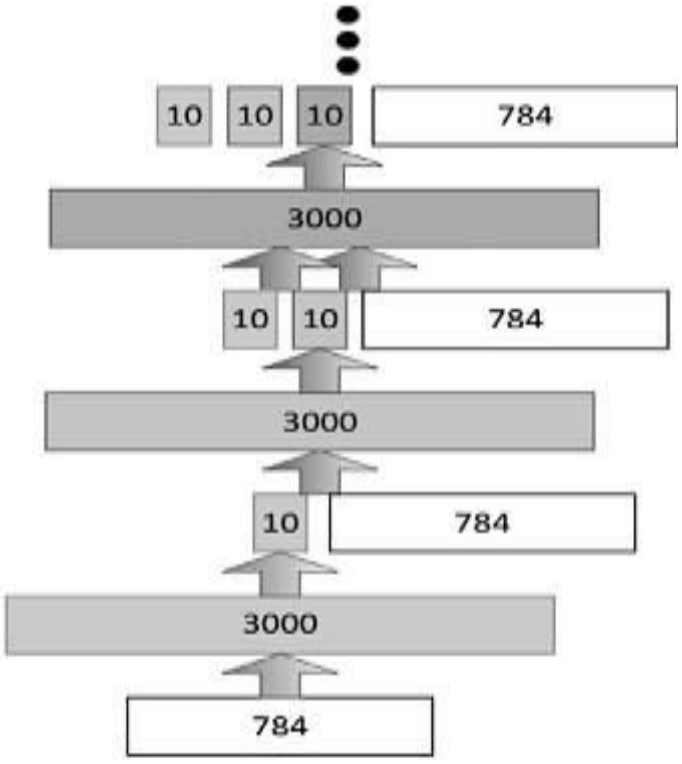


Fig. 4.1 Deep stacking network with three modules (L. Deng et al.)[3]

With three modules, a deep stacking network is given as an example in Fig. 4.1. The modules obtained in this stacking network are considered as a neural network with two sets of weight and also with a hidden layer as exhibited in Fig. 4.1. The three layers of lowest module are obtained in an ascending order. The lowermost layer is linear in which the original data is employed as input. The second layer is non-linear which contains hidden neurons and so called as hidden layer. The topmost layer is also identified as linear layer; it contains C output neurons to represent the classification target. Like all other deep learning methods, for mapping the input layer with hidden layer along with a bias vector and weight matrix, a sigmoid function is applied in the deep stacking network. The concatenated vector is defined by joining the previous output layer(s) with the original data. Above the lowest module, in each module the concatenated vector is utilized as an input. Consider an example, if an n-dimensional vector is applied to represent original data object and it contains 'D' dimension, class type 'C' of the input vector of the i^{th} counting from the lower to upper layer. Then the dimension is evaluated as $D= n + c \times (m- 1)$.

This deep stacking network model is paralleled so this method can be effectively applied to perform training process. The training efficiency is furthermore enhanced by introducing a tensor deep stacking network moreover on the CPU clusters. DistBelief a recent framework of software which was introduced in various machines in parallel order to perform the training process in large-scale deep learning approaches. DistBelief includes combination of both model parallelism and data parallelism to perform the training process with huge amount of data and free parameters. The model parallelism is achieved by partitioning the deep learning process into some sub-blocks, whereas this each sub-block is provided with computer to perform the training process.

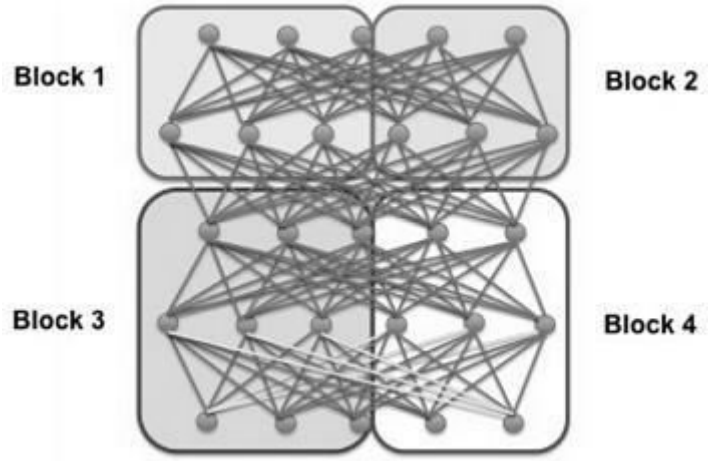


Fig. 4.2 Example of DistBelief with four blocks (J. Dean et al.)[5]

A solitary sample for this DistBelief with four blocks is presented in Fig. 4.2 (X. Chen)[4]. To perform the training process in deep learning approach, DistBelief have to transfer the data between the computers. This data transfer improves the communication particularly, for fully connected networks like deep belief network and stacked auto-encoder. In (J. Dean et al.)[5], the deep model was partitioned into 144 blocks, for enhancing the efficiency of training of DistBelief. Two optimization strategies like Sandblaster and Downpour are implemented by this

DistBelief to achieve the data parallelism process. The sandblaster is widely applied for batch optimization, whereas the Downpour is for online optimization. A high speed is achieved by this DistBelief process to perform the training process in the deep learning approach which is extensive. Consider an example; it achieves 12 times speedup in CNN along with 16 million images and 1.7 billion parameters on 81 machines instead of applying single machine. Besides, a significant improvement was achieved in training efficiency for an alternative deep learning framework having 14 million number of images, everyone with 16 CPU cores, 200 × 200 pixels size on 1000 machines (X. Chen and X. Lin) [6]. Therefore, this DistBelief is identified as a significant process to learn the features in deep learning as it has the capacity to scale-up enormous computers. This is identified as an important advantage of DistBelief.

The training efficiency of both DistBelief and Deep stacking network is improved by employing the multiple CPU cores for the deep learning methods which are extensive. Multiple CPU cores are applied to scale up the deep belief networks and some details regarding SSE2 instructions and data layout implementation were discussed in (V. Vanhoucke) [7]. Recently, some GPUs (graphic processors units) based large-scale deep learning architectures were explored. GPUs are perfectly appropriate for the parallel computing process of extensive deep learning methods, because they are equipped with huge memory bandwidth and maximum computing power. The advanced merits of this GPU based large-scale deep learning frameworks is demonstrated by few experiments. The best suitable example is, a deep learning architecture was developed by (R Raina) [8] to accomplish parallel training process based on GPUs in both sparse coding and deep belief networks by applying millions of training objects and more than 100 million parameters. The efficiency of parallelizing process performed in linear model was improved by applying some specialized schemes in the learning framework developed by (R Raina) [8]. For example, some parameters and training objects are included within the global memory for minimizing the data transfer.

Moreover, the sampling parameters p (x |h) and p (h | x) are produced to implement the parallel Gibbs sapling of both visible and hidden neurons. The speed of DBN developed by multiple restricted Boltzmann machines along with this architecture each having 1 million training objects and 45 million free parameters is increased by a factor of 70□.

2.2. Deep learning for heterogeneous data

In big data, the variety is identified as an individual characteristic, which inferred that big data is gathered from enormous sources in different layouts like data of unstructured and structured, also in the semi- structured form. Mostly the multi-model objects are found in the big datasets. For instance, both the text and image are contained instantaneously in the webpage. The multimedia objects like video clips containing text, audio, and still images are also identified as an example for multi-model approach. In multi-modal objects, diverse characteristics are included in each modality, which produce complexity in heterogeneous data. So, this heterogeneous data is identified as another challenge in the deep learning approach. Some multi-model deep learning approaches were developed to learn the representation of heterogeneous data. For instance, a multi-modal deep learning approach was advanced by (J Ngiam et al.)[9] to perform the feature learning process in audio-video objects. The framework for this multimodal deep learning model is depicted in Fig. 4.3. The representation and feature of both video and audio are learned separately by (J Ngiam et al.)[9] applying the restricted Boltzmann

machines. After that, for several multi-modal objects the learned features are then fused to form joint representation. Finally, the recognition or classification process of deep learning process or logical regression layer is performed by providing this obtained joint representation as an input.

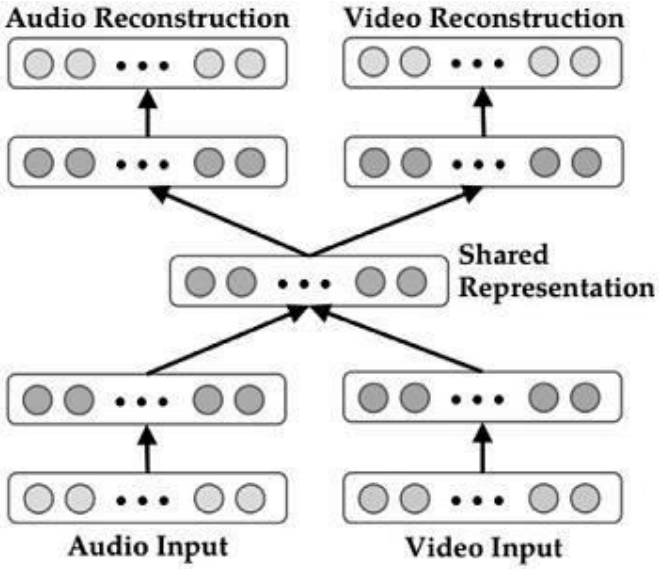


Fig. 4.3 Architecture of the multi-model deep learning model (J Ngiam et al.)[9]

Another one model for multi-model deep learning approach commonly referred as bi-modal deep Boltzmann machine was introduced by the authors Srivastava and Salakhutdinov, for performing the feature learning process in the objects of both the text and image as depicted in Fig. 4.4 (J. Xue et al.)[10]. Two deep Boltzmann machines are developed in this approach to execute the feature learning process for each modality of image as well as text respectively. In the same way, the joint representation is developed by concatenating the learned features of both image and text into a vector. The SVM classifier is trained along with the joint input representation to achieve the classification process.

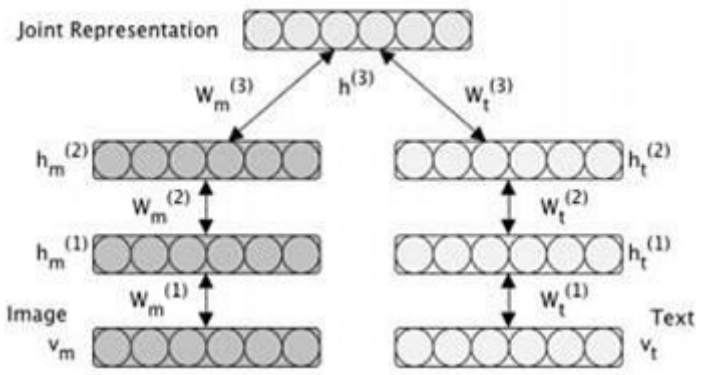


Fig. 4.4 Architecture of the bi-modal deep Boltzmann machine (W Ouyang et al.)[11]

(W Ouyang et al.)[11] delivered a multi-modal deep learning approach commonly referred as multi- source deep learning technique to estimate the human pose. Apart from the above said multi-model method, the primary objective of this multisource deep learning method is to learn the non-linear representation to estimate the human pose from various data sources like human

body clothing and articulation. In this technique, the features are separately extracted by applying the several data sources as input for the deep learning approaches having two hidden layers. To obtain the joint representation the extracted features are fused. The heterogeneous DNN along with conditional random fields are then included in the multi-modal deep learning approaches to recognise the Chinese dialogue act and also the multimodal DNN is provided with sparse group lasso to perform the heterogeneous feature selection process (L. Zhao)[12] and so on (A. Wang et al.)[13] (N. Neverova et al.) [14] (S. Rastegar et al.) [15]. Even though they have diverse architectures, but their concepts are similar. Initially, the features are learned by multi-modal deep learning technique to accomplish the single modality, after that, for several multi-modal object the learned features are then fused to form joint representation. Finally, the recognition or classification process of deep learning process or logical regression layer is performed by providing this obtained joint representation as an input.

2.3. Deep learning for real-time data

The most important characteristics obtained in big data are high velocity, so in real time, it is identified as an essential characteristic to perform the analysis big data. In big data, the rapid speed of data gathering is identified as an important challenging issue in real-time processing. Unfortunately, a high computational complexity is obtained in various deep learning techniques particularly in large-scale DNN, so a large number of parameters are comprised in this method to perform the feature learning process in big data. So, in real-time the feature learning and big data representation process are identified as a difficult process in big data of classical deep learning approaches. The feature learning process performed with high velocity is developed for various incremental learning approaches. An object arrived by stabilising the network structure is very much valuable for the online learning approach is known as the sorts of incremental learning technique for apprising the parameters. The limited weight adjustment was applied by (L.M Fu et al.)[16] for implementing an incremental learning approach. The information regarding the present network is covered over the newly arrived objects, in which for preventing the over-training the weights are not changed. This method is very much sensitive for the former knowledge, so defining the bound beforehand is identified as a difficult task. The bounded weight modification process is suitable only for the NNs that having only two layers, but this process is found to be a difficult approach for an incremental deep learning approach.

2.4. Deep learning for low-quality data

For the feature learning process of big data, one more challenge is emerged from its veracity. In big data, only the data having low-quality is obtained particularly due to the presence of noise, imprecise objects, redundant objects, huge quantity of incomplete data, and inaccurate objects. The development of data with low quality are need to the different reason exist in here. There are different causes for the development of data with low quality. For instance, an enormous amount of information is obtained from sensors, whereas some incomplete objects are collected due to the presence of broken sensors. In big data, the transmission fault that obtained in the network also introduces some noise.

The data having low-quality is not taken into account by various deep learning approaches. Otherwise it can be said in another way, i.e., methods of the deep learning are developed only to study the features of data having high quality. (P Vincent et al.)[17] presented a stacked denoising auto-encoder approach to perform the feature learning process in noise corrupted data. The

original data is reconstructed from the corrupted data by the denoising auto-encoder approach to train the parameters as showed in Fig. 4.5.

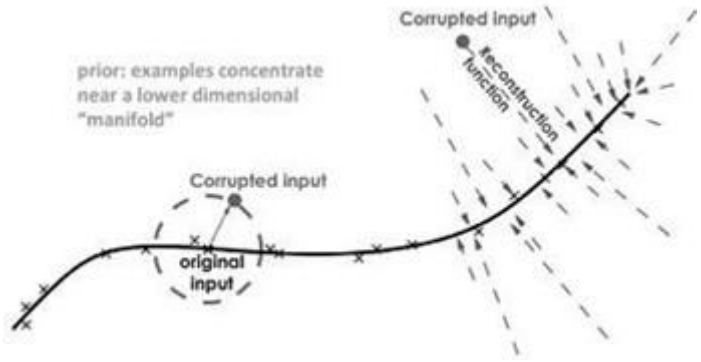


Fig. 4.5 Denoising auto-encoder (P Vincent et al.)[18]

The denoising auto-encoders objective function is eqn. (4.1) for the reconstruction of original input.

$$J = \sum_{t} E_{q(x\ x^{(t)})}\left[L(x^{(t)}, g\theta\,(f\theta\,(x)))\right] \tag{4.1}$$

Where, the corrupted cases of x(t) is represented as $\bar{x}$ on the basis of corruption $q(x\ x^{(t)})$ and the averages that are obtained over the instances $\bar{x}$ are represented as, $E_{q(x\bar{x})}[\,]$. .

The gradient descent methodology is applied to train the parameters of objective function. The isotropic pepper noise or Gaussian noise is added to obtain $\_x$. (P Vincent et al.)[18] offered a stacked denoising auto-encoder approach to perform the feature learning process in noise corrupted data. The features of incomplete object were learned as shown in Fig. 4.6, by the imputation auto-encoder model developed by (Q Zhang et al.)[19]. The small part of the attribute values that are obtained from the object x to 0, is located to generate the simulated incomplete object x′.

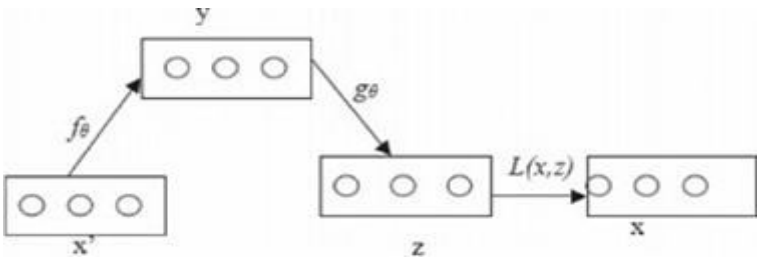


Fig. 4.6 Imputation auto-encoder model (Q Zhang et al.)[19]

The imputation auto-encoder approach output the reconstructed object z by considering an incomplete object x′ as an input.

$$z = g_\theta(f_\theta(x)) \tag{4.2}$$

The subsequent objective function is minimized to train the parameter θ:

$$J = L(x, z) \tag{4.3}$$

In Fig. 4.7, each imputation auto-encoders are stacked within the developed deep imputation network to learn the features of incomplete objects. Recently, a non-local auto-encoder was introduced by Wang and Tao to perform the reliable feature learning in corrupted data. The neurological observation motivates their work, where the human brain is stimulated by similar input to obtain the same response. Therefore, the similar hidden patterns are provided by this neural network for parallel input objects. In detail, suppose that h_1, h_2 and h_3 are the learned representations of x_1, x_2 and x_3, respectively. If $x_1 \sqsubset x_2 \sqsubset x_1 \sqsubset x_3$, $h_1 \sqsubset h_2$ should be smaller than $h_1 \sqsubset h_3$. However, the relationship $h_1 \sqsubset h_2 < h_1 \sqsubset h_3$ was not always guaranteed by the neural network, because the Sigmoid function i.e, non-linear function is employed in the neural networks as a function for activate the process.

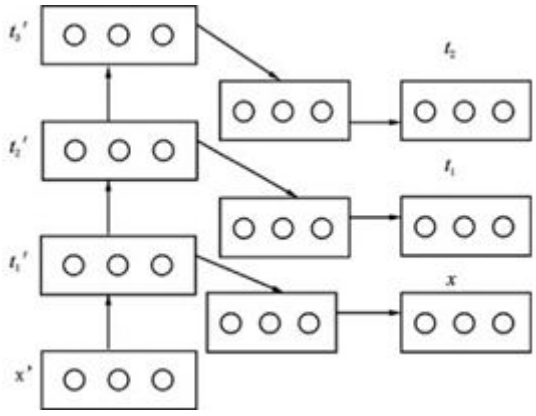


Fig. 4.7 Deep imputation network (Q Zhang et al.)[19].

A regularization term is presented by Wang and Tao by including the regularization term within the objective function to learn the reliable representations. Then, the training sample x is provided, so that a corrupted sample x_i is obtained by including a noise within x. The hidden representations for both x and x_i are represented as h_i and h respectively. Then, regularization is represented as in eqn. (4.4),

$$\sum_i \omega_i \left\| h - h_i \right\|_p \tag{4.4}$$

Where the weight of the i^{th} corrupted sample is represented as $\sqsubset_i$. The regularization term is applied to obtain the non-local auto-encoder method"s objective function.

$$J = x\left\| - g_\phi\left(f_\theta(x)\right)\right\|_2^2 + \lambda \sum_i \omega_i \left\| h - h_{ip} \right\| \tag{4.5}$$

This experiment demonstrates that superior performance is achieved in both image restoration and denoising process by the non-local auto-encoder approach.

3. CONCLUSION

The Deep Learning has the benefit of possibly delivering a solution for addressing the data assessment and learning difficulties discovered in the huge volume of input data as resisted to feature manufacturing algorithms and additional conventional machine learning. Specifically, it supports owing to segregating the representation of complex data from the extensive volume of unsupervised data. For the BDA, this creates it as an essential tool that contains information research from the huge gatherings of raw data which is commonly unsupervised and unclassified. Deep learning method proposed in this model was used for the feature learning of big data.

4. REFERENCES

[1] M. Vamsi Krishna, Sangram Keshari Swain and Srinivas Prasad, Unsupervised Feature Learning On Big Data Based On Deep Learning With Weighted Softmax Regression, International Journal of Innovative Technology and Exploring Engineering (IJITEE) 8, no.5 (2019): 752-761.

[2] Mladenic Dunja, Automatic word lemmatization. Proceedings B of the 5th International Multi- Conference Information Society IS. 2002, 153-159.

[3] L. Deng, D. Yu, J. Platt, Scalable stacking and learning for building deep architectures, Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2012. 2133 C2136.

[4] X. Chen, X. Lin, Big data deep learning: challenges and perspectives, IEEE Access 2 (2014): 514–525.

[5] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, M. Ranzato, A. Senior, P. Tucker, K. Yang, Q.V. Le, A.Y. Ng, Large scale distributed deep networks, Proceedings of Advances in Neural Information Processing Systems, MIT, 2012, pp. 1223–1231.

[6] X. Chen, X. Lin, Big data deep learning: challenges and perspectives, IEEE Access 2 (2014): 514–525.

[7] V. Vanhoucke, A. Senior, M. Mao, Improving the speed of neural networks on CPUs, Proceedings of Deep Learning and Unsupervised Feature Learning Workshop, MIT, 2011, pp. 1–8.

[8] R. Raina, A. Madhavan, A. Ng, Large-scale deep unsupervised learning using graphics processors, Proceedings of International Conference on Machine Learning, (2009), pp. 873 C880. ACM.

[9] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, A.Y. Ng, Multimodal deep learning, Proceedings of International Conference on Machine Learning, ACM, 2011, pp. 689–696.

[10] J. Xue, J. Li, Y. Gong, Restructuring of deep neural network acoustic models with singular value decomposition, Proceedings of Conference of the International Speech Communication Association, Springer, 2013. pp. 2365 C2369.

[11] W. Ouyang, X. Chu, X. Wang, Multi-source deep learning for human pose estimation, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2014, pp. 2337–2344.

[12] L. Zhao, Q. Hu, W. Wang, Heterogeneous feature selection with multi-modal deep neural networks and sparse group LASSO, IEEE Trans. Multimedia 17 (11) (2015): 1936–1948.

[13] A. Wang, J. Lu, J. Cai, T. Cham, G. Wang, Large-margin multi-modal deep learning for RGB-d object recognition, IEEE Trans. Multimedia 17 (11) (2015): 1887–1898.

[14] N. Neverova, C. Wolf, G. Taylor, F. Nebout, Moddrop: adaptive multi-modal gesture recognition, IEEE Trans. Pattern Anal. Mach. Intell. 38 (8) (2016) 1692–1706.

[15] S. Rastegar, M.S. Baghshah, H.R. Rabiee, S.M. Shojaee, MDL-CW: a multimodal deep learning framework with crossweights, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2016, pp. 2601–2609.

[16] L.M. Fu, H. Huang, J.C. Principe, Incremental backpropagation learning networks, IEEE Trans. Neural Netw. 7 (3) (1996): 757–761.

[17] P. Vincent, H. Larochelle, Y. Bengio, P. Manzagol, Extracting and composing robust features with denoising autoencoders, Proceedings of International Conference on Machine learning, ACM, 2008, pp. 1096–1103.

[18] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P. Manzagol, Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion, J. Mach. Learn. Res. 11 (2010): 3371–3408.

[19] Zhang, Q., Yang, L.T., Chen, Z. and Li, P., 2018. High-order possibilistic c-means algorithms based on tensor decompositions for big data in IoT. Information Fusion, 39, pp.72-80, 2018.

ABOUT THE EDITOR

Dr. Abdelgadir Elamin Eltom is an Associate Professor in Medical Laboratory Sciences Program of College of Health Sciences, Gulf Medical University. He has more than 18 years of teaching and research experience and published several scientific papers in reputed journals. He had been awarded "Academic Excellence Award" in the B.Sc. (MLS) final year examinations at university level.

ABOUT THE BOOK

We are living in a generation of fast-growing technology. We find a great emergence of new trends into the market which aids humans and creates a user-friendly environment for optimum utilization of technology.

When it comes to the present date, Artificial Intelligence (AI) and the Internet of things (IoT) are the two trends that have created a mark for them. Artificial intelligence (AI) is all about adding the feature of intelligence to the machines and making them smart and intelligent as humans. The Internet of Things (IoT) is all about connecting all electronic devices to a single network which are interlinked.

AI also has scientific and engineering goals. AI technology has not only automated the industrial process for making efficient and effective output but helps the individuals and machines to work together. AI is being enforced not just into manufacturing but also into other departments such as sales, marketing, customer support, and research and development.

The Internet of things (IoT) is another emerging trend in computer science, IoT is all about an ecosystem of electronics devices connected to one single network through the internet and wireless connectivity. All the connected devices can send information through the inbuilt intelligent technology which helps in creating smart devices and many more such as creating wearable devices, smart cities, and smart homes.

This Edited Book on "Application of Artificial Intelligent and IoT" is an attempt to bring the research work done in the field of Artificial Intelligent and IoT to a world-wide audience. I am sure that the innovations mentioned in this book will proof to be very helpful.

EMPYREAL PUBLISHING HOUSE

ISBN 978-93-93810-31-1

India | UAE | Nigeria | Uzbekistan | Montenegro | Iraq | Egypt | Thailand | Uganda | Philippines | Indonesia

www.empyrealpublishinghouse.com