Proceedings of

FUTURISTIC TECHNOLOGIES AND INDUSTRY 5.0

ICFTI - 22

EDITORS

DR. K. RAJKUMAR

DR. P. S. ELIAHIM JEEVARAJ

DR. M. P. ANURADHA

Proceeding of

Futuristic Technologies & Industry 5.0

Editors

Dr. K. Rajkumar

Dr. P. S. Eliahim Jeevaraj

Dr. M. P. Anuradha

Proceedings of Futuristic Technologies & Industry 5.0

Editors: Dr. K. Rajkumar, Dr. P.S. Eliahim Jeevaraj, Dr. M.P. Anuradha

Published by The Department of Computer Science, Bishop Heber College (Autonomos), Tiruchirappalli.

College Profile

Bishop Heber College is a religious minority educational institution established by the Tiruchirappalli - Thanjavur Diocese of the Church of South India. Its main objective is to cater for the higher educational needs of the Christians and the students of all faiths and religions.

The College now offers 29 Under-Graduate, 22 Post-Graduate besides research programmes leading to M.Phil. and Ph.D. The College was accredited with 'A' Grade with a CGPA of 3.58 on a 4 Point Scale on 3rd March 2015, and our college tops among all the Affiliated Colleges of the Bharathidasan University, Triuchirappalli.

The College has been granted an amount of Rs.1 Crore by the DST-FIST for the scheme "College as Whole" in the year 2016. th College has been awarded "Star College Status" by DBT in the year 2016. The College has secured 4th rank at National Level Awarded by MHRD in the year 2017.

Department Profile

The PG Department of Computer Science is the pioneer in computer science courses, being the first college to offer a degree course in Computer Science in the city of Tiruchirapalli from the year 1984.

The Department offers B.Sc. Computer Science, M.Sc. Computer Science, M.Phil. and Ph.D Programmes. The faculty members of the department are highly qualified and act as mentors, guiding students to help them develop the necessary technical skills and attitudes required to become successful IT managers and corporate citizens. They are also involved in research activities in various fields like Networking, IoT, Computer Vision, Data Analysis and Big Data Analysis. The citation index of the department is more than 750 and h-index of the department is 25.

About ICFTI – 2022

Recent years, the computer technologies has established as an application-oriented domain and attracts scientists working in diverse fields of interest medicine, education, governance, space technology, Industries and Entertainment due to the impact it has created in such fields. The computing technologies are unavoidable one in the day-to-day life and it contributes in all the times in the application domain. In this backdrop, it is planned to conduct the ICFTI -2022 at Bishop Heber College (Virtual Mode).

Objectives

- To encourage interdisciplinary and multidisciplinary research in the area of Computing and Information Sciences.
- To afford a common platform for academicians from universities / colleges and experts from industries / national laboratories for exchange of ideas related to research in Computer Science Domain.
- To explore the new interdisciplinary areas of research for the applications of Computer Science.

Themes of the Conferences

- Computer Networks

- Software Engineering

- Signal & Image Processing

- Data Mining

- Data Science

- Big Data

- Internet of Things

- Could Computing

- Wireless Networks

- Industry 5.0

- Artificial Intelligence

- Network Security

- Video Processing

- Pattern Recognition

Organizing Committee

CHIEF PATRON
Rt. Rev. Dr. D. Chandrasekaran
Bishop, CSI Trichy -Thanjavur Diocese

PATRON

Dr. D. Paul Dhayabaran
Principal

CONFERENCE CHAIRS

Dr. J.G.R. Sathiaseelan
Head, Dept. of CS (Aided)

Dr. K. Rajkumar
Head, Dept. of CS (Shift -I)

Dr. G. Sobers Smiles David
Head, Dept. of CS (Shift - II)

ORGANIZING SECRETARIES

Dr. P.S. Eliahim Jeevaraj
Asst. Professor, PG Dept. of CS(SF-I).

Dr. M. P. Anuradha
Asst. Professor, PG Dept. of CS(SF-I).

Local Organization Members

Dr. S. Sophia, *Assistant Professor, Dept. of CS, BHC*

Dr. D. Kirubai, *Assistant Professor, Dept. of CS, BHC*

Prof. C. Sathish Kumar, *Assistant Professor, Dept. of CS, BHC*

Dr. B. Karthikeyan, *Assistant Professor, Dept. of CS, BHC*

Dr. R. Ramah Sivakumar, *Assistant Professor, Dept. of CS, BHC*

Prof. V. Bhuvaneswari, *Assistant Professor, Dept. of CS, BHC*

Prof. L. Prisicilla, *Assistant Professor, Dept. of CS, BHC*

Prof. G. Lavanya, *Assistant Professor, Dept. of CS, BHC*

Advisory Committee Members

Dr. Asai Asaithambi, Professor, University of North Florida, USA.

Dr. Mariia Skulysh, Professor, National Technical Univ. of Ukraine ,Kyiv.

Dr. Suriyapriya Asaithambi, Professor, National Univ. of Singapore..

Dr. P. Shanmugavadivu, Professor, GRI-DTBU, Gandhigram.

Dr. S.Rajasekaran, Professor, Univ. of Tech.& App.Sciences, Ibri, Oman.

Dr. V. Sivakumar, Asia Pacific Univ. of Tech. and Innovation, Malaysia.

Dr. R. Sivashankar, Chiang Mai University, Thailand

Dr. Hemalatha Mahalingam, King AbdulAziz Univeristy, Saudi Arabia

Dr. E. George Dharma Prakash, Asso. Prof. SCEA, BDU,Trichy

Dr. Vasudevan, Asst. Prof., NIT, Calicut

Dr. A. Muruganantham, General Manager, Haxaware, Chennai

Mr.V. Sridharan, Assoc. Vice-President, Global Logic, Bangaluru

Mr.S. Kamalanathan, BOSCH Global Software Tech, Coimbatore

Mr.Rahul Ganapathy, CEO, Atsuya Technologies, Chennai

Dynamics of the Conference

This is the first edition of ICFTI 2022. Recent years, the computer technologies have established as an application-oriented domain and attracts scientists working in diverse fields of interest medicine, education, governance, space technology, Industries and Entertainment due to the impact it has created in such fields. The computing technologies are unavoidable one in the day-to-day life and it contributes in all the times in the application domain. In this backdrop, it is planned to conduct the ICFTI -2022 at Bishop Heber College.

Tracks

- Machine Learning & Deep Learning
- Computer Networks & Security
- Cloud & IoT

Paper Details

Totally 66 Papers are received. The authors submitted the research articles from States of India, China and South Asia such as Malaysia, Thailand.

Keynote details

The Guest of Honour of the conference Prof. George Ghinea, Professor, Brunel University have participated and delivered the chief guest address.

The conference includes two keynote address. The First Keynote address on "AI Healthcare" will be delivered by Dr. V. Sivakumar, Associate Professor, Asia Pacific University, Kuala Lumpur, Malaysia.

The Second Keynote Address on "Challenges and Responsibilities of IT people for next few years towards the Growth of South Asia" by Dr. R. Sivashankar, Lecturer, International College of Digital Innovation, Chiang Mai University, Thailand.

TABLE OF CONTENTS

The Effect of Flipped Teaching and Acceptance of Curriculum Innovation on Students' Entrepreneurship Motivation: A Case Study from Youth Entrepreneurial

Chao Jin

Innovation College North-Chiang Mai University, Chiang Mai,

169, Moo3, Nong Kaew, Hong Dong, Chiang Mai ,50230,Thailand

innovation.med@northcm.ac.th

Abstract

In modern society, the increase in the weight of knowledge, the acceleration of substitution, and the emergence of new disciplines have contributed to the renewal of teaching content and curriculum reform, calling for the inevitable requirements of lifelong education. To become a knowledgeable teacher, we must adapt to the needs of modern education, constantly learning new things and updating our knowledge structures. The emergence of the knowledge economy requires a new way of production and a further education to keep up with the times. These changes cannot be separated from the cultivation and training of knowledge and innovative talents in higher education. Therefore, the breakthrough of comprehensive quality education has become a creative activity where everyone can participate.

College students are an essential group to promote the development of the national entrepreneurial economy. The number of college students who choose to start their own business is less; society needs to guide college students to create their own business actively. Individual behavior is induced by personal motivation. Their entrepreneurial motivation also drives college students' entrepreneurial behavior. Therefore, it is vital to study the entrepreneurial reason of college students, which has important theoretical and practical significance. The development of entrepreneurship education is conducive to promoting and guiding students' entrepreneurship. The research on entrepreneurship education is of great guiding importance to the expansion of educational concepts in colleges and universities, and the construction of a new theoretical system of a college education can better play the role of guiding students' entrepreneurship, which is of great practical significance to the development of national entrepreneurship economy.

This study focuses on the impact of flipped teaching and curriculum innovation acceptance on students' entrepreneurial motivation by examining the integration model of university students' counseling courses. The first part is the background and significance, the second part is the literature review, the third part is the design and methodology of the study, the fourth part is the results and analysis of the study, and the fifth part is the conclusion and recommendations. The research of variables demonstrates that flipped teaching and curriculum innovation positively impact students' entrepreneurial motivation.

Keywords: *Flipped Education, Curriculum Innovation, Entrepreneurial Motivation*

1. Introduction

Renovation education as an educational principle has different requirements at different academic levels. Institutions of higher education are not only the talent training base but are also a place for the generation of knowledge and technology innovation, renovation of education more performance for cultivating the ability of the students' mastery of knowledge transformation and the creation of new knowledge, and constantly learning new knowledge, and continuously update their knowledge structure, namely train pedagogue independent learning process, the bold exploration, innovation ability, therefore, It is of great significance to explore and implement the integration of ideological and political education and employment education.

Flip education through the internalization of knowledge and knowledge reverse arrangement, change the roles between teachers and students in the traditional teaching the use of classroom time to plan, implement the innovation of conventional education to the standard education model usually includes knowledge and knowledge internalization in two stages. Knowledge imparting is accomplished by teachers' teaching in class, while knowledge internalization is performed by students' homework, operation, or practice after class. In terms of flipped instruction, this form is overturned. Knowledge is taught after class through information technology, and knowledge is internalized in course with the help of teachers and classmates, thus forming flipped education. However, curriculum innovation is an education that cultivates people's innovative spirit and ability as the fundamental value orientation.

There should be innovation in the teaching process so that the potential of young people can be better developed. The best education is the education that excavates the potential of people. These qualities can be cultivated through proper physical, emotional, and intellectual handling. In traditional education, authority figures (teachers) want answers and solutions that they know well but don't want to see things that surprise them, so it's hard to foster creativity. In the teaching process, students have a significant feature active, combined with the characteristics in classroom teaching, can use of the dynamic method, on the one hand, can satisfy the desire of students' psychological dynamic, on the other hand, can let students in learning dynamic cultivate interest in learning, the potential of imagination, creativity, and other aspects.

This is consistent with the pace of development of human society. In the internal renewal of education, teachers should pay attention to the continuous development of students and gradually accept students who are different today and yesterday. We should recognize that everyone is an individual and a concrete person. Therefore, flip education is to cultivate innovative people. Flip education is to follow the principle that people are still human selves, that is, concrete people whose intrinsic nature and world outlook are unique in the performance of mutual relations with others. Therefore, we should pay attention to the cultivation of students' self-esteem. In the renovation of the education of students, we should understand that students are unique people and try to make them outstanding in the field of labor creation.

Renovation education benefits society and society can also benefit from renovation education. Promoting social development is the due righteousness and sacred responsibility of renovation education. The purpose of renovation education is to meet the individual employment demand and the objective demand of the job post, then promote the development of social productivity, and accelerate the adjustment and transformation of the national industrial structure. Students' horizons are becoming wider, and with the support of their parents, many students are leaving the traditional education system to find ways of education and careers that suit their interests. Therefore, professional and targeted renovation education has become a more and more popular option.

2. Objective

Governments worldwide have fully realized that college students' entrepreneurship plays a vital role in the better and faster development of their economy, solving the problem of college students' employment and providing more employment opportunities for society. However, how to effectively promote entrepreneurship among college students? Society, government, and schools are thinking about this problem and actively seek solutions. The author is also thinking about this problem, hoping to find effective solutions through the research of this paper. This paper takes college students as the research object, through the analysis of university flipped education, the interpretation of entrepreneurship education, to explore the degree of entrepreneurship education, the acceptance of curriculum innovation on entrepreneurial motivation, and the mechanism of action. It also discusses how to design entrepreneurship education projects better to cultivate students' entrepreneurial spirit and sense of adventure, effectively improve the success rate of entrepreneurship, and provides basic suggestions and measures for universities and governments to formulate entrepreneurship education projects and policies.

To discuss the feasibility of applying the Chinese experience to Thailand, China's flipped education aims to promote youth employment to serve society and meet the group's human needs. The essential link between renovation education and economic development, productivity cohesion, and human resources as the center of all forces, increased production from education, and renovation education to promote national world economic development of modern renovation education. Therefore, China's flipping experience can be applied in Thailand and promote youth entrepreneurship development.

Once established, flipped education will become a guide for people to act. Pointed out the development direction of it is not only for educates, due to the development result, but also for education workers work pointed out the direction and goals. Flip education profoundly affects the education workers education. Students view, inspiring education workers to cultivate talents, according to the terms of the education policy and education purpose profoundly affects the education of education workers and students, The inspection of education and teaching quality, the evaluation of teachers' teaching quality and work effect; The assessment of students' learning quality and development degree must be based on teaching purpose. Therefore, the educational purpose is goal-oriented for both the educated and the educator.

The purpose of education is to meet the individual employment demand and the objective needs of the job and then promote the development of social productivity to speed up the adjustment and transformation of the national industrial structure. Flip education is the product of social development, human civilization, and the development of human beings themselves. And it was the product of a particular period of development. The purpose of flipped education is to train applied talents and workers with a certain level of education and professional knowledge and skills. Compared with general education and adult education, flipped education focuses on cultivating practical skills and practical working ability. Therefore, flipped education is the overall goal of cultivating human autonomy. It is related to what kind of social role and quality the educates will be trained into, and it is the starting point of educational practice activities. It is determined according to the needs of certain social productive forces and people's development and advocates youth entrepreneurship as the goal.

3. Expected Benefits

Flip educational knowledge also plays an important role in cultivating students' motivation for autonomous learning. And flip education trains young people to start their businesses. Flip education not only trains young people to impart cultural knowledge independently but also has the habit of being good at thinking and carries out

ideological and moral education to shape the young generation's beautiful mind through the ability of independent learning to create a certain grasp of cultural knowledge, with a certain moral quality and spiritual outlook of the young members of the society. Therefore, Flip education in China has a crucial impact on the moral education of young people. Through the cultivation of independent cognition of young people, oral education, and independent thinking behavior, the goal of choosing their own business can be achieved.

Flip education plays an integral role in schools. School is to cultivate talents in every respect, sets up the flip education reflects the characteristics of innovation education. That is to highlight the ethos of the school and give the society to cultivate entrepreneurial youth. Therefore, flipped education plays an important role in schools. The conclusion of this paper can be a reference for the construction and entrepreneurship education in colleges and universities. In carrying out entrepreneurship education, colleges and universities can train students with entrepreneurial motivation to start their businesses and improve their success rate of entrepreneurship. This will help ease the pressure on employment and contribute to society's harmonious and stable development.

4. Research Methodology

Hypotheticals are hypothetical answers to questions posed by the topic, and statements about the relationship between variables are made based on these answers. A hypothesis is a hypothetical explanation of an unknown fact based on existing factual materials and scientific principles.

First, Educational research plans can be formulated by consulting data carefully, thinking carefully, and discussing and revising carefully. Second, be specific. Educational and scientific research plans must be clear and specific, but there should be no jargon, nonsense, and unnecessary modifiers. Third, scientific. The formulation of the research plan should conform to the requirements of educational research methods and be carried out based on grasping certain theoretical and factual materials. The formulation of the research plan should be feasible, fully considering their research ability and research conditions.

This study is about questionnaire collection and testing. It focuses on the design of the questionnaire in terms of the dimensions of the variables and the analysis of the information collected on the data through SPSS 21.0, the reliability testing of the two variables involved in this study, and the validity verification of the scales using AMOS. The data were also analyzed and processed. The main purpose was to perform correlation and regression analysis of the two variables through SPSS 21.0 software, illustrate the relationship between the variables in this study, and test the hypotheses formulated.

The frequency analysis shows that for gender. 106 or 51.5% of the students were male. The number of female students was 100, or 48.5%. This indicates that the number of male students was slightly higher. The distribution chart is shown in Figure 1.

The frequency analysis shows that 46 people (22.3%) were in first year. In sophomore year, there were 60 students (29.1%). There were 57 juniors (27.7%). In the fourth year, there were 43 people, accounting for 20.9%. This means that the largest number of respondents were sophomores.

The frequency analysis shows that 38 people, or 18.4%, were in science and technology. In agriculture and forestry, there were 53 students, accounting for 25.7%. There were 56 people in Medicine, accounting for 27.2%. There were 36 people in the Humanities category, accounting for 17.5%. 16 people in the Economics and Management category, accounting for 7.8%. There were seven people in the architecture category, accounting for 3.4%. This means that the largest number of respondents were in the medical category and the smallest number were

in the architecture category.
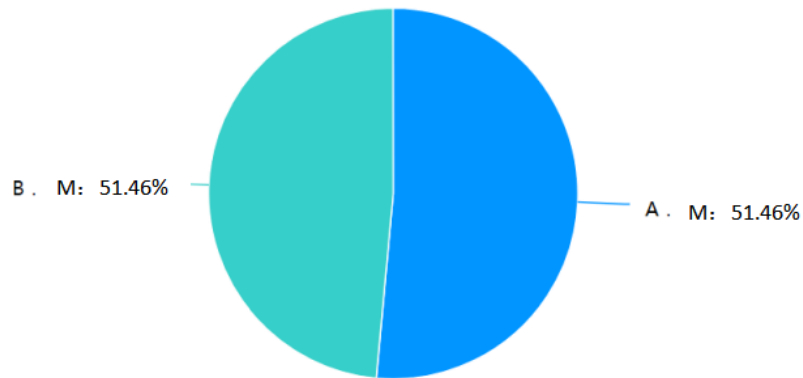
Figure 1. Pie charts for gender
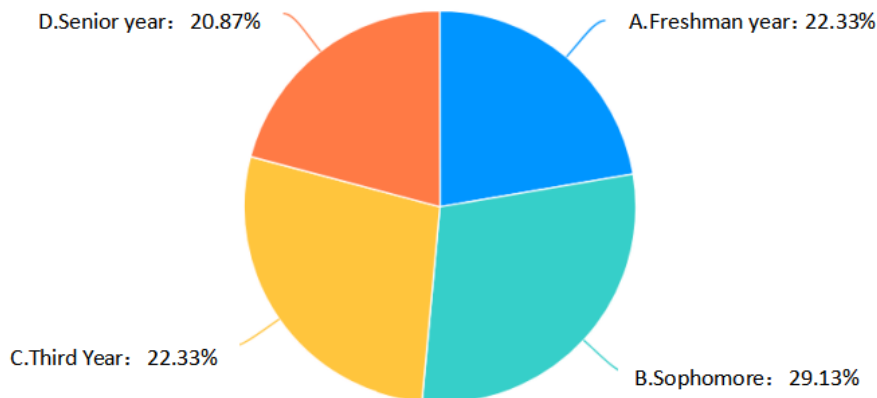
Figure 2. Pie charts for grades

From the analysis of the statistical graphics, we can see that it shows that the largest number of respondents are sophomores. This is also in line with the general data distribution pattern.

The existing demographic analysis provides few results, and we need to apply other statistical methods to extract more useful information

Summary of the Study

The core purpose is to discuss the feasibility of applying the Chinese experience to Thailand. The form of the flipped classroom is no longer strange. Student-centered teachers adopt task-driven teaching and turn classroom tasks into cooperative projects between students to master more skills in learning and put theory into practice. For students majoring in preschool education, the ability to practice innovation is particularly important. They should master not only teachers' basic theoretical ability but also master professional technical ability.

Most importantly, they should learn to deal with various problems in their future careers, so they need independent thinking ability. Renovation education can make students master scientific and cultural theory knowledge and pay more attention to the cultivation of students' practical operation skills and professional ability. The renovation of

education is to cultivate professional talents with the solid beginning ability and professional skills; the young people only through skills training can skills talents adapt to market demand and realize the transferred business. Therefore, China's experience can be implemented in Thailand.

Discussions

The purpose of renovation education is not to make students invent many new things but to cultivate students' innovative consciousness, innovative concept, and innovative attitude and shape their creative ability through effective education and teaching approaches. Knowledge weight and upgrading to accelerate the growth of modern society, the emergence of new disciplines, promote the renewal of teaching contents and curriculum reform, calling for education, lifelong learning becomes the inevitable requirement of modern, become an imparter of knowledge, more want to meet the needs of the development of modern education, constantly learning new knowledge and constantly update their knowledge structure. Therefore, inheritance is learning; Innovation is learning.

The regression analysis in the previous section shows that entrepreneurship education positively impacts entrepreneurial motivation. Of these, personal factors have a smaller impact on entrepreneurial motivation than school factors. Individuals are mainly subjective in their willingness to receive entrepreneurship education and are more influenced by their abilities and sense of efficacy. However, the shaping of a good environmental climate can effectively motivate university students to do something. According to the organic integration and causal orientation theories of self-determination theory, the external environment can influence an individual's behavior and stimulate extrinsic motivation, which can be effectively transformed into intrinsic motivation and lead to increased motivation. Therefore, when students are influenced by the entrepreneurial atmosphere on campus, they will be motivated to start their own business. Their motivation will increase, which will be more effective in promoting students to act for entrepreneurship.

Recommendations

Entrepreneurship education in colleges and universities has a greater influence on college students' entrepreneurship. Colleges and universities should further deepen and reform entrepreneurship education, which aims to cultivate and improve college students' entrepreneurial knowledge, entrepreneurial skills, and entrepreneurial quality, create an atmosphere for college students' entrepreneurship, provide technical support and financial assistance, and help students to start their businesses successfully.

A perfect organizational structure and institutional guarantee are prerequisites for entrepreneurship education's smooth implementation and execution. Chinese universities should continuously improve their organizational structure. According to the data analysis of the questionnaire, a good external environment has a positive impact on college students.'

This study finds that the degree of participation of university students in entrepreneurship education has a certain influence on the strength of entrepreneurial motivation. This requires close integration of entrepreneurship education with professional education and a seamless connection between entrepreneurship education and inducing entrepreneurial motivation to induce entrepreneurial behavior. The questionnaire found that some students do not participate much in entrepreneurship seminars or forums. Some of the open-ended questions indicated that there were not many entrepreneurship-related lectures in their majors, so they hoped that the university would strengthen

the integration of entrepreneurship education with their majors. Based on the students' suggestions and the organic integration of internal and external motivation in self-determination theory, the author believes that universities can strengthen this aspect in the following ways, through theoretical learning, to meet the self-confidence of university entrepreneurship, and to improve the cognitive evaluation of internal and external motivation into entrepreneurial motivation.

The study results and the results of the questionnaire survey show that many respondents believe that entrepreneurship guidance for college students in colleges and universities mostly stays in theoretical aspects, with relatively few practical courses. As a result, many students have a good grasp of the theory but cannot put it into practice. Practice is the only criterion to test the truth. Therefore, according to self-determination theory, Chinese universities should continue to increase and enrich the establishment of practical courses in entrepreneurship education, continue to meet the basic psychological needs of students, enhance the sense of autonomy and competence, and provide more practical activities and opportunities for entrepreneurial activities to students, effectively transform the external motivation of entrepreneurship into intrinsic motivation, so that students' hands-on practical ability in entrepreneurship can be continuously enhanced, and continuously promote.

Through this study, I have a general understanding of the current relationship between entrepreneurship education, the flipped classroom, entrepreneurial motivation and the current state of research. However, not many Chinese scholars have studied the relationship between the three, and even fewer have done so in China's specific cultural context, leaving much room for research in this area. Therefore, this study suggests that future research could be conducted in the following areas.

Existing foreign research has shown that many factors influence entrepreneurial motivation. The entrepreneurship education and flipped classroom mentioned in this paper are also influential factors. And many foreign studies have studied culture as one of its influencing factors. Therefore, Chinese scholars should explore the factors influencing entrepreneurial motivation in future research based on China's cultural background, the state of China's economic development, and the context of the "double creation" era. In addition, comparative studies can be conducted with foreign studies.

There is still no unified definition of the dimensions of the flipped classroom, entrepreneurial motivation and entrepreneurship education in the existing studies. Based on different contexts and theories, the dimensions classified are different. Therefore, it is hoped that Chinese scholars can make a reasonable division of the dimensions of the flipped classroom, entrepreneurial motivation, and entrepreneurship education based on the cultural background of China, combined with theories of Chinese construction and economic and social development.

References

[1] Chai, Yue-jiao (1995). Revitalizing education: An important way to relieve employment pressure in China. Economic problems.

[2] Cong, Xinqiang (2018). The Religious Dimension of Contemporary Chinese Literature. Literary theory and Critical Literature. China Social Sciences Press.

[3] Chen jiangsheng, Zhang Xiaodong, Shi Jianglan, Fan Xiurong.(2005) An analysis of educational investment structure in western China. Journal of Northwest A&F University social Science Edition.

[4] Du yuhong (2000) Study on the imbalance of educational development. Beijing Normal University Press.

[5] Du qinying (2021). Innovative practice of renovation education based on modern apprenticeship. Xiamen Business Tourism School.

[6] Fu chuanming (2004). The widening trend of individual income distribution gap in China: Theoretical analysis and countermeasures. Exploration of economic issues.

[7]   Guo, qijia (1987). The History of Chinese Educational Thought. Education Science Press.

[8]   Johnson, D. (1998). Education, labor market and income distribution. Economic research.

[9]   Li jingwen (1997). Application of theoretical model of modern human resource economic analysis. Beijing. Social Sciences Academic Press.

[10] Li Haijuan (2021.) Exploring the talent training path of distance renovation education under the guidance of innovation and entrepreneurship. Heilongjiang University of Finance and Economics humanities Science Department

[11]]Li Peilin (2003). Changes in China: Current Social Development and Existing problems. China Economic Times.

[12] Li Fengliang, Xia Guisong, Zhao Xihong, Zhang Shaogang (2009). An empirical study on the profitability of distance higher education in China: A case study of Nanjing. Harbin Central Television University undergraduate

[13] Luo Rundong, Yang Xinming (2009). Comparison of human capital formation patterns in pearl River Delta, Yangtze River Delta and Beijing-Tianjin-Hebei Region. East China Economic Management.

[14] Lenin (2009) Collection of monographs: On the proletarian Party. Beijing: People's Publishing House.

[15] Meng Dahu (2005). Institutional segmentation, stickiness of human capital and unemployment. Take the Northeast region for example. Changbai Journal.

[16] Selected works of Marx and Engels (1995) (vol. 1). Beijing People's Publishing House.

[17] Sun Peiqing (2009). History of Chinese Education. Education Science Press.

[18] Tao xingzhi (2008) Anthology of Tao Xingzhi. Jiangsu Education Publishing House.

[19] Wang meiyan (2009). Education return and allocation of urban and rural education resources.

[20] Wu Binru (2021). Promoting high-quality development of renovation education. Guizhou Daily

[21] Wei ge (2020). The Generation of Teacher Practical Knowledge. Education Science Press.

[22] Wei Min (2021). On the status quo and development trend of preschool education in higher vocational colleges. Yunnan Vocational College of Culture and Art

[23] Wang yun-duo (2009). An Empirical Analysis of the Impact of education level difference on income distribution. Education Science Edition.

[24] Wang yun-duo (2008). The quantile mean test of education Minser's return rate. Journal of Inner Mongolia Normal University education Science edition

[25] Wu keming, Cheng Gang (2008). The income effect of education: a new perspective on labor mobility. Education and the economy.

[26] Xu Dongdong (2018). Practice and operation mechanism of university-enterprise cooperation mode in renovation education. Scientific Research Department of Shenzhen Polytechnic

[27] Xi Jinping (2019). On Upholding the Party's leadership over all work. Beijing Central Literature Publishing House.

[28] Xi Jinping (2017) On The Governance of China (Volume 2). Beijing Foreign Languages Press.

[29] Xi Jinping (2019) Study Outline of Socialism with Chinese Characteristics for a New Era. Beijing Learning Press, People's Publishing House.

Flow Control Using Distance Vector Routing

B. Karthikeyan, Benin S

Department of Computer Science, Bishop Heber Collge, Tiruchirappalli.

karthikeyan.cs@bhc.edu.in

Abstract

This paper simulates a network traffic in a given topology where the distance vector algorithm is used for path computations and routing. The topology of the network is given and cannot be changed by the user. User can change the path costs between each router, and specify any number of flows from any router to the other. Package size also can be changed. After the user specifies, the program can run the network traffic. The main prepose of paper is Flood attacks prevention. The Flood attack is also known as Denial of Service (DoS) attacks. In a flood attack, attackers send a very high volume of traffic to a system so that it cannot examine and allow permitted network traffic. Network flooding attacks have long been a standard part of an attacker's toolbox for denying service. The basic concept is either send a massive amount of traffic at a particular server or service with the aim of exhausting all its resources trying to respond to bogus traffic so that it cannot process legitimate requests for service.

Keywords: *Distance vector algorithm, Topology, Network traffic, Path computations and Routing, Flood attacks prevention*.

Introduction

In information technology, a network is defined as the connection of at least two computer systems, either by a cable or a wireless connection. The simplest network is a combination of two computers connected by a cable. This type of network is called a peer-to- peer network. There is no hierarchy in this network; both participants have equal privileges. Each computer has access to the data of the other device and can share resources such as disk space, applications or peripheral devices (printers, etc.)

Today's networks tend to be a bit more complex and don't just consist of two computers. Systems with more than ten participants usually use client-server networks. In these networks, a central computer (server) provides resources to the other participants in the network (clients).

In Mobile Ad hoc Network (MANET), nodes do not know the topology of their network; instead, they have to discover it by their own as the topology in the ad-hoc network is dynamic topology. The basic rules is that a new node whenever enters into an ad-hoc network must announce its arrival and presence and should also listen to similar announcement broadcasts made by other mobile nodes.

In a computer network, flooding occurs when a router uses a nonadaptive routing algorithm to send an incoming packet to every outgoing link except the node on which the packet arrived. Flooding is a way to distribute routing protocols updates quickly to every node in a large network. Examples of these protocols include the Open Shortest Path First and Distance Vector Multicast Routing Protocol.

Distance Vector Routing

A distance-vector routing (DVR) protocol requires that a router inform its neighbors of topology changes periodically. Historically known as the old ARPANET routing algorithm (or known as Bellman-Ford algorithm). Bellman Ford Basics – Each router maintains a Distance Vector table containing the distance between itself and ALL possible destination nodes. Distances, based on a chosen metric, are computed using information from the neighbors' distance vectors.

DVR Protocol Works

- In DVR, each router maintains a routing table. It contains only one entry for each router. It contains two parts − a preferred outgoing line to use for that destination and an estimate of time (delay). Tables are updated by exchanging the information with the neighbor's nodes.

- Each router knows the delay in reaching its neighbors (Ex − send echo request).

- Routers periodically exchange routing tables with each of their neighbors.

- It compares the delay in its local table with the delay in the neighbor's table and the cost of reaching that neighbor.

- If the path via the neighbor has a lower cost, then the router updates its local table to forward packets to the neighbor.

Literature Review

Jorge A. Cobb et al [1]: In this paper, we presented a randomized distance vector routing protocol to distribute the data traffic randomly over all available paths to a destination in the network for data load balancing. Our simulation results show that the proposed routing protocol has an improved performance with respect to throughput, message loss and link utilization under both light and heavy loads.

Jyotsana Pandey et al [2]: We have seen that the method shown above ie dynamic processing method yields the optimal method for the packets to travel in the network in a very efficient way. This is a method different from the method followed frequently and traditionally for working with the routing table. This method proves to be efficient and optimal and also always finds the correct result for the most efficient path for visiting the nodes. Apart from this method, one can try more processes to find out the same which may prove to be even better than this

K. Uday Kumar Reddy et al [3]: This paper presents the proposed AODV routing protocol with link prediction for ad hoc networks. The routing protocol which presents a common structure for optimization solutions to dynamic routing problems in MANETs. The proposed prediction function is that predicts link breaks based on signal strength of the successive traditional packets and threshold signal strength. This Work presented a detailed comparative analysis in terms of protocol design and simulation related parameters for Proposed and existing protocols. The performance of simulation results are analyzed investigated and compared the proposed AODVLP model with AODV. The simulation results are observed that the proposed algorithm AODLP performs

is better than the AODV.

Proposed System

Distance Vector Routing Algorithm

- The Distance vector algorithm is iterative, asynchronous and distributed.

- Distributed: It is distributed in that each node receives information from one or more of its directly attached neighbors, performs calculation and then distributes the result back to its neighbors.

- Iterative: It is iterative in that its process continues until no more information is available to be exchanged between neighbors.

- Asynchronous: It does not require that all of its nodes operate in the lock step with each other.

- The Distance vector algorithm is a dynamic algorithm.

- It is mainly used in ARPANET, and RIP.

- Each router maintains a distance table known as Vector.

Three Keys to understand the working of Distance Vector Routing Algorithm:

- Knowledge about the whole network: Each router shares its knowledge through the entire network. The Router sends its collected knowledge about the network to its neighbors.

- Routing only to neighbors: The router sends its knowledge about the network to only those routers which have direct links. The router sends whatever it has about the network through the ports. The information is received by the router and uses the information to update its own routing table.

- Information sharing at regular intervals: Within 30 seconds, the router sends the information to the neighboring routers.

Distance Vector Routing Algorithm

Let $d_X(y)$ be the cost of the least-cost path from node x to node y. The least costs are related by Bellman-Ford equation,

$d_X(y) = min_V\{c(x,v) + d_V(y)\}$

where the min_V is the equation taken for all x neighbors. After traveling from x to v, if we consider the least-cost path from v to y, the path cost will be $c(x,v)+d_V(y)$. The least cost from x to y is the minimum of $c(x,v)+d_V(y)$ taken over all neighbors.

With the Distance Vector Routing algorithm, the node x contains the following routing information:

o For each neighbor v, the cost $c(x,v)$ is the path cost from x to directly attached neighbor, v.

    o   The distance vector x, i.e., $D_X = [D_X(y) : y$ in N], containing its cost to all destinations, y, in N.

    o   The distance vector of each of its neighbors, i.e., $D_V = [D_V(y) : y$ in N] for each neighbor v of x.

Distance vector routing is an asynchronous algorithm in which node x sends the copy of its distance vector to all its neighbors. When node x receives the new distance vector from one of its neighboring vector, v, it saves the distance vector of v and uses the Bellman-Ford equation to update its own distance vector. The equation is given below:

$d_X(y) = \min_V\{ c(x,v) + d_V(y)\}$ for each node y in N

The node x has updated its own distance vector table by using the above equation and sends its updated table to all its neighbors so that they can update their own distance vectors.

Results & Discussion

Completion Status of Flows: [true, true, true, true] Flow
Routing has done successfully at time = 40.0.

        f0's final path: [2, 0, 1, 4], done at time = 20.0.

        f1's final path: [3, 2, 1, 0], done at time = 30.0.

        f2's final path: [3, 4, 1], done at time = 40.0.

        f3's final path: [3, 2, 0], done at time = 35.0.

Distance vector routing works as follows. Each router maintains a routing table. Each entry of the table contains a specific destination, a metric (the shortest distance to the destination), and the next hop on the shortest path from the current router to the destination.

Conclusion

In this paper presented a distance vector routing protocol to find the data traffic over all available paths to a destination in the nodes for flow routing. Our simulation results show that the proposed routing protocol has an improved performance with respect to throughput.

References

[1]   Sangman Bak, Jorge A. Cobb, "*Randomized Distance-Vector Routing Protocol*", ResearchGate, Conference Paper, Jan 1999.

[2]   Jyotsana Pandey, Ashish, "*Finding the Routing Table by Dynamic Processing Method in Distance Vector Routing*", International Journal Of Engineering And Computer Science, ISSN: 2319-7242, Volume 6 Issue 1, pp: 20007-20010, Jan. 2017.

[3]   Bhavna Rathi, Er. Farminder Singh, "*Performance Analysis of Distance Vector and Link State Routing Protocols*", International Journal of Computer Science Trends and technology (IJCST), ISSN: 2347-8578, Volume 3 Issue 4, pp: 23-32, Jul-Aug 2015.

[4]   K. Uday Kumar Reddy, Dr. P. Subbaiah, A. Madhavi, "*Performance of Ad hoc On-Demand Distance Vector Routing*", International Journal of Applied Engineering Research, ISSN 0973- 4562, Volume 13, pp: 6414-6418, 2018.

[5]   Anush Baskaran, Sushant Ramesh, Ronnie D. Caytiles and N.Ch.S.N Iyengar, " *ABAODV: An Accumulator Based Ad Hoc on Demand Distance Vector Routing Protocol for Vehicular Ad Hoc Networks*", International Journal of Future Generation Communication and Networking, ISSN: 2233-7857, Vol. 9, pp: 213-228, 2016.

A Comprehensive Study on Various Network Security Attacks with Counter Measures in Network

C. Linda hepsiba, Dr. R. Jemima Priyadarsini, J. Jenifer,

Department of Computer Science,

Bishop Heber College, Trichy-17, Tamil Nadu,

Lindahepsiba.cs@bhc.edu.in, jemimapriyadarsini.cs@bhc.edu.in, jenifer.john3@gmail.com

Abstract

The term "network security" refers to a diverse set of technologies, devices, and processes.. Today's network structure is complicated, and it must contend with an ever-changing threat environment and intruders that are always looking for and exploiting weaknesses. These vulnerabilities can sometimes be found in various locations, including computers, information, mobile applications and users. As a result, several vulnerability management tools and software are now in use to address specific threats and exploits.Various Network Security Attacks and Countermeasures in Network Security is the subject of this extensive study.

Keywords: Network, Security, Attacks, Passive Attacks and Active Attacks

1. Introduction

People are becoming more conscious of the necessity of network security as computernetwork technology advances. As internet technology advances at a faster pace, people are becoming more aware of the importance of network security.The network must havebeen protected against intruders and hackers.Network security refers to the monitoring and control of illegal access, misuse, and unauthorized changes to the networking system.Network intrusions are commonly used by malicious hackers to alter, destroy, or steal private data.Active and passive attacks are the two types of attacks based on network security. In an active attack, an attacker attempts to change the original data's content. In a passive attack, an intruder observes and copies the communications.

This article is incorporated in the following ways:Section II provides related works of Network security attacks. Network work security and Various Network security attacks are disseminated in the section III and IV.The countermeasures for Network security attacks are discussed in the section V. Finally, the conclusion of section VI.

2. Related Work

This section examined existing articles relating to the field of Network Security Attacks based on Network Security. The authors present the Rate Decreasing Algorithm (RDA) for avoiding DOS attacks in [3], and RDA efficiently controls DOS attacks in networks.Snehal et al. present an RSSI mechanism for identifying wormhole attack nodes, which is implemented inContiki OS using the Cooja Simulator [4]. Shukla describes a Machine Learning-based Intrusion Detection System (ML-IDS) approach for detecting wormhole attacks in IoT [5]. Shiranzaei etal. offer an IDS for detecting sinkhole attacks and attack prevention mechanisms [6]. Stephen et al. [7] Deist Mechanism designed to detect sinkhole attacks in a network.The authors of [8] describe various types of sybil attacks and their defensive mechanisms within the Internet of Things.

Sowmya et al. propose a PKI-secured unique session key to indicate the presence of the Sybil attacker [9]. The authors proposed a mechanism for detecting spoofing attacks in aeronautical ad-hoc networks using deep autoencoders [10]. Dwoskin et al. furnish two low-cost hardware-based frameworks to improve key management scheme security against sensor node fabrication attacks for something like a lightweight mobile Ad-Hoc network [11]. In[12], five lightweight feature vectors for network traffic classification and anomaly detection were tested.Vashist et al. deliver an identification and protection mechanism for WiNoCs against jamming-based eavesdropping and DoS attacks originally derived from either an external or an internal attacker by predicting the number of burst errors in packet delivery over wireless interconnects using burst

error correction codes [13].

3. Network Security

The term "network security" refers not only to the protection of a single network, but also to the security of any network or network connection. It is a collection of guidelines and protocols designed to protect the computer network's integrity, availability and confidentiality.

a. Integrity

Integrity security mechanisms keep information safe from illegitimate modification. These metrics improve the credibility and consistency of data.Data must not have been modified in transit to protect the integrity, and actions must always be taken to protect cannot be modified by an unidentified user or program.Required to implement access control policies and version control, for example, can help to prevent fallacious modifications or data leakage by legitimate user.

b. Availability

Availability means that persons in need have prompt and dependable access to information, depending upon the time of the day, geography or other considerations.Some of the most basic vulnerabilities to availability seem to be non-malicious by nature, such as communication bandwidth issues, hardware failures and unexpected software downtime.

c. Confidentiality

Confidentiality Measures to protect data from theft and misuse are in place.It pertains to network security since it necessitates network access to guarded data. Confidentiality necessitates safeguards to ensure that the appropriate individuals have access to the data, and that unauthorized parties are refused entry.

4. Network Security Attacks

During the data transfer process, network security is a major concern. Attackers utilize a variety of techniques to exploit network vulnerabilities and weak spots. Passive and active intrusions are the two basic forms of network security attacks.

A. ACTIVE ATTACKS

It tends to involve compromising a user or network using information gathered during a passive attack. An active attack, as compared to a passive attack, is more likely to be discovered by the target shortly after it is carried out.

a. Spoofing

Spoofing is a technique used by attackers and intruders of all types to impersonate people, businesses, and even computers in order to trick people into providing personal information in exchange for access to something valuable.Spoofing is the act of misrepresenting amessageoutofanunknownsourceasone from aknown, reliableone. Spoofing can be applied to emails, phone calls, and online platforms, or it might be more technical, like as IP spoofing, Address Resolution Protocol (ARP) spoofing, or DNS server spoofing.

b. Modification

Modification attacks require us to tamper with our assets. This kind of attack may be classified as an integrity attack, but they may also be classified as an availability attack.There are three kinds of modifications.such as changing existing information, removing existing information, and inserting new information..The integrity of the original data is jeopardized by modification.

c. Wormhole

In network security, a wormhole attack is a significant network security attack. It is made up of two key components: tunnels and vengeful nodes. More than one attacker node forms a tunnel in a wormhole assault [2]. The other nodes are misled by this tunnel, which generates a misleading path. This is one of the most difficult attacks to detect or prevent, and it also changes the routing behavior.

d. Fabrication

Fabrication attacks, which are primarily an attack on confidentiality, permit unauthorized users to access our information, apps, or contexts. One of the seven broad categories used to classify active attacks and threats is fabrication.Within a system, a fabrication attack generates unauthorized information, communications, processes, or other data.

e. Denial of Service

A denial of service (DoS) assault is a Network attack in which the attacker attempts to deplete a legitimate user's resources or bandwidth. A DoS attack is an attack that overloads an online service with traffic from dozens or hundreds of devices in a short period of time. Intruders generate massive amounts of traffic that a server or network cannot handle, preventing regular users from accessing the server or network.

f. Sinkhole

A sink node is an attacker node. The attacker's main goal in this attack is to generate a high volume of traffic out of this specific node [2]. Sinkhole attack is one of most dangerous network assaults; it is not detrimental when used alone, but it has a significant impact when combined with other attacks such as wormhole attacks.

g. Sybil Attack

A Sybil attack is when a node pretends to be many people. Sybil nodes are the names given to these false identities. Sybil attackers can make up a large percentage of network identities [8]. In networks, there are three types of Sybil assaults. Directly and Indirectly Communication, Stolen and Fabricated Identities, and Non-Simultaneous & Simultaneous are just a few examples.

B. Passive Attacks

An attacker conducts a passive attack by monitoring a network or systemcommunications and scanning for ports and services and other security flaws.Once the attacker has gained network access, they can acquire information in many different ways, including forcing the user to enter their credentials and recording them for future use.

a. Traffic Analysis

Traffic analysis is one of the network threats. Even if a packet is encrypted, an attacker can deduce its content by analyzing network traffic patterns.In this category, an attacker tracks channels of communication to collect various information, including people and machines identities, locations of such identities.

b. Eavesdropping

When an attacker intercepts, deletes, or changes data sent between two devices, it is called an eavesdropping assault. Eavesdropping, also referred to as snooping or sniffing, relies on unencrypted network interactions. The data is sent across an open platform, giving an intruder the opportunity to attack a defect and intercept it using a variety of methods. Eavesdropping assaults are notoriously difficult to detect.

5. Countermeasures for Network Securityattacks

There are several existing solutions to network security attacks, including active and passive attacks. Encryption and Authentication, Router filtering, Regenerating of Session ID after a Successful Login,String as Session key,Cross-Site Scripting, Port security and Various IDS mechanisms are used against active and passive network security attacks.

Table 1. Counter Measures for Network Security Attacks

Network Security Attacks	Active Attacks						Passive Attacks	
	Spoofing	Worm hole	Fabri cation	Denial of service	Sinkhole	Sybil Attack	Traffic Analysis	Eavesdroppi ng
[3]				✓				
[4]		✓						
[5]		✓						
[6]					✓			
[7]					✓			
[8]						✓		
[9]						✓		
[10]	✓							
[11]			✓					
[12]							✓	
[13]				✓				✓

6. Conclusion

The process of preventing and detecting malicious attackers on a network is known as network security. These attacks result in data security issues such as data loss. This paper attempts to define network security and demonstrates it through various network security attacks and countermeasures. The precautionary measures are not yet complete. In the future, countermeasures must improve against all sorts of attacks and ensure better system prevention.

References

[1]   Pawar, Mohan V.,and J.Anuradha."Networksecurityandtypesofattacksinnetwork.", *Procedia Computer Science* 48 (2015): 503-506.

[2]   Hepsiba, C. Linda, R. Jemima Priyadarsini, and S. Titus. "A Comprehensive Study on Routing Attacks with Countermeasures in Internet of Things." *Solid State Technology* 63.4 (2020): 7993-

7999.

[3]   Malla, Adil Mudasir, and Ravi Kant Sahu. "Security attacks with an effective solution for dos attacks in VANET." *International Journal of Computer Applications* 66.22(2013).

[4]   Deshmukh-Bhosale, Snehal, and Santosh S. Sonavane. "A Real-Time Intrusion Detection System for Wormhole Attack in the RPL based Internet of Things." Procedia Manufacturing 32 (2019): 840-847.

[5]   Shukla, Prachi. "Ml-ids: A machine learning approach to detect wormhole attacks in internet of things." In 2017 Intelligent Systems Conference (IntelliSys), pp. 234-240. IEEE, 2017.

[6]   Shiranzaei, Atena, and Rafiqul Zaman Khan. "An Approach to Discover the Sinkhole and Selective Forwarding Attack in IoT."(2018).

[7]   Stephen, Raju, A. Dalvin Vinoth Kumar, and L. Arockiam. "Deist: dynamic detection of Sinkhole attack for Internet of Things." International Journal of Engineering And ComputerScience 5, no. 12 (2016): 19358-19362.

[8]   Suresh, Mr S. Alexander, and R. Jemima Priyadarsini. "A Comprehensive Study on Sybil Attacks and Its Defence Mechanisms in Internet of Things." *Solid State Technology* 63.4 (2020): 7966-7974.

[9]   Sowmya, P., and V. Anitha. "Defence Mechanism for Sybil Attacks in MANETS using ABR Protocol." International Journal of Advanced Computer Research 4.2 (2014):494.

[10]   Hoang, Tiep M., et al. "Detection of Spoofing Attacks in Aeronautical Ad-Hoc Networks Using Deep Autoencoders." *IEEE Transactions on Information Forensics and Security* 17 (2022): 1010-1023.

[11]   Dwoskin, Jeffrey S., et al. "Secure Key Management Architecture Against Sensor-Node Fabrication Attacks." *GLOBECOM*.2007.

[12]   Meghdouri, Fares, Tanja Zseby, and Félix Iglesias. "Analysis of lightweight feature vectors for attack detection in network traffic." *Applied Sciences* 8.11 (2018):2196.

[13]   Vashist,Abhishek, et al. "Securing a wireless network-on-chip against jamming-based denial-of-service and eavesdropping attacks." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 27.12 (2019): 2781-2791.

A Novel Framework for Blockchain Parallel Mining using Proof-of-Work Mechanism

K. Lino Fathima ChinnaRani and M.P. Anuradha

Department of Computer Science, Bishop Heber College,
Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.
linofathima.ca@bhc.edu.in, mpanuradha.cs@bhc.edu.in

Abstract

The prompt and furious study on enamours Blockchain transactions is required for several applications like tracking deceitful money transactions. Here, the transactions are uninterruptedly rising and therefore it is arranged as series of transaction Blocks. This setup requires efficient decentralized data structures. The proposed Parallel Proof of Luck algorithm, the traditional setup is converted as a scalable cluster-based architecture that builds a parallel distributed transaction graph as well as deploy a several transaction analysis algorithms. In contrast to tradition transaction system, Blockchain is stand apart with unique features such as transparency and immutable data structures. The proposed architecture is framed to permit multiple miners to authenticate the transaction Blocks, ultimately it raises the throughput by the terms of transaction per second. The participation of more miners can eventually enlarge the availability of the network and prevent the bottleneck issues. This work strengthens with fair share technique along with parallel mining algorithm to transfer the blockchain network into parallel miners operating system, therefore several miners could be able to contribute their processing computational resources. It leads to the reduction of transaction confirmation time for the users.

Keywords: *Parallel Mining, Proof-of-Luck, Throughput, TPS, Scalability.*

1   Introduction

This offers a distributed authenticated record which comprises the confirmed transactional history. It also provides a public disseminated network, Here the clients have the provisions to exchange the authenticated transactional records among themselves without integrate on the third-party Association [2]. Furthermore, Blockchain maintains the privacy amid of all the transactional information and maintains the replica of constantly growing ledger to all the client of the system in the network. Nevertheless, when compare to other systems, Blockchain offers many applications, yet the major issue is scalability. There are many cryptocurrencies on the market presently implement the Blockchain network for mining the transactions and maintaining the distributed ledgers among the network [1].

The current issue is scalability problem. The TPS (Transactions Per Second) speed can be varied in diverse Digital Coins because of their corresponding etiquettes. For Illustration, Proof-of-Work consensus mechanism can be applied in most of the Bitcoins while Ethereum implements the Proof-of-Stake mechanism [3].

Cryptocurrency	Transaction per Second	Average Transaction Confirmation Time
BTC	4	60 Mins
Ethereum	15	6 Mins
FANTOM	300000	1 Min
IoTA	500-800	2 Mins
Dash	10 - 28	15 Mins
Stellar	1000	2-5 S
LiteCoin	56	30 Mins
Ripple	1500	4 sec
XRP	1500	1 min

There are some existence models have been used to renovating the Blockchain computing into Parallel execution through the alteration of consensus Mechanism of Distributed Ledger. For Illustration: Foldingcoin

generated from Stanford University and GridCoin from UC Berkeley. Here, Foldingcoin has formed on the basis of "Proof-of-Fold" and GridCoin originated through "Proof-of-Research" to validate the transactions and also promote executional power of each contributed Blockchain Node [4].

In traditional Blockchain, Participants share their executional resources to the Scientific research on the Blockchain Network. Yet, Both FoldingCoin and GridCoin are not well-suited with Current Commercial Blockchain System so that they required to be replaced with present Blockchain Consensus Protocol. These two different decentralized Consensus Mechanism employed on various Consensus executional tasks on the individual node which are accomplished in a parallel Mining. Although, only a limited set of Computational tasks can be activated as a Consensus Execution. Therefore, the Applications of this technique is very partial for distributed and Parallel Computing [5].

2   Literature Survey

Chencheng Zhou et.al [1] have focused on the dependency modelling architecture and the investigation of Bitcoin. They have also examined the selfish miner processing based on the various significant parameters like selfish miners processing power, and miner's interrupts regaining competence through results. They stated the various chances and the Bit-coin dependency by Continuous -time Markov Chain based Approach.

Shihab Shahriar Hazari et al [2] have introduced the Parallel proof of Work, here, each miner in the Blockchain network can together crack the puzzles through participate in the competitions. They have given the assurance that the basic aspects of the Blockchain will not be disturbed. They have showed the primary outcomes through the raises of Scalability of PoW up to 34% related to the existing Systems.

Zonyin Shae etal.,[3] have investigated distributed parallel computing architecture by renovating traditional Blockchain Computing. The outcomes derived by decentralized parallel Computing architecture have a feature that fasten the data validation distributed management. Distributed data sharing, Distributed execution analytics and studies are the basic processing of the proposed architecture. They also explained the various applications, methods and use cases of this paradigm.

Shrey Baheti et al [4] have proposed a decentralized Miner-Follower Framework DiPETrans to examine the Block of transactions parallelly among more than one system on the same Blockchain Network. They have also tested protype model which there is actual transactions are take out from Ethereum Blockchain. They have also achieved in relational expansions based on the numbers of followers' performance. They have evaluated DiPETrans among the several workloads, here, the number of transactions varies from 100 to 500 in a Block   with several Contract to monetary Transactions.

Xuechao Wang et al [5] have presented a common mechanism which describes that any Parallel Chain Protocols can be transferred from fixed difficulty to the variable difficulty architecture. They have also Proved that the proposed settings use the new Proof methodology which examines the interconnection between the pivot and Non-pivot Chains. They Proposed the Novel Protocol called Ledger-Combiners which implements Parallel-Chains for strongly Coupling More than one -Ledgers added with attaining Low Latency.

Baran Kılıç et al [6] have focused on the Blockchain Ethereum Mainet which implements the execution of Smart Contracts. Even though Blockchain throughputs are increased, enormous transactional data have been gathered. All the above said improvements give the signal that scalable and Parallel system computing is required that can examine the massive Blockchain graph transactional data.

3   Proposed Work

When implement the Proof-of-Work, most of the important data of miners are alike, some of them are Bitcoin Index, Previous Block hash value, and timestamp but the transaction data and the nonce value selected by the miners may vary, the proposed method constructed based on the fact that all the miners might use the similar transactional data yet a diverse nonce value. This method describes the concept that all the miners might use the similar transactional content excluding the nonce for some of the blocks, therefore, the proposed method gives the assurance that no more than one miner is going to do the same work [7].

To afford such a framework, a leader is needed to confirm that not multiple miners implement the similar nonce value and that all the miners in the network use the similar transactional data. The Leader, who will be selected from the miners will be changed at every period. here, the period means the time intermission between two Blocks. Significantly, the miner will select the nonce value to compute for validation but not the Leader [8]. The leader will take the responsibility that he has to generate and provide the transactional hash for a certain Block along with nonce value for the miners. At the same time, in contrast to the nonce, the transactional hash value should be identical for all the miners.
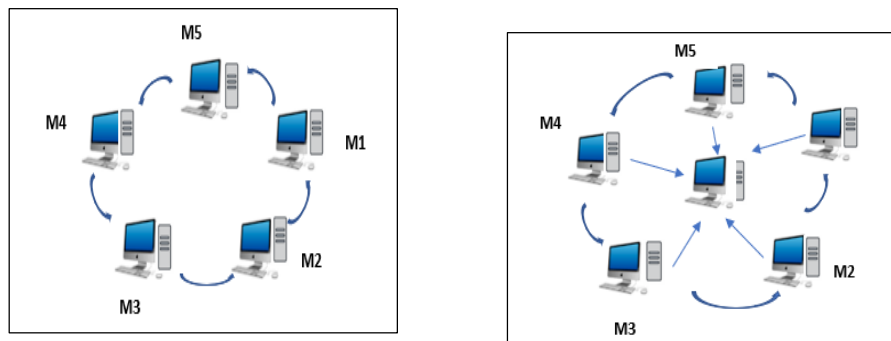

Figure 1 depicts the transformation between earlier systems and Figure 2 explains the proposed solution on the basis of ring Structure. In the Conventional Model, all the systems are linked to other directly or through other system. In the proposed system, they will be linked with each other and also directly associated with the manager. At the Initial stage, there must be a Genesis block present at the flinch of the Blockchain with no computational transactions. Subsequently, the process starts for selecting a miner randomly as a leader, he is the one who is going to be solve the next block (Block 1), and also for the remaining blocks.

All the miners in the network are now try to solve the genesis Block like traditional methodology. While the first Block (genesis Block) is cracked by the miner, subsequently, the time to crack the next block will start. The proposed method activated at this period of time.

Distributed Data

At the Beginning, the leader will generate a computational hash value along with un authorized transactions. Simultaneously, he will create the nonce values for miner groups. The range of nonce value is set to all the groups in the network. There is no identical nonce values should be assigned in cluster of groups [9]. Generally, if 'n' number of miners actively participated in the network, the leader should create and register minimum 'n' number of groups. Therefore, each miner in the Blockchain got the transactional hash along with sets of nonces by the Leader.

The Blockchain system confirms that no two miners in the network along the similar group. Except the Leader, all the Leaders are trying to find the answer for next Block embedded with transactional data and the nonce values assigned to each of the miner. On the other flip of the side, the leader will create and acknowledge all the groups of nonces in the network. In this Process, when the miner has used all the allocated Nonce values, then he will request the Leader for a new nonce range value. the Leader will give the idle nonce range to the miner.

In the Mid of this Process, if a new miner enters into the network, then the Leader acknowledge him and provide him a required data and the Leader allocates him to a new group of nonce values. For this Purpose, the Leader must create as much as group of nonce values as possible. This executional process will remain till a solution for the existing nonce is found.
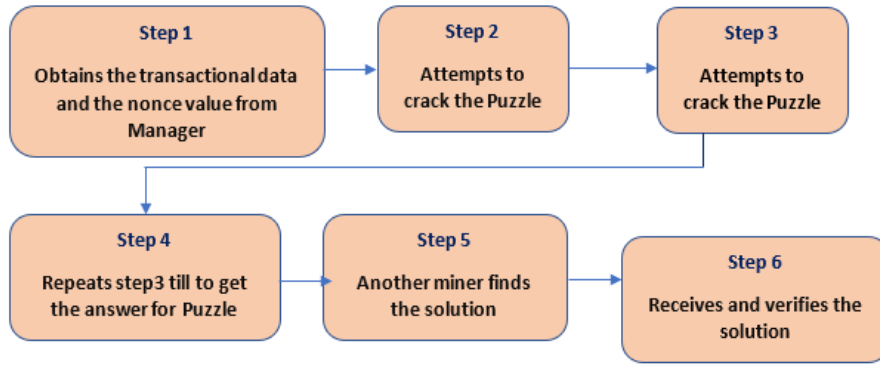
Fig 3. Work Flow of a Miner

4   Implementation and Evaluation of Results

To execute this approach in real time, a Docker has been used. Docker given a Linux based setup with its own Integrated Development Environment. A separate network is created for Docker where all the systems will be linked. The execution has been achieving in Ubuntu Operating System with Core i5-5200U CPU 2.2 GHz, RAM has 4.00 GB. All the miners in the network are assigned with 10% of total resources for ensuring each miner has equivalent in processing power. The alternative analogous framework has established by using the similar resources and components in order to compare the results with an existing system [10].

The test has been steered on the various numbers of systems for both single and Parallel mining by means of diverse difficulty Levels. The difficulty level is described by the least number of zeros begins at the hash values.
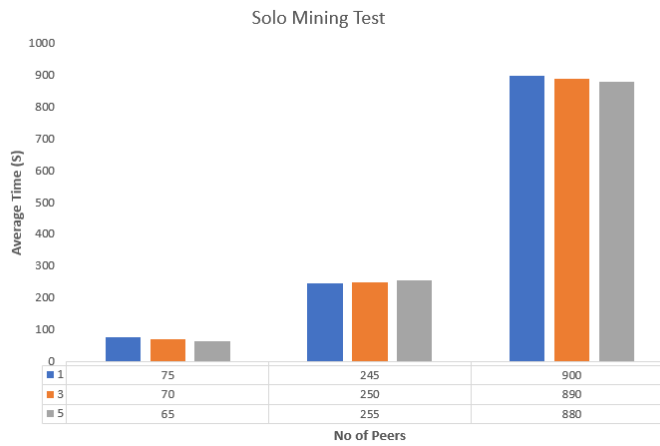
Fig 5. Solo Mining Test Results
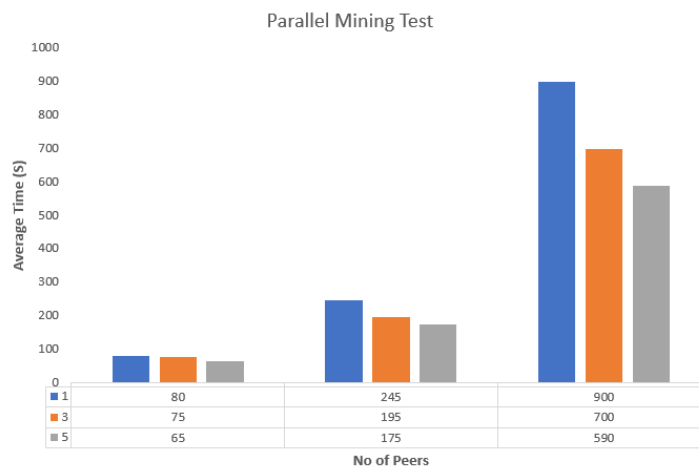
Fig 6. Parallel Mining Test Results

Figs 5 and 6 describe the test results generated on solo and Parallel Mining. Here, Transaction Per Second TPS means the average time taken to crack the Block in seconds. This can be computed after executing various tests based on the same conditions, at the end computing the average of all the results. The Index value, timestamp, transactional hash, Previous Hash and nonce value are reserved as a input to get the solution. All the miners in the networks have the identical timestamp and previous hash values for certain Blocks, In Contrast, Parallel Mining also, the transactional hash value is same for all the Miners.

5   Conclusion

The framework of the Blockchain or Proof of work algorithm required to change for scaling the Blockchain systems. The proposed solution can present the Parallel-Proof-Work mechanism among each miner in the Blockchain network combinedly crack the puzzle by means of participate in the competition. The proposed solution cannot disturb the core characteristics of Blockchain Technology. The Framework must always distribute, Peer to Peer and Protected architecture. Therefore, the execution Framework should stick on to the all the procedures described by Bitcoin.

In future, the same work will be implemented in the Proof-of Luck with Fair share concept, which is another substitute of Bitcoin Blockchain developers. Significantly, the proposed method can be estimate on the transaction allotment and the energy consumption. Therefore, It enhances the Transaction Per Second (TPS) value, the proposed solution will absorb ultimately less power per Block authentication when compared to the existing system.

References

[1] Zhou, C., Xing, L., Guo, J., & Liu, Q. (2022). Bitcoin Selfish Mining Modelling and Dependability Analysis. International Journal of Mathematical, Engineering, and Management Sciences, 72(1).

[2] Hazari, S. S., & Mahmoud, Q. H. (2019, January). A parallel proof of work to improve transaction speed and scalability in blockchain systems. In 2019 IEEE 9th annual computing and communication workshop and conference (CCWC) (pp. 0916-0921). IEEE.

[3] Shae, Z., & Tsai, J. (2018, July). Transform blockchain into distributed parallel computing architecture for precision medicine. In 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS) (pp. 1290-1299). IEEE.

[4] Baheti, S., Anjana, P. S., Peri, S., &Simmhan, Y. (2022). DiPETrans: A framework for distributed parallel execution of transactions of blocks in blockchains. Concurrency and Computation: Practice and Experience, 34(10), e6804.

[5] Wang, X., Muppirala, V. V., Yang, L., Kannan, S., & Viswanath, P. (2021, November). Securing parallel-chain protocols under variable mining power. In Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (pp. 1700-1721).

[6] Kılıç, B., Özturan, C., & Sen, A. (2022). Parallel analysis of Ethereum blockchain transaction data using cluster computing. Cluster Computing, 1-14.

[7] Popov, Serguei. "The tangle." http://www.tangleblog.com/wpcontent/uploads/2016/11/IOTA_Whitepaper.pdf. April 3, 2016. Version 0.6.

[8] Le Lann, Gérard. "Distributed Systems-Towards a Formal Approach." IFIP congress. Vol. 7. 1977.

[9] Dwork, Cynthia, Nancy Lynch, and Larry Stockmeyer. "Consensus in the presence of partial synchrony." Journal of the ACM (JACM) 35.2 (1988): 288-323.

[10] Boyen, Xavier, Christopher Carr, and Thomas Haines. "Blockchain-Free Cryptocurrencies. A Rational Framework for Truly Decentralised Fast Transactions." IACR Cryptology ePrint Archive 2016 (2016): 871.

[11] Clark, Joseph Bonneau Andrew Miller Jeremy, Arvind Narayanan Joshua A. Kroll Edward, and W. Felten, "Research perspectives and challenges for bitcoin and cryptocurrencies." Security and Privacy (SP), IEEE Symposium on. IEEE, 2015.

Intrusion Detection Based on Generalized Variable Precision Rough Sets Feature Selection Techniques

R. Rajeshwari and M.P. Anuradha

Department of Computer Science, Bishop Heber College,
Affiliated to Bharathidasan University, Trichy-17, Tamilnadu.India.
rajeshwari.ca@bhc.edu.in, mpanuradha.cs@bhc.edu.in

Abstract. Intrusion detection systems (IDS) are becoming particularly crucial as the world's communication moves over the internet. For many years, network intrusion detection becomes a substantial research area. An intrusion detection system identifies the anomaly of threats or intrusion activities. The growing attack size and significant improvements in machine training are a major research topic and a good solution for cyber security in cooperation between the intrusion detection system and machine learning. Massive amounts of test data are typically used in the training phase. A large amount of information has a negative impact on the training and classification of attacks in machine learning techniques. Effective method for selecting optimal features in order to eliminate insufficient and unwanted data set features. In this paper a new feature selection model based on feature selection Generalized Variable precision Rough Sets techniques are proposed for selecting an optimal feature. Using the Support Vector Machine classifier to determine output after applying feature selection. NSL-KDD data set is used to investigate important characteristics for the detection of network intrusions with data mining and machine training techniques. The output of the experiments exhibited that the model proposed improves detection rates and also speeds the detection process.

Keywords: *Intrusion detection System, Generalized Variable precision Rough Sets, Support Vector Machine.*

1   Introduction

Intrusion detection systems (IDS) become vital and commonly used to ensure network security. The functionality has complicated relationships, so the amount of data to be processed must be reduced. Data filtering, clustering, and selection of functions can result in reductions. Attributes comprises different types of wrong correlations that impede the detection process in complex classification domains. Additional features can increase the calculation time and can affect IDS accuracy. By searching for the subset of features that best classify the training data, selection improves classification [1]. The literature contains some works that employees develop IDS using various fuzzy logic techniques, machine learning methods and intelligent system. [2][3]. Authors of [4] have exposed that many features are not important and can be eliminated without reducing the presentation of the IDS significantly. In a machine learning context, IDS tasks are often modelled as a classification problem.

Feature selection is a technique of machine learning that reduces the amount of data that needs to be analyzed. Key features (or attributes) of a data set can be identified. Machine learning algorithms can improve the effectiveness of classification predictions by reducing the size of the data set that includes only the most important functions. This efficiency is especially relevant to Intrusion Detection which has demanded real-time performance. According to Nassar et al. [5], Even 1 Gbps of sustained packet filtering traffic poses big data issues for IDS with With Network Monitoring. This effectiveness is particularly important for intrusion detection that has real-time performance requirements.

When selecting a feature, the machine learning methods predictive accuracy is not significantly reduced. As discussed in the work, classification performance improves with fewer estimated features in the data set. Feature selection from network traffic record is currently a significant challenge for intrusion detection, and researchers attempted to identify dynamic feature selection techniques using classification problems.

The amount of correct classification of an occurrence from a dataset with the correct class label by a "classifier" is referred to as classification performance. The records in the intrusion detection data sets are separated into two categories: Attack for malicious attack and Normal for non-malicious incidents.

This paper aims to assess the effect on the NSL-KDD data package for classifying intrusions of future selection. The Generalized Variable Precision Rough Set (GVPRS) techniques is used to select an optimal feature selection. The SVM classifier classify the different types of attacks.

2   Related Works

In the [7] focused on essential essential attributes to create an efficient IDS in this work. To select a best feature in the dataset using majority vote method (correlation based, information gain and chi-square) and decision tree classifier is used classify the attributes. The results indicate that using reduced attributes from the novel feature selection method to construct a computationally efficient IDS system performs better.

In the [8] proposed two models to detect and classify intrusions TIDCS (Trust-based Intrusion Detection and Classification System) and TIDCS-A Accelerated) are safe network intrusion detection and classification systems.

To evaluate these high dimensional data,[9] used the attribute correlation approach. In the proposed work proses a correlation cut-off-value methodto choose an optimal attributes and classifier obtain a good result. The result show cut-off-value in recursive feature is 0.2, it acquires (99.26 %) precision. The selection function can reduce the running system time.

In the [10] proposed system uses KDD CUP 99 dataset. This system used four classifier models, like LDA, SVM, Random Forest, and Adaboast, which give an Adaboost sensitivity of 99.75% and a specialty of 95.69% which, in comparison with a different classifier, is higher. This system can also detect unknown future attacks.

3   NSL-KDD Dataset

In this work, a high dimensional NSL KDD dataset used with 41 attributes [11] has been used.22 attacks contained in the training phase and 17 attacks are contained the testing phase.Following different types of attacks are

Denial of Service (DoS)-Suspicious activity aimed at halting the operation of a system or networks services and support.

Remote to Local (R2L)-In the network packet, there is an illegal access.

Probe- acquire knowledge on security holes and subsequently attack those systems

User to Root (U2R)-exploited the administrative leverage [12]

In the Table [1] exhibit NSL-KDD dataset datatypes and their group.

4   Proposed Architecture for IDS

The NSL-KDD high dimensional dataset is used in this work. The Generalized Variable Precision Rough Set (GVPRS) method is used to choose an optimal feature in a high dimensional dataset. It is a computer-based feature selection process [13].$\boldsymbol{\beta}$ is defined in classification range from $0.0 \le \beta \le 0.5$. The GVPRS method consists of following steps. In the Fig 1 shows the proposed work IDS architecture

Steps:

1. Selecting Precision Value (β)

2. Find the complete set of (β) reduct.

3. Remove Duplicate Item.

4. Eliminate no longer attributes.

5. Follows the rules.

Support Vector Machine (SVM) widely used classification algorithm [14]. It is used to select an anomaly detected features in the network. The Sigmoid Kernel (SK) is used in this work.

Table 1. NSL KDD dataset attributes datatypes and group.

Group	Attack Name	Attack Type	Group	Attack Name	Attack Type
Symbolic	Duration	Numeric	**Content**	count	Numeric
	Src_Bytes			is_guest_login	Binary
	Des_Bytes			serror_rate	Numeric
	Urgent			srv_count	
	Wrong Fragment			rerror_rate	
	Types of the protocol	Nominal		srv_rerror_rate	
	Service			diff_srv_rate	
	Flag			same_srv_rate	
	Land	Binary		srv_diff_host_rate	
Content	Count_Failed_Login	Numeric	**Traffic**	serror_rate	Numeric
	Logged_in			dst_host_same_srv_rate	
	Hot			dst_host_srv_count	
	Comprised_Number			dst_host_same_src_port_rate	
	root_shell			dst_host_differ_srv_rate	
	root_no			dst_host_srv_differ_rate	
	attempted_su			dst_host_srv_serror_rate	
	shell_no			dst_host_srv_rerror_rate	
	file_no_creation			dst_host_rerror_rate	
	access_file_no			class_	
	outbound_cmd_no			dst_host_same_srv_rate	
	Host_login	Binary		dst_host_srv_count	

$SK = tan\_h(\alpha\ x^t x_i) + \beta$ Where α represents growing parameter and β represents varying parameter.
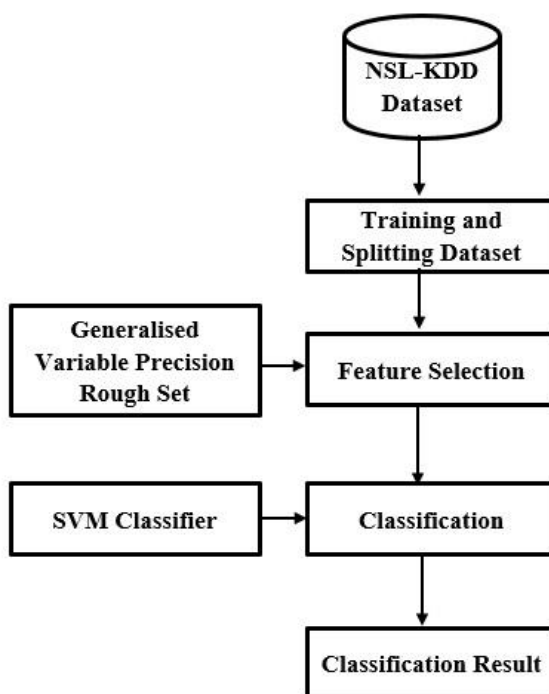
Fig 1. Proposed GVPRS IDS

5　　Evaluation Metrics

Despite the confusion matrixes representational power in its classification, it is not a very useful tool for comparing NIDS.The confusion matrix variables defines numerous performance indicators to overcome this problem. These measures provide quantitatively comparable values, which are briefly explained in the following paragraphs.

True Positive (TP): Properly identified the attacks.

True Negative (TF): Properly identified the normal data instances.

False Positive (FP): Wrongly identified the attacks [15].

False Negative (FN): Wrongly identified the normal data instances.

Precision (Pre): The proportion of threats appropriately predicted to all the threats

$$Pre = \frac{TP}{TP + FP}$$

Recall (Rec): The proportion of threats appropriately predicted to all the incorrectly predicted threats.

$$Rec = \frac{TP}{TP + fN}$$

F-Measures (FM): P and R defines the Harmonic Mean.

$$FM = \frac{2 * Pre * Rec}{Pre + Rec}$$

6　　Performance Evaluation

In the NSL-KDD dataset, five various attacks are DoS, R2L, U2R and Probes. The training and the testing records are 25192 and 21136 respectively. To identify 23 optimum attributes in the NSL-KDD dataset using the GVPRS features selection approach. The SVM Classifier is used to categorize the various kinds of attacks in the dataset. The accuracy estimates for the DoS and Probe class attack remained good. In the Table [2] describe evaluation for accuracy metrics. The accuracy metrics for the U2R class need to be improved. In the Fig 2 exhibit overall demonstration of the proposed work GVPRS.

Table 2. Evaluation for Accuracy Metrics

Accuracy Metrics	Dos	R2L	U2R	Probe
Precision	95.4	40.8	40.1	89.8
Recall	96.5	85.9	38.7	83.2
F-Measures	93	50.6	43.1	75.9
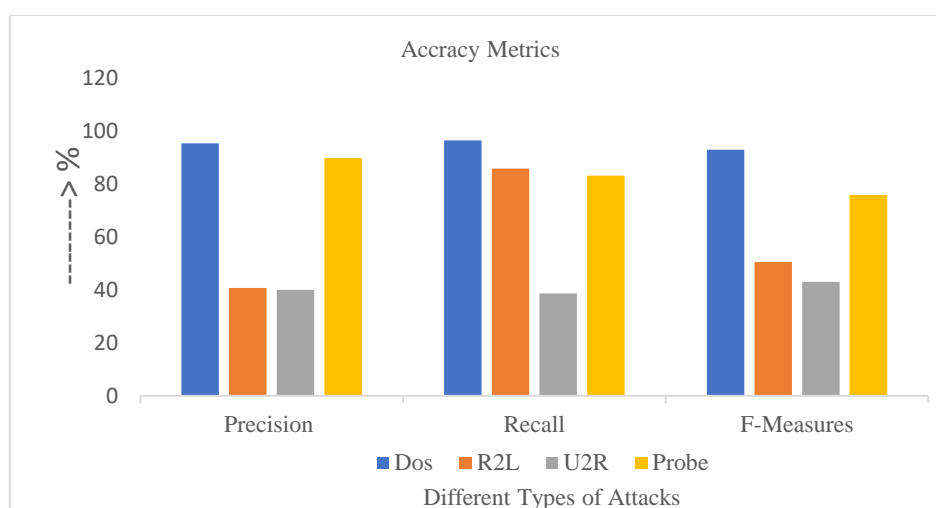
Fig 2. Overall Demonstration of the proposed work GVPRS.

7   Conclusion

A new method GVPRS is used to reduce the anomaly identify features in the NSL-KDD dataset. A SVM Classifier is used to classify the different types of attacks in intrusion detection. The number of features reduced from 23 from 42 in the NSL-KDD dataset. The proposed method GVPRS for intrusion detection is precise and well-organized in a categorize of attack class. The GVPRS techniques outperformed with DoS and Probe attacks. The overall detection metrics have been reduced as a result of insufficient data for R2L and U2R attacks. Future research will focus on expanding a more reliable classification algorithm for R2L and U2R attacks, as well as selecting a hybrid features technique.

References

[1] Foroushani, Zohreh Abtahi, and Yue Li. "Intrusion detection system by using a hybrid algorithm of data mining technique." In Proceedings of the 2018 7th International Conference on Software and Computer Applications, pp. 119-123. 2018.

[2] Lee, Wenke, Salvatore J. Stolfo, and Kui W. Mok. "A data mining framework for building intrusion detection models." In Proceedings of the 1999 IEEE Symposium on Security and Privacy (Cat. No. 99CB36344), pp. 120-132. IEEE, 1999.

[3] Riyaz, B., and Sannasi Ganapathy. "A deep learning approach for effective intrusion detection in wireless networks using CNN." Soft Computing (2020): 1-14.

[4] Zhang, Yuening, Yiming Zhang, Nan Zhang, and Mingzhong Xiao. "A network intrusion detection method based on deep learning with higher accuracy." Procedia Computer Science 174 (2020): 50-54.

[5] Ziarko, Wojciech. "Variable precision rough set model." Journal of computer and system sciences 46, no. 1 (1993): 39-59.

[6] An, Aijun, Ning Shan, Christine Chan, Nick Cercone, and Wojciech Ziarko. "Discovering rules for water demand prediction: an enhanced rough-set approach." Engineering Applications of Artificial Intelligence 9, no. 6 (1996): 645-653.

[7] Ahmadi, S. Sareh, Sherif Rashad, and Heba Elgazzar. "Efficient Feature Selection for Intrusion Detection Systems." In 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 1029-1034. IEEE, 2019.

[8] Chkirbene, Zina, AimanErbad, RidhaHamila, Amr Mohamed, Mohsen Guizani, and Mounir Hamdi. "TIDCS: A dynamic intrusion detection and classification system-based feature selection." IEEE Access 8 (2020): 95864-95877.

[9] Sugianela, Yuna, and Tohari Ahmad. "Pearson Correlation Attribute Evaluation-based Feature Selection for Intrusion Detection System." In 2020 International Conference on Smart Technology and Applications (ICoSTA), pp. 1-5. IEEE, 2020.

[10] WS, Jenif D. Souza, B. Parvathavarthini, and S. Arunmozhi. "Machine Learning based Intrusion Detection Framework using Recursive Feature Elimination Method." In 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1-4. IEEE, 2020.

[11] Kumar, Vikash, Ditipriya Sinha, Ayan Kumar Das, Subhash Chandra Pandey, and Radha Tamal Goswami. "An integrated rule-based intrusion detection system: analysis on the UNSW-NB15 data set and the real-time online dataset." Cluster Computing 23, no. 2 (2020): 1397-1418.

[12] Chen, Ke, Feng-Yu Zhou, and Xian-Feng Yuan. "Hybrid particle swarm optimization with a spiral-shaped mechanism for feature selection." Expert Systems with Applications 128 (2019): 140-156.

[13] Duarte, Humberto Martins Mendonça, and Rafael Lima de Carvalho. "Hybrid particle swarm optimization with a spiral-shaped mechanism for solving high-dimension problems." Academic Journal on Computing, Engineering and Applied Mathematics 1, no. 1 (2020): 1-6.

[14] Shone, Nathan, Tran Nguyen Ngoc, Vu DinhPhai, and Qi Shi. "A deep learning approach to network intrusion detection." IEEE transactions on emerging topics in computational intelligence 2, no. 1 (2018): 41-50.

[15] Karatas, Gozde, and OzgurKoraySahingoz. "Neural network-based intrusion detection systems with different training functions." In 2018 6th International Symposium on Digital Forensic and Security (ISDFS), pp. 1-6. IEEE, 2018.

.

Digital Platform for Chinese Ancient Arts Based on Blockchain and Sentimental Analysis

Sheng Dong Zhou [1],Siva ShankarRamasamy*[2], Bibi She*[3]

[1,2,3]International College of Digital Innovation, Chiang Mai University, Chiang Mai-50200, Thailand.
shendong_zhou@cmu.ac.th, sivashankar@cmu.ac.th, bibi_she@cmu.ac.th

Abstract. Ancient Artwork has not been adequately researched. It is only essential for future researchers interested in the same topic to have a reference point to help ease their research challenges. After the pandemic, all the countries are searching for their history and saving their arts and culture because of their cultural proof. For that time, every art is important, like art, dress, dance, painting, statue, song, etc. because it reflects the language of their country. At present they are searching for roots that have roots from their history, hence blockchain will be useful for searching roots for artworks. Blockchain provided evidence for digital transactions and will aid in the preservation of history through digital transactions in the future. With the support of blockchain, ancient artworks are still relatively new concepts in the business world, this study will assist the reader to understand where the technology came from and how it has progressed over time.

Keywords: Ancient, Artwork, Blockchain, Digital transactions, Business.

1   Introduction

The world is changing at a faster pace, but culture of each country remains same because of its own arts and culture. Every art and culture have its own country's roots. Art and culture can be preserved with the help of new technologies [1]. Artists who have yet to accept technology will undoubtedly become a new phenomenon. The primary goal of this research is to learn how copyrights can be implemented in the arts and how they can affect the industry's economic and social status [2]. Traditional art buying and selling exposed artists to security risks such as copying, imitating, theft, and destruction by fire, water, and other natural elements; however, blockchain technology has transformed how business is conducted by converting the copyright of any work or specific Art into digital traces [3]. Hence, Blockchain is definitely a cutting-edge technology that has the potential to transform standard business operations.

2   Literature Review

W.Yu and S Huang advocated [4] using RFID and geographical data to think about the traceability of artwork goods. The blockchain technology concept, which scans the product's traceable QR code with smart devices and obtains data information from all transaction nodes of the goods, is equally relevant to artworks. In terms of platform architecture, this paper used K. Balasubramanyam's paper "Adaptation of Blockchain Technology in Manufacturing Industry," which was presented at the 2020 International Conference on Mainstream Blockchain (ICOMBI) as Blockchain Technology for the New Internet [5]. Through the validity and integrity of data, this novel notion can assure transparent transactions. And the concept of such a system is widely regarded as the most secure, traceable, and decentralized method of transaction. Simultaneously, S.P. Gayathri, S. Vijayalakshmi, and Siva Shankar Ramasamy [6] stated in 2021 that blockchain technology can eliminate electronic tracking by supplying transaction keys, allowing for a reliable project or data tracking or monitoring. Each transaction has a vast amount of data stored

in several links or nodes. Even if only one or two copies of the data remain on the network, the digital ledger will have other copies as proof. If any manufacturing organization implements the blockchain concept in its IoT devices, consumers and companies should take it to a new level of quality and consistency. It also promptedto contemplate a citation evaluation system to improve the quality of consumer and art evaluations.

In the artwork circulation link as a product,the packagingis one of the major issues in distribution networks, whose productivity is harmed by a significant amount of human labor, according to P. Kuhlang and A. Sunk [7]. The value stream design and methodology - time measurement (MTM) approach has been demonstrated in practice to increase work and hence productivity, particularly in manual work systems. YK Ro, JK Liker, and SK Fixson [8]also presented a more gradual technique employed by Toyota in the supply chain system for comparison, because it incorporates modularity and moves to a build-to-order model. J. Block, T. Tyrberg, and Yuan Fuqing [9] also suggested a method for determining the best time to stop repairing repairable devices in order to save money on repairs. It predicts the number of units available using a Poisson distribution based on the NHPP (Non-Homogeneous Poisson Process) while meeting spare parts needs for the remaining systems. To formulate the problem, a notion called minimum margin is established, and non-linear programming is proposed to find the best solution. Finally, the method is demonstrated with a numerical example.

While D. Sathya, S. Nithyaroopa, D. Jagadeesan, and IJ Jacob[10]proposed the use of blockchain technology and the decentralization of supply chains to improve security, immutability, and tamper resistance, the art transaction itself has the chaos of being counterfeited, special materials, and different genres, while D. Sathya, S. Nithyaroopa, D. Jagadeesan, and IJ The suggested work employs Ethereum smart contracts for supply chain management systems, preventing data falsification, database corruption, and external threats. SW Sheng and S. Wicha[11] suggested a traceability technique to determine teak validity in the teak supply chain using distributed ledger technology due to some wood art (DLT). The fundamental properties of DLT, also known as Blockchain Technology (BCT), are the preservation of dissemination archives, digital encryption, traceable data, and non-absolute block materials, which are appropriate for the teak supply chain and can give remedial actions. To conclude, blockchains are made up of an ever-growing list of records called blocks, each of which is encrypted and distributed over the network, as described by CP Jagrat and J. Channegowda[12]. This aspect of the blockchain adds an extra layer of security to the data being communicated, which has been frequently used in the financial sector. The documents reviewed above provide valuable feasible options for the construction of a blockchain traceability platform for artworks, which will be incorporated into the future design of the blockchain platform.

3   Proposed Method

The proposed method uses Blockchain and sentimental analysis for the Chinese Artwork. The collected data is used in the operations such as Blockchain and Sentimental Analysis. Defined usage of the blockchain and sentimental analysis are having their own abstract. The transaction ID is used for holding the data and sharing it in the future. A transaction ID will be known to everyone. But Sentimental Analysis is used by the platform owner to analyze

his or her business for future predictions. Operations on data are associated with Sentiment parameters. The parameters are given as positive weightages such as Great, Love, Success, Family, Art, Culture, High, Status, and Legacy.

This work proposed a digital platform to search, identify, verify, sell, and purchase ancient artworks. This business is trending in China, where many Chinese people are interested in buying ancient Chinese artworks such as paintings, amulets, Coins, Asian paintings, Thai Amulets, and Thai paintings. Same time, Many South Asians are interested in Chinese Ancient arts and artworks [11,14]. But people may not know the originality of every art and artwork. So, the proposed method can identify the original painters who are approved by ancient schools related to every unique art and artwork. This platform can be operated in China, Hong Kong, and Thailand. The creators must register as the artists on this platform, and they can share their qualifications, school they are related to the city, artworks with images, videos, and their unique identity. All this information will be stored and given with a blockchain transparent ID. So, the blockchain ID will be given to every single unique artwork, so no one can copy or cheat the unique artwork [14,15]. Every art will have the artist's name, the school or university, or the company that is dealing with the artwork. anyone who claimed that they have a particular artifact cannot be able to imitate the value will be first given by the seller or the artist based on the demand the value will be increasing or changing Aunt like painting antique weapons antique tools antique maps antique jewelry antique dressers and so on can be used in this platform every city has their own antique culture a style of their own artifacts**Error! Reference source not found.**. This proposed digital platform may be considered as a cross-border e-commerce platform, where the user can register, search, and choose their artifacts based on their interests based on their budget also. The payment may be done by Chinese Yuan, Thai baht, Indian Rupee, Visa cards, WeChat, and Alipay methods. For identifications, the users may be requested to provide their passport ID are any accepted government ID, mobile number, and address for transparency in the blockchain.

4   Conceptual Framework

Following the completion of the many research activities for this study, it is necessary to comprehend and collect information or elements in order to continue the research on Arts and Blockchain technology. Only this platform can verify the date of the art's creation, the artist's name, the date of production, the date of sale, and so on. Quantitative data will be created and kept as a result. The data was employed in this study, and the quantitative data was crucial to understanding the economic success of blockchain and the arts, as well as the technology's future path. The existing data utilized is from the user inputs, feedback, and comments associated with the proposed platform in the research. The data relating to Artist, art, date, type, value, demand count, number of views, number of likes, and number of comments towards the Artwork. The chosen research methods were based on the limited data available on the selected field of research, and future research would be essential to help elaborate this research further.
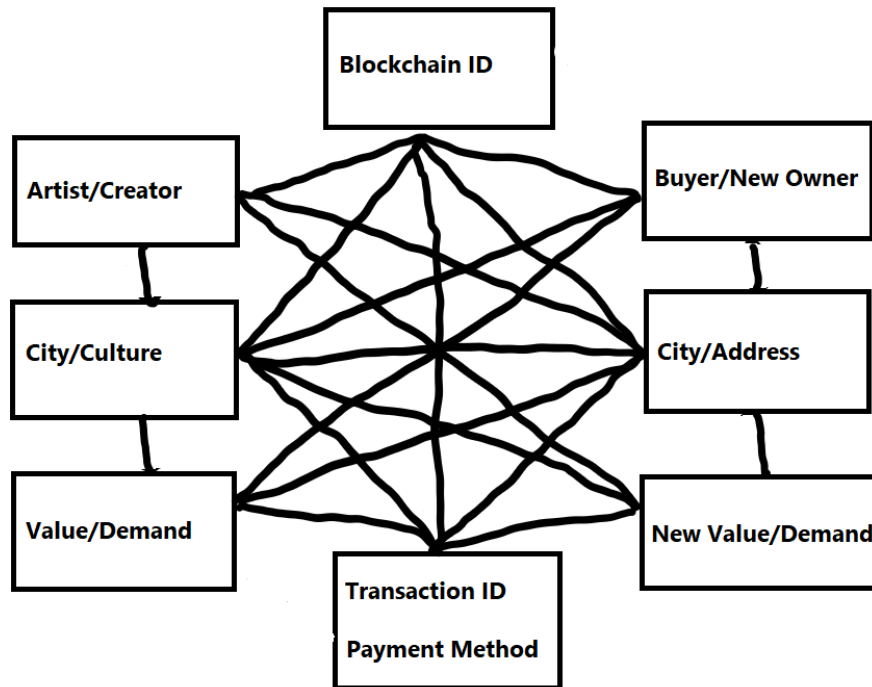
Fig. 1. Elements in the platform

The study relied on descriptive research to gather data in order to better understand the relationship between the arts and blockchain technology, and both qualitative and quantitative data were employed. To understand the current condition of concerns and challenges facing ancient arts and artists, qualitative data was gathered through secondary data. Quantitative data were gathered from government publications and books in order to better comprehend the use of blockchain technology in this industry and the technology's future trajectory. The SWOT analysis was used in this study to identify problems and potential solutions for improving the ancient artworks sector using the Blockchain system, as well as to aid in the future development of this research, as it is critical to adapt to technologies that will protect artists' work from fraud. When artists have control over how their work is delivered, they create demand, which serves to raise the value of art. Artists, art enterprises, and art galleries can use this research to find and integrate new technology into their organizations.

As explained, Blockchain technology is a type of technology that is decentralized. Diego goes on to say that no third party can claim ownership of the system, which lowers implementation costs and gives the user entire control; this eliminates transaction fees imposed by third-party providers. Because blockchain technology is immutable, it means that transactions can be tracked by the vendor and supplier, which is critical in preventing theft and counterfeiting. It's similar to a digital ledger that uses a decentralized system to carry out transactions, demonstrating that the system is highly efficient and that, with the help of the internet, transactions can be carried out quickly and easily; with the availability of smart contracts on the blockchain network, processes can be automated, whereby once pre-contract negotiations have been agreed upon and met, the contract automatically triggers itself, reducing human

interaction. Diego goes on to say that blockchain technology is also environmentallybeneficial because it uses less energy, which cuts down on operating costs.
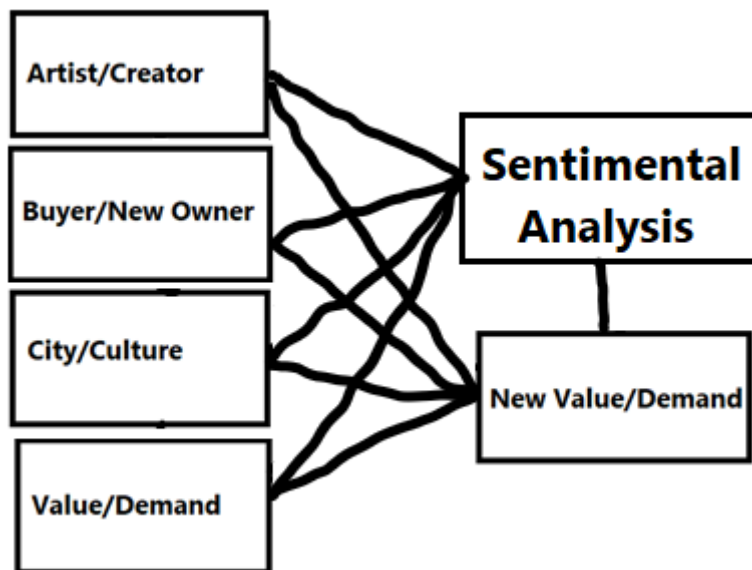
Fig. 2. Sentimental analysis associated with art and values

An artist's sentiment analysis is vital information that a collector is looking for. Sentiment analysis is the study of people's attitudes, sentiments, evaluations, appraisals, attitudes, and feelings about things like products, services, and organizations, as well as their attributes. It encompasses a significant problem area. Sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, effect analysis, emotion analysis, review mining, and so on all have distinct names and tasks [13].

The primary goal of combining art and blockchain technology [14] is to better comprehend how ancient artwork has revolutionized how artists distribute their work and secure it from harmful individuals. This proposed work shows that how technology has changed the demand for art to see if this new technology will be efficient in the future. With the rise in popularity of digital arts in the marketing industry, such as infographics, posters, and banners, there is an increase in demand for art, which generates a market niche for supply. It is critical to adapt to solutions that will protect artists' work from fraud as technology trends continue to evolve. When artists have control over how their work is delivered, they create demand, which serves to raise the value of art [15]. Artists, art enterprises, and art galleries can use this research to find and integrate new technology into their organizations. Quantitative and qualitative statistics will be critical in determining whether ancient artworks have a future in the art world and if they are effective in assisting artists in increasing their profits.

5   Study Area and Duration

This platform can be operated in China, Hong Kong, and Thailand. Further, this work will be extended to the US, Japan, and India. For this proposed work, 35 articles were reviewed. Since the proposed work is dealing with arts and crafts, it requires 3 months for the data collection on arts and crafts.

6   Conclusion

The proposed work was handled for following and tracing ancient artwork. The research work proposesa digital platform to cluster the artists, artworks, and art schools and give them digital identification trace or ID. Blockchain supplied proof for digital transactions and will help preserve history in the future through digital transactions. This can be used for further advertisements, increasing the sales, reducing the fake creations, reducing the scams, and, Increasing the tourism and awareness of ancient artworks from China and Thailand. It may be spread in the future. Every art form, such as art, fashion, dance, painting, statue, music, and so on, was significant at the time since it reflected their country's language. They are currently looking for roots that have roots in their past, thus blockchain will be useful for future root searches. Because blockchain and ancient art are still relatively new notions in the corporate world, this research will help the reader understand where the technology comes from and how it has evolved over time.

References

[1] J. Lee, "Development of Craft Copyright Industry Using Blockchain Technology," *2021 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter)*, 2021, pp. 263-264, DOI: 10.1109/SNPDWinter52325.2021.00065.

[2] Garcia-Teruel, R. M., & Simón-Moreno, H. (2021). The digital tokenization of property rights. A comparative perspective. Computer Law & Security Review, 41, 105543.

[3] P. R. Nair and D. R. Dorai, "Evaluation of Performance and Security of Proof of Work and Proof of Stake using Blockchain," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 279-283, DOI: 10.1109/ICICV50876.2021.9388487.

[4] W. Yu and S. Huang, "Traceability of Food Safety Based on Block Chain and RFID Technology," 2018 11th International Symposium on Computational Intelligence and Design (ISCID), 2018, pp. 339-342, DOI: 10.1109/ISCID.2018.00083.

[5] K. Balasubramanyam, "Adaptation of BlockChain Technology in manufacturing sector," 2020 International Conference on Mainstreaming Blockchain Implementation (ICOMBI), 2020, pp. 1-13, DOI: 10.23919/ICOMBI48604.2020.9203545.

[6] S. P. Gayathri, S. Vijayalakshmi, Siva Shankar Ramasamy, "Tracking IoT-Based Devices Parts from Manufacture to End of Life by Leveraging Big Data in Blockchain Environment", Convergence of Blockchain and IoT, Taylor & Francis, Edition 1, pages 20, Nov 2021.

[7] P. Kuhlang and A. Sunk, "Productivity improvement in logistical work systems of the genuine parts supply

chain," 2015 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2015, pp. 280-284, DOI: 10.1109/IEEM.2015.7385652.

[8] Y. K. Ro, J. K. Liker and S. K. Fixson, "Modularity as a Strategy for Supply Chain Coordination: The Case of U.S. Auto," in IEEE Transactions on Engineering Management, vol. 54, no. 1, pp. 172-189, Feb. 2007, DOI: 10.1109/TEM.2006.889075.

[9] J. Block, T. Tyrberg and Yuan Fuquing, "Optimal repair for repairable components during phaseout an aircraft fleet," 2010 IEEE Aerospace Conference, 2010, pp. 1-8, DOI: 10.1109/AERO.2010.5446880.

[10] D. Sathya, S. Nithyaroopa, D. Jagadeesan, and I. J. Jacob, "Block-chain Technology for Food supply chains," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 212-219, DOI: 10.1109/ICICV50876.2021.9388478.

[11] S. W. Sheng and S. Wicha, "The Proposed of a Smart Traceability System for Teak Supply Chain Based on Blockchain Technology," 2021 Joint International Conference on Digital Arts, Media, and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, 2021, pp. 59-64, DOI: 10.1109/ECTIDAMTNCON51128.2021.9425780.

[12] C. P. Jagrat and J. Channegowda, "A Survey of Blockchain Based Government Infrastructure Information," 2020 International Conference on Mainstreaming Blockchain Implementation (ICOMBI), 2020, pp. 1-5.

[13] Joanna Kołodziej& Horacio González-Vélez (Eds.) (2019). High-Performance Modelling and Simulation for Big Data Applications. Springer Switzerland.

[14] Kochetkova, Marina. "BLOCKCHAIN IN THE ART MARKET: Opportunities and Challenges." (2020).

[15] Whitaker, Amy. "Art and blockchain: A primer, history, and taxonomy of blockchain use cases in the arts." Artivate 8, no. 2 (2019): 21-46.

B2C E-CommercePlatforms and Marketing Strategiesof Thailand Education Products Towards China's 2 deduction EducationPolicy

Zhe Tao[1], Siva Shankar Ramasamy[*2], Bibi She[*3]

International College of Digital Innovation, Chiang Mai University, Chiang Mai,

239 Nimmanhaemin Road, Suthep, Muang,Chiang Mai,50200,Thailand,

zhe_tao@cmu.ac.th[1], sivashankar.r@cmu.ac.th[*2], bibi_s@cmu.ac.th[*]

Abstract

Following the covid pandemic, every country is attempting to rebuild its economy and support its local population. Improving education in a country benefits it both directly and indirectly. In the meantime, China is attempting to restrict the policy on k12 education providers in the country. As a result, the current project proposes a solution based on a Digital Platform, which focuses on strengthening the e-commerce platform for education providers in Thailand in relation to China following the two deduction policy. We need to focus on the economic environment, consumption methods, consumption structure, consumer demands, language used, money payment issues, and unique cross-border trade policies to find a more efficient way to improve the specified strategies, including cross-border IT platforms and marketing strategy. A framework of B2C E-Commerce based marketing strategies is introduced, which will better serve Thailand education providers to export education services to China, based on interdisciplinary research from IT and Economics and based on WeChat mini program, computer programming language, and management information system.

Keywords: cross-border e-commerce, marketing strategies, China 2 deduction education policy, WeChat mini

1. Introduction

China has a stable increasingGDP after Covid 19, whose GDP is 98651.5billion RMB in 2019.It has a stable GDP increase in 2020 and 2021, which is 101356.7 billionRMB and 114367 billion RMB (National Bureau of Statistics of China, 2022). [1]China's total e-commerce trade has increased from 21790 billion RMB in 2015 to 37210 billion RMB in 2020, even after CovidPandemic in 2019, China still had a 4.5% increase in E-commerce trade in 2020(InternetSocietyof China, 2021). [2] From China Central administration of Customs, China has a cross-border e-commerce trade of 1980 billion RMB in 2021, increased by 15% compared with 2020(Askci.com,2022).[3] Therefore, China has a strong cross-border commerce scale and the demand for cross-border services and products is still high. In 2021, the total trade between China and Thailand had increased to 131 billion US dollars, which made China the top trade partner to Thailand compared with other trading countries (Sun, 2022). [4] Therefore, based on the scale of China and the close relationship between Thailand and China, enhancing e-commerce trade could be one opportunity for Thailand to recover its economy from Covid Pandemic. However, due to competition with other developed countries, how to find a more efficient way and innovative channel to achieve a significant educational industry share in cross-border commerce needs to be verified by research. This article discussed 3important aspects in order to improve strategies: market data analysis, platform design and selection, strategy selection, and innovative challenges.

2. Literature Review

2.1 Research on IT, Education, and Trade Environment

According to the data published by China MIIT, until 2021, China has 1.643 billion registered mobile accounts, mobile using coverage rate is 116.3%. In 2021, mobile Internet usage is 212.5 billion GB, an increaseof 35.5% compared with 2020 (MIIT, 2022). [5] China has very strong demand of usage of Internet, which means that e-commerce platform had a large number of potential users. On the basis of China's 2 deduction policy, the market share of China's K12 industries is significantly affected and it will give more opportunities based on substitution theory (Mankiw, 2020). [6] On 24 July 2021, China Government Central Office published reduce K12 students' study pressure policy announcement, and all of the commercial private and public education and training companies (K12 subjects related) would be closed or temporarily terminated with strict limiting reopen policy (General Office of the State Council of the People's Republic of China, 2021). [7]

This article would also havea deep research on the opportunities for Thailand Merchants after China education companieswere affected by the policy and find the relationships.In China, the examination for entering into the next levelof school has been very stressful for students in China, China government would like to help improve the students' study environment and solve the psychological stress issue. Various research and examination methods and related models would be used to analyze the current situation.

According tothe iResearch Report, until 2020, the total K12 education industry scale was 700 billion RMB(iResearch, 2021).[8] Due to the effect of Covid 19, the China K12 training industry suffered a downturn in 2020. **See Figure 1:**

2016-2020 China K12 Industry Scale
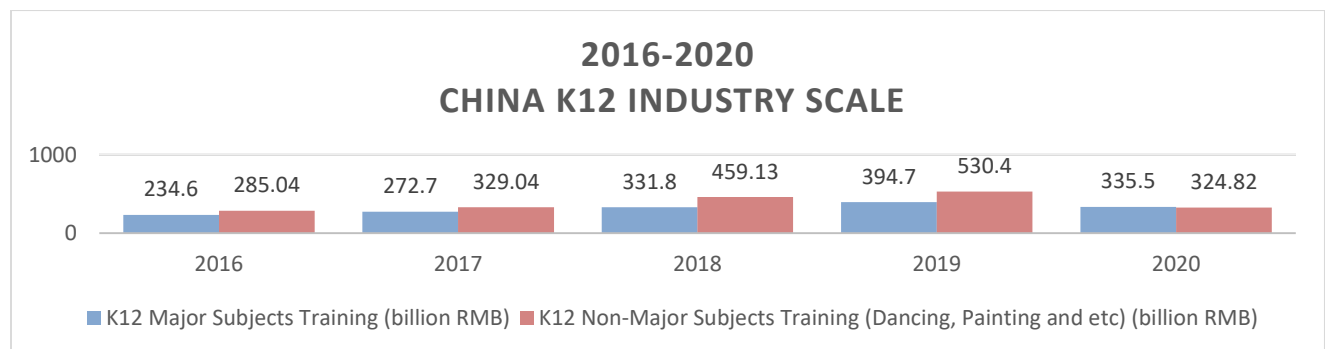
Fig. 1. 2016 to 2020 China K12 Industry Scale

Due to the high needfor K12 education, more and more companies are joining this area for profit.From the research report published on 2018 from KLBK (Key laboratory of Big Data Mining and Knowledge Management of Chinese Academy of Science),1230 family samples from different level city groups were investigated. From the sample, 70% of the parents would like to use online K12 education, and the average expenditure from each family would be 6432.2 RMB (Key laboratory of Big Data Mining and Knowledge Management of Chinese Academy of

Science, 2018). [9] According to China Statistic Bureau, the average revenue per person in China is only 5600 RMB,which means that at least more than one month's salary was spent on the K12 training in these samples. However, this is only online K12 expenditure, offline K12 expenditure has not been added to this number. In China, there is a proverb called'learning better means living better, which means that entering a good university means that you could get an entrance of getting a better job. Yue, Wen and Ding (2004) concluded that a better university would give students more opportunities to get better jobs. [10]Therefore, to enter better public schools and universities, students have to face 3 important exams, one is from Little School to Junior School, one is from Junior school to higher school, and the 3rd exam from higher school to better university. K12 training 2 deduction policy would force the parents to find another way to improve their children's study results and due to the unequal education system had not been changed. The single policy on K12 seems to cause unexpected results in another issue but will give more opportunities to Thailand's education industries. Therefore, based on the situation available, innovative design of educational product platform might enhance the competitiveness of Thailand's education export industry. According to CNNIC(China Internet Network Information Center)'s, there are 325 million education online users in China in June 2021.[11]From the world bank database, we could find that Thailand has a stable fluctuation of expenditure on education based on its GDP, while it has a downward expenditure trend in the proportion of GDP (UNESCO Institute for Statistic, 2021). [12] However, from China's data graph, we would also find that there is a downward trend in expenditure in proportion to GDP. And they have a similar fluctuation trend in expenditures.Inmacroeconomic analysis, Thailand's government had a long-term stable investment in education expenditures, which means that the political strategy on education is very stable. Zhou (2014) had researched the China-Thailand Cooperation based on theASEAN and China Cooperation Political view, and also argued the China Government's "one belt one road" policy effect on the long-term economic development between the two countries. [13] Shi (2003) had implemented statistical methods to verify the export structure between China and ASEAN, and concluded that they have similar structures, and it implies that it had a trend that they would focus on a long-run similar export structure based on their similar resources.[14] Zhouet al. (2015)had proved the increasing trend of cooperation with the "one belt one road" related countries due to nearby traffic effects. [15] Fan and Tang (2003) acknowledged that Thailand had an International Education advantage due to its cooperation with English developed countries by learning developed countries' education models.[16]

2.2 Research on Management and IT theory Cross Application：

Heinz Weihrich supported the SWOT analysis as one important theory used in strategy design (Weihrich, 2010). [17]SWOT theory has its internal strength inanalyzing the internal and external environment. Akhobadze (2018) researched on wasted industry water by using SWOT analysis as a practical way. [18]Therefore, our research team has a SWOT analysis of the education type analysis.

We need to discusswhich kind of education product is better for promotion, one is cross-border promotion and domestic education, and the other is online education (offshore education) through e-commerce. Due to the effect of Covid 19, online education could be one good option. Zhang (2017) discussed the 3 different cross-border e-commerce standards: the first one is B2B, B2C and C2C type，the second is vertical or multiple ones, and the third is a self-

selling platform, an agent platform, multiple selling platforms, and multiple agent platforms. [19] Therefore, the product sellers should choose to match the suitable cross-border channel in order to match its advantage based on the structures and SWOT.Based on Michael Porter's competitive advantage theory (Porter,2011), we recommend that it would use Porters' five forces theory to measure the products offered to Chinese users. [20]

2.3 Research on WeChat mini platform application and design

Until June 2021, China has a significant proportion of video shoppers 0.888 billion people, increased by 14.4 million people compared with 2020, which stands at 87.8% of total local internet users.Short-video users' average use time is 125 minutes per user per day, and 53.5% of the short video users would watch every day (CNNIC,2022). [11] Yu (2016) discussed the short video app had brought users multiple feelings and attracted more users based on psychological analysis.[21]Therefore, the short-video platform could be one option that markets Thailand products in one efficient way. Based on the sample data analysis of iresearchiclick in 2019, 1416 users' samples were researched, and 90.4 of them had used Wechat mini program before, 48.7% of them had used the Wechat program more than 5 times (iResearch Inc, 2019). [22] According to the Tencent Financial Reportof the third quarter of 2021, the quantity of WeChat users had increased to 1.26 billion in China 2021. It showed that WeChat had a strong user group in China (Tencent Inc, 2021). [23]
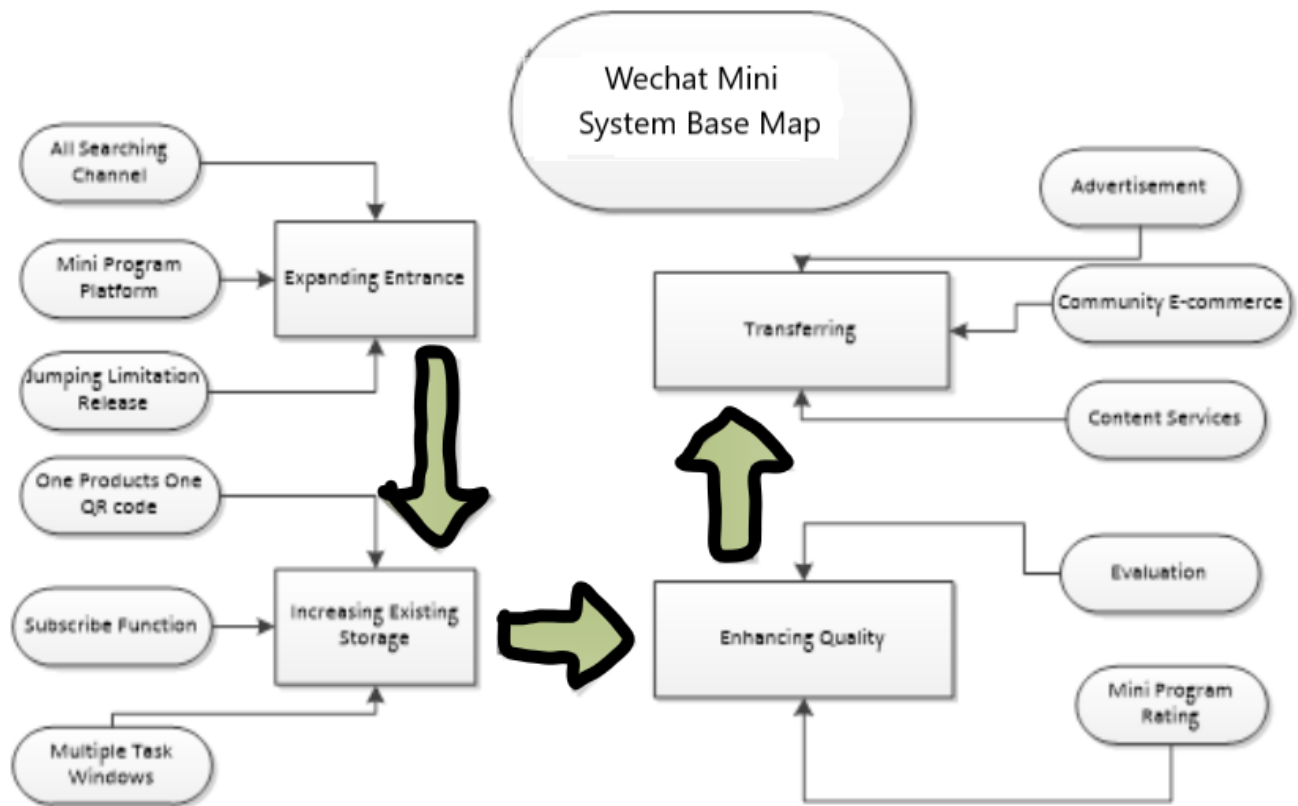

Fig. 2. WeChat Mini System Base Map

3. Applying Innovation and Future Strategies

3.1 Marketing and Promotion Strategy

To find out the factors that might motivate a person to do better in life and work and to understand what encourages the most, psychologist Abraham Maslow in 1934 postulated Maslow's Pyramid or theory of human behavior. Based on the theory of Maslow's hierarchy of needs, five needs identification are used to explain the needs of human beings. Certain lower needs should be satisfied before higher needs could be satisfied. And from lower to higher, they are psychological needs, safety needs, belonging needs, esteem needs, and self-actualization needs (Gawel,1996).[24]SanchitaGhatak and Surabhi Singh (2019) researched on Maslow's Theory in the social media adoption and concluded that companies who are willing to promote social media adoption must focus more on customers with belonging needs and esteem needs.[25] Sinclair and Dowdy (2005) have suggested that emotional belonging leads to identifying closeness in a group and developing better interpersonal relationships with others in their research. [26] Therefore, to use WeChat mini app to design the marketing strategy, it would suggest more concern on the customers' belonging needs and esteem needs. Chang and Zhu (2011) reviewed the previous social networking service studies and found that sociality and entertainment are the two dominant motivating factors to adopt social networking services and the third one is information motivation. [27]Chung and Austria (2010) in their research investigating gratifications underlying social media usage, argued that social media usage derived from entertainment, social interaction, and information gratification anda positive relationship to attitudes toward social media marketing messages was found from social interaction and information gratification.[28]Therefore, when designing the strategies, the Thailand merchants could design strategies based on the group socialization oriented and entertainment-oriented base.

We will use affordance theory-based framework to evaluate the app design framework after mini app set up.(Gibson, 2014)[29]. Volkoff&Strong (2013) had researched on affordance theory in IT area and illustrate how affordance-based theories informed by critical realism enhance ability to explain IT-associated organizational change. [30]. The principles could be used in identifying the logic design of WeChat Mini program.

3.2 IT Platform Design

Based on the strategies and frame design, we decide to test on WeChat mini program as our selected tools.

Short Video Marketing (Customer Attraction Purpose)	The platform uses the short video of recording videos or designing advertising education products videos of Chiangmai (as a sample target city), to promote local characteristics through Douyin (China Tiktok), Weishi, Weibo, Bilibili, Kuaishou，and other channels to attract potential consumers and conduct inflow of attractions of customers
Scan code or Advertising Button Press from the screen by authorization	After scanning the QR code by Wechat mini program, the user authorizes Wechat to log in and enter the homepage from their Wechat account.
Display the information available for users and Match relative products with user's needs	The WeChat mini program would have two functions on the first appearing page, one button would direct the user to see the list to products and the other would show the search key function, which could give the user a function of typing text, recording voice, and delete functions.
Function one: Listing Products	The listing page would appear, and show the choosing options with more details including an introduction, pictures, and videos on the list, the user could press the product button they need and then add it to the shopping cart. Or they could also press the go to homepage button for another choice.
Function two: Searching Products Directly	The user types the keywords into the form and searches for the product they need, and classified contents would appear on the page. The user could press the product button they need and then add it to the shopping cart. Or they could also press the go to homepage button for another choice.
Related products and good education products recommendations	The user could view the "Star" products on the current platform in the scroll bar on the homepage. After sliding the home to the downside, you could view the review and recommendations of the products or the user could also choose to view the top-rated page only.
Complete online merchandise order	After the product selection is completed, users can enter the shopping cart page and place an order online through the platform.
Inquire about the completion of the transaction, make an evaluation after the sale	Users could click the order function button, and see the details of the process, and their account would be automatically activated for the products they had already paid. They could give recommendations after buying. Logistics information would appear after orders. Before and after buying the products, the user could type on the customer service chat form to consult with the service managers.
Collect high-quality feedback videos or texts to receive voucher rewards or account credit points	After completing the transaction, the user could record the product feedback video or write some suggestions, and upload to the sharing ideas feedback chart. WeChat mini program would automatically evaluate the quality of the feedback by program, and give back some vouchers or discount for the next time purchase.

Fig. 3. WeChat platform design based on our research group

Discussion

The Proposed work followed the theory-based strategy and follow the process of designing innovative framework of platform marketing, Followed by Cheng et al.（2019）'s research, on the perspective of affordance theory, both perceived affordance and real affordance can be provided to users via a WeChat mini programin WeChat ecosystem and no extra cost on downloading apps.[31] However, due to the limitation of mini program set up cost, suppliers could follow the strategies but still need to concern other hypothesis factors.

Conclusion

Implementing strict policy on k12 education providers in China indirectly reflect the economy. The goal was sincere, but the outcome demonstrated that we needed to reform or find a reasonable answer. The proposed solution provided a B2C based E-Commerce Platform, WeChat mini, to increase Thailand's education providers' ties with China in keeping with the two deduction policy. The economic environment, consumption techniques, consumption structure, consumer expectations, language used, money payment concerns, and distinctive cross-border trade policies all seem to be part of the structure of B2C E-Commerce marketing tactics. WeChat Frame structure process design combined with Marketing Strategies could be one solution for education merchants to enhance their sales. Two countries are effective in managing digital platforms and enhancing specific strategies, such as cross-border IT platforms and marketing strategies. Future interdisciplinary study in IT, Economics, and Agriculture-based marketing could improve the proposed framework.

References

[1] National Bureau of Statistics of China.(2022). *Statistical Communiqué of the People's Republic of China on National Economic and Social Development in 2021*. http://www.gov.cn/xinwen/2022-02/28/content_5676015.htm on April 08 2022.

[2] Internet Society of China. (2021). *China Internet Development Report.*https://www.isc.org.cn/article/40203.html

[3] Askci.com.（2022）.*Research Report on Market Prospects and Investment Opportunities of China's Export Cross-border E-commerce Industry*, https://m.askci.com/news/finance/20220117/1126381726159.shtml

[4] Sun G.（2022，April19）. Jointly building the 'Belt and Road' gives Thailand unprecedented opportunities for development.*People'sDaily:China Edition.* http://world.people.com.cn/n1/2022/0419/c1002-32402213.html

[5] Ministry of Industry and Information Technology. (2022).*2021 Statistical Bulletin of the Communication Industry*. https://www.miit.gov.cn/gxsj/tjfx/txy/art/2022/art_e8b64ba8f29d4ce18a1003c4f4d88234.html

[6] Mankiw, N. G. (2020). *Principles of macroeconomics*. Cengage learning.

[7] General Office of the State Council of the People's Republic of China. 2021.Opinions *on further reducing the burden of homework and off-campus training for students in K12 education*, http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202107/t20210724_546576.html?ivk_sa=1024320 u

[8] iResearchInc. (2021). K12 education Research Report. https://report.iresearch.cn/report_pdf.aspx?id=3780

[9] Key laboratory of Big Data Mining and Knowledge Management of Chinese Academy of Science. (2018). *China K12 Online Education Market Research Report.*https://www.sohu.com/a/252287436_386633?_f=index_chan25news_104

[10] Yue C, Wen D, Ding X. (2004). Job Search and Starting Salary: An Empirical Analysis of the Employment Competitiveness of College Graduates.*Management World* (11),53-61. doi: 10.19744/j.cnki.11-

1235/f.2004.11.008.

[11] China Internet Network Information Center (CNNIC). （2021）.*No.48 Statistical Report on Internet Development in China.*http://www.cnnic.cn/gywm/xwzx/rdxw/20172017_7084/202109/t20210923_71551.htm

[12] UNESCO Institute for Statistics. (2021). *World GDP database.* https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS?end=2019&start=2009

[13] Zhou F. (2014). The role of strategic pivot in ASEAN cooperation in China-Thailand relations: From the perspective of the Century Maritime Silk Road, *Southeast Asian Affairs* (03),17-22+40. doi: 10.14073/j.cnki.nywtyj.2014.03.003.

[14] Shi Z. (2003). Export Similarity and Trade Competition: A Comparative Study of China and ASEAN. *Finance & Trade Economics* (09),53-57+97. doi: 10.19795/j.cnki.cn11-1166/f.2003.09.011.

[15] Zhou J.,LiuC.,Yin G. & Tang Z.(2015). The trade pattern and economic contribution between China and the countries along the "One Belt and One Road". *Progress in Geography* (05),598-605.

[16] FanW.&TangY. (2003). Guangxi-ASEAN Higher Education Internationalization Strategic Thinking. *Crossroads:Southeast Asian Studies* (11),19-22.

[17] Weihrich, H., &Cannice, M. V. (2010). *Management*. Tata McGraw-Hill Education.

[18] Akhobadze, G. N. (2018, November). SWOT analysis of industrial wastewater. In *IOP Conference Series: Materials Science and Engineering* (Vol. 451, No. 1, p. 012212). IOP Publishing.

[19] Zhang X. (2017).Types and Operation Modes of Cross-border E-commerce. *China Business and Market* (01),76-83. doi: 10.14089/j.cnki.cn11-3664/f.2017.01.010.

[20] Porter, M. E. (2011). *Competitive advantage of nations: creating and sustaining superior performance*. simon and schuster.

[21] Yu X. (2016).Analysis on the dissemination and marketing mode of short videos in the mobile social era. *View on Publishing* (24),57-59. doi: 10.16491/j.cnki.cn45-1216/g2.2017.0082.

[22] iResearchInc. (2019).*WeChat Mini Program Market Research Report.*https://report.iresearch.cn/report_pdf.aspx?id=3428

[23] TencentInc. (2021).*Tencent Announces 2021 Third Quarter Results Report*https://static.www.tencent.com/uploads/2021/11/10/d89fcf7a69f0dea47aff543fbe030e4d.pdf

[24] Gawel, J. E. (1996). Herzberg's theory of motivation and Maslow's hierarchy of needs. *Practical Assessment, Research, and Evaluation*, *5*(1), 11.

[25] Ghatak, S., & Singh, S. (2019). Examining Maslow's hierarchy need theory in the social media adoption. *FIIB Business Review*, *8*(4), 292-302.

[26] Sinclair, V. G., & Dowdy, S. W. (2005). Development and validation of the Emotional Intimacy Scale. *Journal of nursing measurement*, *13*(3), 193-206.

[27] Chang, Y. P., & Zhu, D. H. (2011). Understanding social networking sites adoption in China: A comparison of pre-adoption and post-adoption. *Computers in Human behavior*, *27*(5), 1840-1848.

[28] Chung, C., & Austria, K. (2010). Social Media Gratification and Attitude toward Social Media Marketing

Messages: A Study of the Effect of Social Media Marketing Messages on Online Shopping Value. *Proceedings of the Northeast Business & Economics Association*.

[29] Gibson, J. J. (2014). *The ecological approach to visual perception: classic edition*. Psychology Press.

[30] Volkoff, O., & Strong, D. M. (2013). Critical realism and affordances: Theorizing IT-associated organizatio nal change processes. *MIS quarterly*, 819-834.

[31] Cheng, A., Renb, G., Hongc, T., &Kood, C. (2019). The WeChat mini program for smart tourism. *Asia Pacific Journal of Information Systems*, *29*(3), 489-502.

Dynamic Domain Name System 'A' Record Update Using Cloudflare

B. Karthikeyan, Ganesh Kumar R

Department of Computer Science, Bishop Heber Collge, Tiruchirappalli.

karthikeyan.cs@bhc.edu.in

Abstract

Disclosed are various embodiments for validating updates to dynamic domain name system (DDNS) A request is received to modify at least one DDNS record associated with a domain owned by a domain owner. The request to modify the at least one DDNS record is compared with at least one policy. The at least one policy is configurable by the domain owner. The requested modification to the at least one DDNS record is selectively granted based at least upon the comparison. APIs or websites run on internet connections that have their IP addresses changed frequently; this creates a problem if the operators of those properties want to give a hosted resource a specific domain name, which must then store an IP address in Domain Name System (DNS) records. Dynamic DNS (DDNS) is a service that keeps the DNS updated with a web property's correct IP address, even if that IP address is constantly being updated.

Keywords: Domain Name System, Cloudflare, IP Address;

Introduction

Dynamic DNS (DDNS Pro) allows you to access your devices from the internet via a simple to remember domain name. Example: Instead of connecting to your security camera, DVR, or computer through a difficult to remember IP address like 216.146. Cloud flare provides an API that allows you to manage DNS records programmatically. To set up a Cloud flare dynamic DNS, you'll need to run a process on a client inside your network that does two main actions: get your network's current public IP address and automatically update the corresponding DNS record.

Dynamic DNS (DDNS) is very useful if you need to access internal network services from across the Internet. It isn't designed for hosting a business website, for that you will need standard web hosting. DDNS can be used for many home-lab services as it simply tracks the external IP address of your home network. If you're fortunate enough to have a static external IP address, DDNS will do nothing other than allow you to connect a domain name to your external IP address. However, if you have a dynamic IP address (as most people do), DDNS will allow you to ensure you're always connecting to your external IP address. This can all be accomplished relatively easily by following the instructions below on how to set up DDNS on pf Sense using Cloud flare.

Dynamic Domain Name System

Dynamic DNS (DDNS) is a method of automatically updating a name server in the Domain Name System (DNS), often in real time, with the active DDNS configuration of its configured hostnames, addresses or other information.

The term is used to describe two different concepts. The first is "dynamic DNS updating" which refers to systems that are used to update traditional DNS records without manual editing. These mechanisms are explained in RFC 2136, and use the TSIG mechanism to provide security. The second kind of dynamic DNS permits lightweight and immediate updates often using an update client, which do not use the RFC2136 standard for updating DNS records. These clients provide a persistent addressing method for devices that change their location, configuration or IP address frequently.
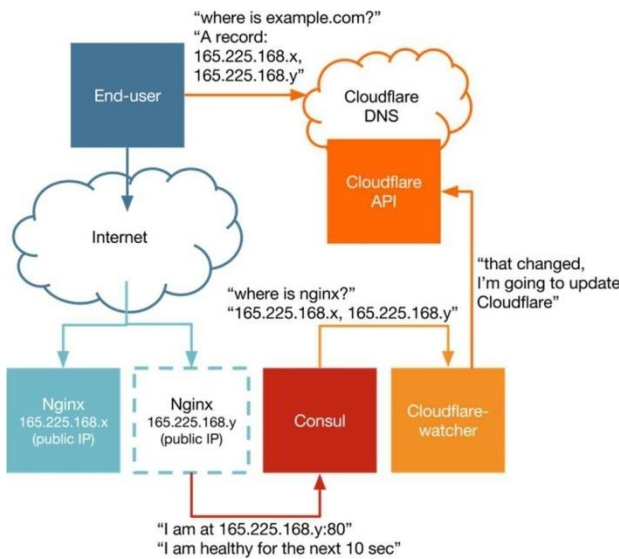

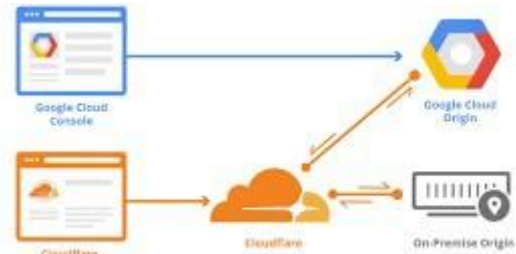Fig. 1. Dynamic DNS   Fig. 2. Setting up Dynamic DNS through Cloudflare

The DDNS works in the following way: The DDNS client monitors the IP address for changes. When the address changes (which it will if you have a dynamic IP address), the DDNS (or Dynamic DNS) service updates your new IP address.

This server will be connected to the internal network, and it will communicate with the Internet via NAT (network address translation) router. The NAT router will provide an internal IP address to the server.To Make it available externally, we will need to perform port forwarding and get external Port (Port ex) and external IP address (IP ex). Now the service you want to share is visible with IP ex and Port ex, and people can use it through the Internet. The problem comes when this IP changes.

Literature Review

ZhengWang [1] present in this paper DNS query for the CN TLD server follows Zip f- like distribution. Through the analytic model and analysis of cache effects, we have inferred the TTL's exponential-like distribution. This distribution lays out a theoretical foundation to improve current DNS systems for lowering the load of DNS CNTLD servers.

WeinaNiu et al [2] present in this paper efficient APT malware C&C domain detection approach capable of handling unmarked data. In our proposed anomaly detection algorithm, information entropy is introduced to indicate the different influence of each feature.

Xiaohui Gong [3] present in this paper teaching quality of national vocal music in Chinese music colleges and universities, it is imperative to innovate the model of national vocal music. Based on this, this study uses the deep recurrent neural network algorithm and the fuzzy evaluation model.

Haiyan Xu et al[4] present in this paper In this paper, we have analyzed the actual impact of the resolution dependence on the DNS, investigated the dependence relationship of 1 million domain names, and found that 86.14% of the domain names are dependent on name servers which are not in their own authorization domain (we do not consider the case of top-level domain and the root domain here).

Mohammed Abdulridha Hussain et al [5] present in this paper. This paper proposes Enc-DNS-HTTP to protect web browsing and to secure client–DNS server and client–web server communications. The scheme is based on sharing a web server public key through the DNS server. The key is signed by a trusted third party such as a web browser program creator.

Proposed System

DDNS, most commonly known as Dynamic DNS, is an automatic method of refreshing a name server. It can dynamically update DDNS without the need for human interaction. It is extremely useful for updating and when the host has changed its IP address.

Imagine this situation. You have a server in your office, and you are providing some service to your employees. You are using a standard/consumer-grade internet from a typical ISP (Internet service provider).

You are getting a temporary IP address that could change the next type you connect or change automatically after some time. To provide a service, you 3 options:

1. A Static IP address that could be expensive.

2. Change manually the IP address every time when it changes.

3. Automatically update the IP addresses – Dynamic DNS or DDNS!

DDNS is a service that automatically and periodically updates your DNS's A (IPv4) or AAAA (IPv6) records when your IP address changes. These IP changes are made by your Internet provider. With DDNS you don't need to worry about the changes in IP addresses!

Conclusion

At this point, you should understand what DNS is and what a name server is, as well as be familiar with technical concepts relating to IP addresses. Many books have been written about and dive deeper into the fascinating world of DNS, and there is so much more to learn. The topics that were not included in this article but are either part of DNS or very related include.

References

[1]   ZhengWang, "*Analysis of DNS Cache Effects on Query Distribution*", Hindawi, ISSN: 938418, Volume : 2013,PP : 01, 14 November 2013.

[2]   WeinaNiu,1,2 XiaosongZhang,1,2 GuoWuYang,2 JiananZhu,3 and ZhongweiRen1, "*Identifying APT Malware Domain Based on Mobile DNS Logging*", Hindawi, ISSN: 4916953, Volume : 2017,PP : 01,6 April 2017.

[3] Xiaohui Gong, "*Research on Discrete Dynamic System Modeling of Vocal Performance Teaching Platform Based on Big Data Environment*" , Hindawi, ISSN: 5111896, Volume : 2022, PP : 01, 24 February 2022.

[4] Haiyan Xu , Zhaoxin Zhang, Jianen Yan , and Xin Ma,"*Evaluating the Impact of Name Resolution Dependence on the DNS*" , Hindawi, ISSN: 8565397, Volume : 2019, PP : 01 9 September 2019.

[5]   Mohammed Abdulridha Hussain,1,2 HaiJin,1 Zaid Alaa Hussien,1,3 Zaid Ameen Abduljabbar,1,2 Salah H. Abbdal,1 and Ayad Ibrahim2 ,"*Enc-DNS-HTTP: Utilising DNS Infrastructure to Secure Web Browsing*" , Hindawi, ISSN: 9479476 ,Volume : 2017, PP : 01 3 April 2017

Classification Algorithms for Disease Diagnosis System

P.Usha[1] and Dr. M.P.Anuradha[2]

[1]Assistant Professor, Department of Information technology, Bishop HeberCollege,

[2] Assistant Professor, Department of Computer Science, Bishop HeberCollege,

[1,2] Affiliated to Bharathidasan University, Tiruchirappalli, India.

[1] usha.it@bhc.edu.in.com, [2] anuradha.cs@bhc.edu.in

Abstract

Machine learning algorithm plays a vital role in many fields such as medical disease analysis, text classification, user smartphone classification, images. Classification is one of the techniques in machine learning where the data can be represented in many forms such as rules, tree or functions. Classification algorithm is used to identify the categorical output of a dataset. In disease diagnosis system classification algorithms are very useful in decision making for further treatment options. In this paper Support vector machine, Decision tree, Naïve bayes, K nearest neighbor and Random forest models were used to predict the efficiency on bench marked datasets. The predictive abilities of classifier models were compared through performance metrics. Random forest models show greater accuracy than other models.

Keywords: Machine learning, Classification, Random forest, Algorithms, Disease analysis.

1. Introduction

Machine learning categorized into supervised, unsupervised and semi supervised or reinforcement learning. In supervised learning technique, inputs are trained and output is obtained from previous experience of a dataset. In unsupervised class labels are unknown inputs form a cluster based on similarity of a data to predict the output of a model. Semi supervised combines the features of both supervised and unsupervised mechanism.

In supervised method, sample labeled data is given as training and the output is predicted based on sample data. Machines are learnt from training and processing, then sample data can be provided to check whether it predicts correct output or not.

In medical diagnosis system, patient medical history, clinical features based on test and physical examination. Gene expressions are also involved in classification process. Classification has been an essential basis in statistics, machine learning, data mining, bioinformatics and Medical science. Classification is most important technique in machine learning. It is the process of separating data objects into categories for better interpretation of results Classification also called as the form of pattern recognition.

This paper is organized as, in section 2, review of different classification algorithms used in decision making, section 3 discuss how the classifier model works, section 4 explain dataset information and section 5 discuss the results obtained from classifiers. Finally the paper is concluded with future direction and references.

2. Literature Review

In [1], Takanori Yamashita, Yoshifumi wakata et al. Propose a Diagnosis Procedure combination & clinical path way to classify the post-operative patient status. To avoid long term hospitalization this epath is used in which Machine Learning algorithms are used in classification. epath is used to standardize the structure of medical records. In [2], Jobeda Jamal Khanam & Simon Y. Foo, compare, Neural Network & Machine Learning algorithm to predict diabetes. In this work, seven machine learning algorithms arte used. Linear Regression & Support Vector Machine performs well and shows 76-78% of accuracy & NN model shows 88-57% of accuracy with two hidden layers.

In [3], Nazin Ahmed, Rayhan Ahammed propose a web based application to predict diabetes based on Machine Learning algorithm using clinical data. In this work dataset is preprocessed by label encoding and normalization to improve accuracy. Using feature selection algorithm important risk factors are identified. Compared to existing Machine Learning algorithm this smart web application increases accuracy from 2.71% to 13.13%. In [4], Md Mahmudul Hasan, Gary J. Young et al. Algorithm propose a Machine Learning framework to identify the risk factor of Opioid disorder from healthcare data. Clinical & Prescription histories are used as features. Chi-squared used as a Feature Selection algorithm. The selected features are scaled, encoded compare with RF, DT, LR & GB MC algorithm RF shows higher accuracy then other models.

In [5], Beibei shi, Hua ye et al. alayse covid-19 severity using Machine Learning algorithm with enhansed optimisation of brain stroming (BSO) . SVM used with BSO & provides 91.91% of accuracy. This BSO-SVM can be treated as a computer aided technique in covid-19 prediction using coagulation index.In [6], Lerina hversano, Mario Luka Bernardi et al use Machine Learning approach to predict treatment for thyroid disease based on hormonal parameters & clinical data. 10 difference Machine Learning algorithm are compared in which extra tree classifier modal shows 84% of accuracy.

In [7], Rachel Hagan, Charless J.Gillan & Fiona Mallett compare Multi-Layer Perceptron (MLP), ensemble methods and SVM to classify cardiovascular disease. In ensemble methods, Decision Tree algorithm shows 90% of accuracy, MLP shows 88.7% and SVM shows 71% of accuracy. Compared to other models tree based ensemble is best suited for classification based on clinical features. In [8], Xiaolin Diao, Yanni Huo et al. propose Light Gradient Boosting Machine (LGBM) as a classifier model along with Grid search and cross-validation used to select optimal hyper parameters. This LGBM is used in primary diagnosis and provide automated ICD coding. This optimal feature with LGBM shows 95.2% accuracy in auto ICD coding of primary diagnosis based on clinical features.

In [9], Mohammed Amine Naji, Sanaa El Filali et al. apply SVM, Random Forest, Logistic Regression, Decision Tree and K Nearest Neighbour machine learning algorithms are used to predict breast cancer. Confusion matrix is used as a performance evaluator and SVM gives better performance obtained through Anaconda platform.. In [10], Amelec Viloria, Yaneth Herazo-Beltran et al, use Support Vector Machine to predict diabetes among Colombian patient. The model shows 99.2% of accuracy & 65.6% with patients from different ethnicity. In this work to predict diabetes body mass index (BMI) & Glucose levels are used.

3. System Model

Dataset is a collection of records which contains attributes. Dataset can be fully cleaned and completed. In some situation, dataset can have missing values, noisy data or irrelevant data. Thus pre-processing is necessary to clean the data. Pre-processing is a technique in which performed earlier in all task like machine learning [11], deep learning or data science. It transforms the data into a convenient format for easy and effective process to predict a result. Different pre-processing techniques are available such as sampling, imputation, transform, denoising, scaling and normalization etc. In this work data is pre-processed by scaling technique in which uniform scale is distributed for all attributes and it will standardize an independent feature in a fixed range. Instead of removing the missing values it will be replaced by mean.

Naïve Bayes [12] is a classification technique which use Bayes theorem which is based on conditional probability. It calculates the possibility of attributes whether it belongs to certain category or not. Bayes theorem represented as,

$$P\left(h/D\right) = \frac{P(D/h)P(h)}{P(D)} \qquad (1)$$

K nearest neighbour [11] algorithm is a pattern recognition algorithm which uses the training dataset to find the most related features. KNN used to classify the features within or its neighbour as a category. Various distance-based techniques are available such as Manhattan distance, Minkowski distance, Euclidean distance, Cosine distance and Jaccard distance. Mostly the distance between features can be calculated using Euclidean distance is denoted as,

$$distance(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (2)$$

Support vector machine [13] uses a hyperplane to separate data into categories. Hyperplane used to maximize the margin between two different classes. The equation for setting plane is,

$$Y = a * x + b \qquad (3)$$

Decision tree [14] is a visual representation of decision making process which use if-then rules to predict the decision. The node in decision tree can be split until to get terminal node. In decision tree all the true values comes right side of the tree and false will be in left side of root node. The node can be splitted using, Information gain or gini index. The entropy is calculated in decision tree by,

$$E(S) = \sum_{i=1}^{n} -p_i \, log_2 p_i \qquad (4)$$

Information Gain [15], separate the training dataset according to the target variable and is calculated using

$$Information \; Gain = 1 - E(S) \qquad (5)$$

Gini index used to split the decision nodes into various branches. It is calculated by,

$$Gini \ = 1 - \sum_{i=1}^{n}(p_i)^2 \qquad\qquad (6)$$

Logistic regression used to find the most probable relationship between dependent and independent variable. It is used to find the binary outcome of independent variable.

$$g\big(E(y)\big) \ = \ \alpha \ + \ \beta x_1 \ + \gamma x_2 \qquad\qquad (7)$$

where E(y) is the target $\alpha \ + \ \beta x_1 \ + \gamma x_2$ is the linear predictor variables.

Random forest is an ensemble learning algorithm which combines individual trees obtained from decision tree constructed at the training time and provide the output class which can either classification or regression based. It is represented as,

$$C_{rf}(x) \ = \ Majority\_Vote\{C_i(x)\} \qquad\qquad (8)$$
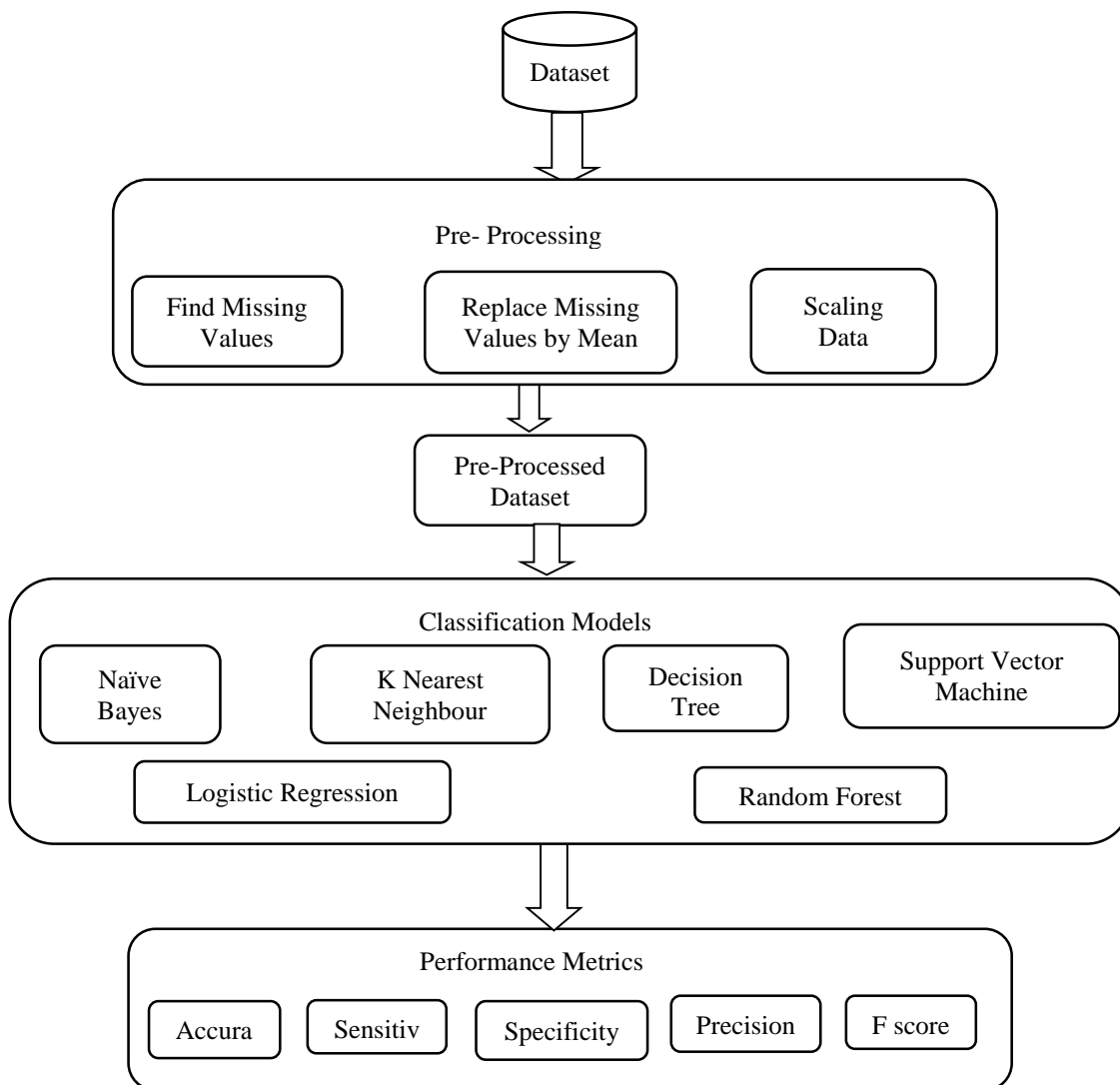
Fig. 1. System Model

The pre-processed data can be used as a input to the classifier models such as Naïve Bayes, KNN, SVM, Decision tree, Logistic regression and Random forest models. The model's performance is evaluated using different performance parameters. The block diagram is shown in Fig. 1.

4. Data Analysis

In this study, two benchmarked datasets breast cancer and diabetes are downloaded from UCI repository. Supervised classification algorithms are applied to predict the data to diagnosis the disease. The experiment results are obtained using Jupyter Notebook. Diabetes dataset contains attribute as Age, Gender, Polyuria, Sudden weight loss, Weakness, Polyphagia, Genital thrush, visual blurring, itching, irritability, delayed healing, musclestiffness, Alopecia and outcome. Here outcome attribute used as a target variable to predict whether the patient have diabetes or not. Likewise breast cancer dataset contains attribute radius_mean, perimeter_mean, area_mean, smoothness_mean, compactness_mean, concavity_mean, concave_points_mean, symmetry_mean, etc. Diagnosis attribute used as a prediction variable.

5. Results and Discussion

The dataset is given as input to the classifier models such as Naïve Bayes, K nearest neighbour, Support Vector Machine, Decision Tree and Random Forest. All models are evaluated in terms of performance metric. The metrics are calculated as mentioned in Table 1.

Table 1. Mathematical Formula for Performance Metrics

Metrics	Formula
Accuracy	$Accuracy = \dfrac{TP + TN}{TP + TN + Fp + FN}$
Sensitivity	$Sensitivity = \dfrac{TP}{TP + FN}$
Specificity	$Specificity = \dfrac{TN}{TN + FP}$
Precision	$Precision = \dfrac{TP}{TP + FP}$
F_score	$F\_score = \dfrac{2 * Precision * Sensitivity}{Precision + Sensitivity}$

In Table 1, TP →True Positive, TN→True Negative, FP→False Positive and FN→False Negative .

Table 2, performance metrics are calculated from confusion matrix obtained from different classifier models through Jupyter platform. From the performance, Decision Tree and Random Forest performs well compared to other classifiers. Random forest and decision tree shows 99% of accuracy, KNN shows 98% and Linear regression model shows 97%, Naïve Bayes algorithm shows 94% and support vector machine shows 68% of accuracy. Other measures are shown in Table 2. Fig 2 shows the performance metrics value of bench marked dataset.

Table 2. Performance Analysis for Benchmarked Datasets

Dataset	Performance Measures in %		
	NB	KNN	SVM

	Acc	Sen	Spe	Rec	Fsc	Acc	Sen	Spe	Rec	Fsc	Acc	Sen	Spe	Rec	Fsc
Breast Cancer	94	83	95	66	74	98	86	100	97	91	68	100	95	100	100
Diabetes	85	99	80	68	80	91	77	97	91	83	89	62	100	100	77

Dataset	Performance Measures in %														
	DT					LR					RF				
	Acc	Sen	Spe	Rec	Fsc	Acc	Sen	Spe	Rec	Fsc	Acc	Sen	Spe	Rec	Fsc
Breast Cancer	99	89	100	78	93	97	71	100	94	81	99	94	100	100	97
Diabetes	95	94	95	89	92	89	64	100	100	78	98	100	97	95	97

Fig 2.
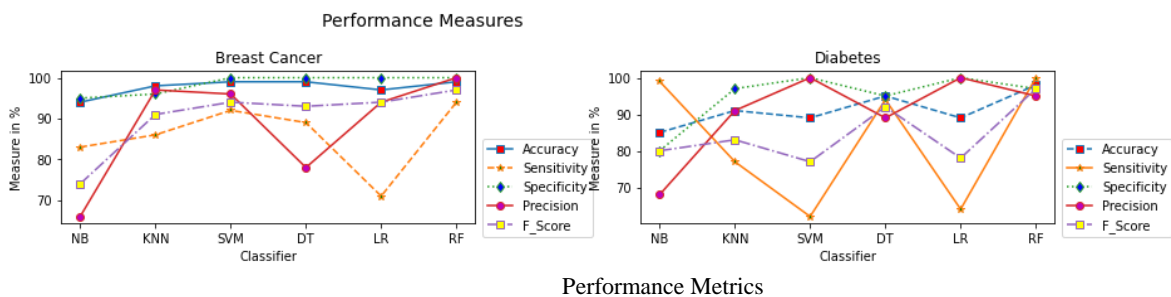
Performance Metrics

ConConclusion

This paper shows how the classification algorithm performs on disease diagnosis system. Different classifier models are evaluated using different performance metrics on benchmarked datasets. Random forest and Decision tree classification algorithms are outperformed than other models and shows 99% of accuracy. Feature selection is the process not only to reduce dataset it also helps to obtain most relevant data for prediction process. In future various feature selection algorithm are applied in prior to the classification process to improve the prediction results.

References

[1] Takanori Yamashita, Yoshifumi Wakata, Hideki Nakaguma et al., " Machine learning for Classification of Postoperative Patient Status using Standartized, Medical Data", Computer Methods and Program in Biomedicine 214 (2022) 106583.

[2] Jobeda Jamal Khanam & Simon Y. Foo," A Comparison of Machine Learning Algorithms for Diabetes Prediction", The Korean Institute of Communications and Information Sciences, ICT Express 7 (2021) 432-439, 2021.

[3] Nazin Ahmed, Rayhan Ahammed, Md. Manowarul Islam et al.," Machine Learning based Diabetes Prediction and Development of Smart Web Application", International Journal of Cognitive Computing in Engineering 2 (2021) 229-241,2021.

[4] Md. Mahmudul Hasan, Gary J. Young, Mehul Rakeshkumar Patel, " A Machine Learning Framework to Predict the Risk of Opioid use Disorder", Machine Learning with Applications 6 (2021) 100144, 2021.

[5] Beibei Shi, Hua Ye, Ali Asghar Heidari et al.," Analysis of COVID-19 Severity from the Perspective of Coagulation Index using Evolutionary Machine Learning with Enhanced Brain Strom Optimization", Journal of King Saud University- Computer and Information Science, 2021.

[6] Lerina Averano, Mario Luca Bernardi, Marta Climitile et al.," Thyroid Disease Treatment Prediction with Machine Learning Approach", Procedia Computer Science 192 (2021) 1031-1040, 2021.

[7] Rachael Hagan, Charles J.Gillan and Fiona Mallett, "Comparison of Machine Learning methods for the Classification of Cardiovascular Disease", Informatics in Medicine Unlocked 24 (2021) 100606, 2021.

[8] Xiaolin Diao, Yanni Huo Shuai Zhao et al., "Automated ICD Coding for Primary Diagnosis via Clinically Interpretable Machine Learning", International Journal of Medical Informatics 153 (2021) 104543, 2021.

[9] Mohammed Amine Naji, Sanaa El Filali et al.," Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis", Procedia Computer Science 191 (2021)487-492, 2021.

[10] Amelec Viloria, Yaneth Herazo-Beltran, Danelys Cabrera and Omar Bonerge Pineda, "Diabetes Diagnostic Prediction Using Vector Support Machine", Procedia Computer Science 170 (2020) 376-381, 2020.

[11] Kaur H, Kumari V. Predictive modelling and analytics for diabetes using a machine learning approach. Appl. Comput. Inform. 2019. https://doi.org/10.1016/j. aci.2018.12.004.

[12] Saritas MM, Yasar A. Performance analysis of ANN and Naive Bayes classification algorithm for data classification. Int. J. Intelligent Syst. Appl. Eng. 2019;7(2): 88–91

[13] Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. CANCER GENOMICS PROTEOMICS 2018;15(1):41–51.

[14] J.J. Khanam,S.Y,Foo,"Comaparison of Machine learning Algorithms for Diabetes Prediction", ICT Express,2021.

[15] H. Lyu, M. Wan, J. Han, R. Liu, C. Wang, A filter feature selection method based on the maximal information coefficient and gram-Schmidt Orthogonalization for biomedical data mining, Comput. Biol. Med. 89 (2020) 264–274.

Phishing Website Detection Techniques Comparison and Implementation

B. Karthikeyan, Dinesh Kumar P

Department of Computer Science, Bishop Heber Collge, Tiruchirappalli.

karthikeyan.cs@bhc.edu.in

Abstract

Phishing attack is a handiest manner to achieve sensitive statistics from harmless users. Aim of the phishers is to collect vital statistics like username, password and bank account details. Cyber safety people are actually searching for sincere and consistent detection strategies for phishing web sites detection. This paper offers with gadget mastering era for detection of phishing URLs via way of means of extracting and reading numerous functions of valid and phishing URLs. Therefore, Decision Tree, random forest, Autoencoder neural network, XG Boost, Multilayer perceptron and Support vector machine algorithms are used to detect phishing websites. Aim of the paper is to stumble on phishing URLs in addition to slim down to best machine learning algorithm by comparing accuracy rate, fake fine and fake bad charge of every set of rules.

Keywords: Phishing, Machine learning, Detection

1. Introduction

A phishing internet site is a not unusual place social engineering approach that mimics trustful uniform useful resource locators (URLs) and web pages. The goal of this mission is to teach system getting to know fashions and deep neural nets at the dataset created to expect phishing web sites. Both phishing and benign URLs of web sites are collected to shape a dataset and from them required URL and internet site content-primarily based totally functions are extracted. The overall performance stage of every version is measures and compared.

2. Data Collection

The set of phishing URLs are accrued from open source carrier known as Phish Tank. This carrier offer a fixed of phishing URLs in more than one code like csv, json etc. that gets updated hourly.

The URLs of phishing websites were collected from www.phishtank.com/developer_info.php. From this dataset, 5000 random phishing URLs are collected to train the ML models. This dataset has a set of benign, spam, phishing, malware & defacement URLs. Out of these types of types, the benign url dataset is taken into consideration for this project. From this dataset, 5000 random valid URLs are accumulated to teach the ML models.

4. Literature Review

Rishikesh Mahajan *et al.*[1] presented in this paper aims to enhance detection method to detect phishing websites using machine learning technology. We achieved 97.14% detection accuracy using random forest algorithm with lowest false positive rate. Also result shows that classifiers give better performance when we used more data as training data. In future hybrid technology will be implemented to detect phishing websites more accurately, for

which random forest algorithm of machine learning technology and blacklist method will be used.

Taskin Kavzoglu *et al.*[2] presented this paper and purpose of this study was to compare the efficiency of three machine learning classifiers, namely RF, SVM and DT using freely available Sentinel-2A imagery covering part of Istanbul, Turkey. In this context, pixel-based classification process was applied to the datasets consisting of different band combinations to delineate the effect of spectral bands. Some important findings were revealed using different datasets and methods. Firstly, it should be noted that Dataset-III was more informative and sufficient for classification, compared with Dataset-I and Dataset-II.

Arun Kulkarnil *et al.*[3] presented In this work, we implemented four classifiers using MATLAB scripts, which are the decision tree, Naïve Bayes' classifier, Support Vector Machine (SVM), and the Neural Network. The classifiers were used to detect phishing URLs. In detecting phishing URLs, there are two steps. The first step is to extract features from the URLs, and the second step is to classify URLs using the model that has been developed with the help of the training set data. In this work, we used the data set that provided the extracted features. The data set, from The University of California, Irvine Machine Learning Repository, contained nine features.

4. Feature Extraction

We have applied python software to extract capabilities from URL. Below are the capabilities that we've extracted for detection of phishing URLs.

4.1. URL with @ symbol

This involves using @ symbol in the link/URL of a website. This @ symbol is used to redirect traffic to a phishing site whose domain name immediately followed the @ symbol. For example, http//amazon.com@gatewayamazon.com will direct a user to Gate way amazon instead of Amazon. The @ symbol usually comes with a shorter domain name unlike some other symbols such as "-" or ".".Therefore, if URL contains @ symbol then phishing otherwise legitimate.

4.2. Using the IPaddress as URL

This process involves the use of IP address to take place of the domain name of a website. Usually, this method is very usual for hiding the original details of a domain name. Therefore, such Internet Protocol address usually denotes phishing or suspicious domain If URL contains Domain path as IPaddress then Phishing, otherwise legitimate.

4.3. URL with Hexadecimal Character Code

Phishers commonly mask phishing URLs by using hexadecimal code to denote the numbers in the IP address. Each hexadecimal code normaly begins with a "%" symbol. For example, http://donefe.000webhostapp.com/auto%20ferify/mail.php which was reported in PhisTank used the hexadecimal Character Code If URL contains hexadecimal character then phishing, otherwise legitimate.

4.4. **URL Length**

This includes getting a URL length which is more than 54 characters. For example, http//womenincoachingsuccess.com is found on the Alexa website database as a benign URL. A detailed observation of the Alexa database notifies that any length of more than 54 is most probably to be phishing. If URL length lesser than 54 then legit otherwise, phishing.

4.5. **URL with multiple"//"**

This includes the using of more than one "//" in the domain name path of a link. A search queryonthe1millionAlexadatabaseof benign URLs in a .csv excel format returns0 for this feature. If URL contains multiple"//"its phishing Otherwise it is legitimate.

5. Machine Learning Algorithm

Six machine learning classification model Decision Tree, Random forest, Support vector, Multilayer Perceptrons, XGBoost and Autoencoder Neural Network machine has been selected to detect phishing websites.

5.1 Decision Tree Algorithm

One of the most widely used algorithm in machine learning technology. Decision tree algorithm is easy to understand and also easy to implement. Decision tree begins its work by choosing best splitter from the available attributes for classification which is considered as a root of the tree. Algorithm continues to build tree until it finds the leaf node. Decision tree creates training model which is used to predict target value or class in tree representation each internal node of the tree belongs to attribute and each leaf node of the tree belongs to class label. In decision tree algorithm, gini index and information gain methods are used to calculate these nodes.

Information Gain= Entropy(S)- [(Weighted Avg) *Entropy(each feature)]

Gini Index= $1- \sum_j P_j^2$ ….1
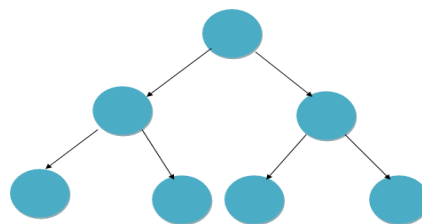
Fig. 1. Decision Tree

5.2 Random Forest Algorithm

Random forest algorithm is one of the most powerful algorithms in machine learning technology and it is based on concept of decision tree algorithm. Random forest algorithm creates the forest with number of decision trees. High number of tree gives high detection accuracy. Creation of trees is based on bootstrap method.

$$Gini\ Index\ =\ 1\ -\ \sum_{i=1}^{n}\left(P_i\right)^2$$
$$=\ 1\ -\ [(P_+)^2 + (P_-)^2]$$

….2
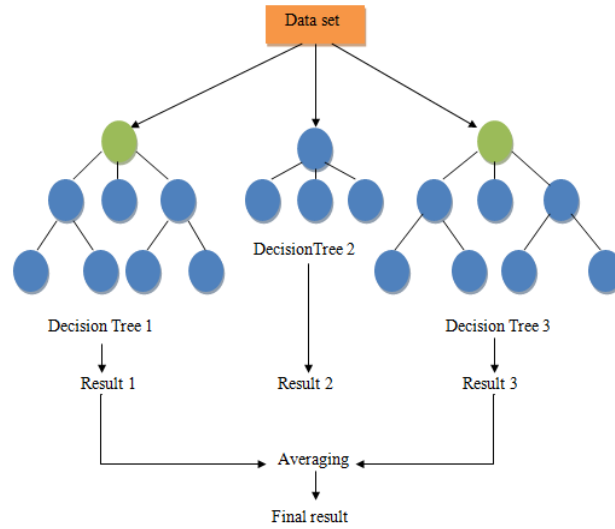
Fig. 2. Random forest

5.3 Support Vector Machine Algorithm

Support vector machine is another powerful algorithm in machine learning technology. In support vector machine algorithm each data item is plotted as a point in n-dimensional space and support vector machine algorithm constructs separating line for classification of two classes, this separating line is well known as hyper plane.

$$R(\alpha) = R_{train}(\alpha)\ +\ \sqrt{\frac{f(h)}{N}}$$

$$f(h) = h\ +\ h\log(2N)\ -\ h\log(h)\ -\ c$$

$$Margin = \rho$$

$$Relative\ Margin = \frac{\rho}{D}$$

$$h \leq\ \min(\left\{d,\ \lceil\frac{D^2}{\rho^2}\rceil\right\})\ +\ 1$$

…...3
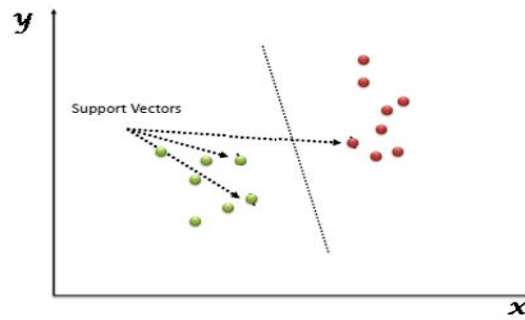
Fig. 3. Support Vector Machine

5.4   Multilayer Perceptrons

Multilayer perceptrons (MLPs) are also known as feed-forward neural networks, or sometimes just neural networks. Multilayer perceptrons can be applied for both classification and regression problems. MLPs can be viewed as generalizations of linear models that perform multiple stages of processing to come to a decision.

$$a = \phi\left(\sum_j w_j x_j + b\right),$$

.....4
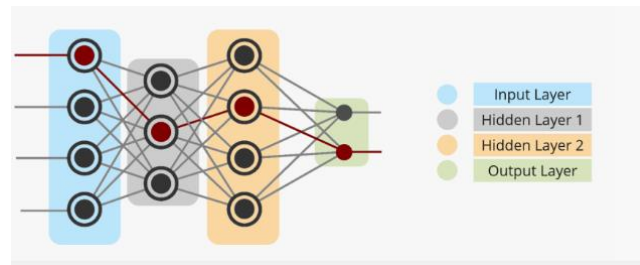
Fig. 4. MLPs

5.5   XG Boost

XG Boost stands for eXtreme Gradient Boosting. It is an application of gradient boosted decision trees, which is intended for its speed and performance. Boosting is an ensemble learning method where advanced techniques are included in order to rectify the errors made by the already proposed models. Models are included consecutively till we find that no additional enhancement can be carried out. While adding new models it uses a gradient descent technique to minimize the loss. The application of this algorithm is to provide efficient computational time and memory supplies. The aim of this design was to produce the best necessity of the accessible sources to train the model. Execution Speed and Model Performance are the two main reasons to work with XG Boost. This approach can support both classification and regression models.

$$obj(\theta) = \sum_i^n l(y_i - \hat{y}_i) + \sum_{j=1}^{j} \Omega(f_j)$$

….5
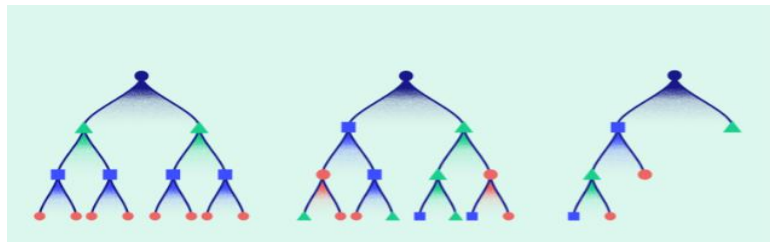
Fig. 5. XG Boost

5.6 Autoencoder Neural Network

An auto encoder is a neural network that has the same number of input neurons as it does outputs. The hidden layers of the neural network will have fewer neurons than the input/output neurons. Because there are fewer neurons, the auto-encoder must learn to encode the input to the fewer hidden neurons. The input layer(I), Output layer(Z), Encoder(X) and Decoder(X)They are exactly the same in an auto encoder.
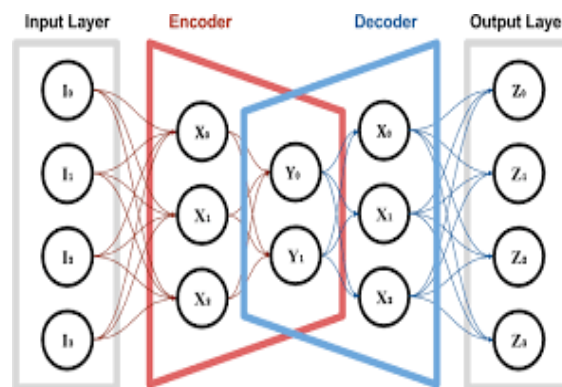
Fig. 6. Autoencoder Neural Network

6. Results

6.1 Detection Accuracy Comparison

Table 1. Detection accuracy comparison

Algorithm	Train Accuracy	Test Accuracy
Decision Tree	0.814	0.804
Random Forest	0.817	0.81
Multilayer Perceptrons	0.866	0.85
XGBoost	**0.87**	**0.85**
Autoencoder Neural Network	0.023	0.019
Support Vector Machines	0.803	0.799
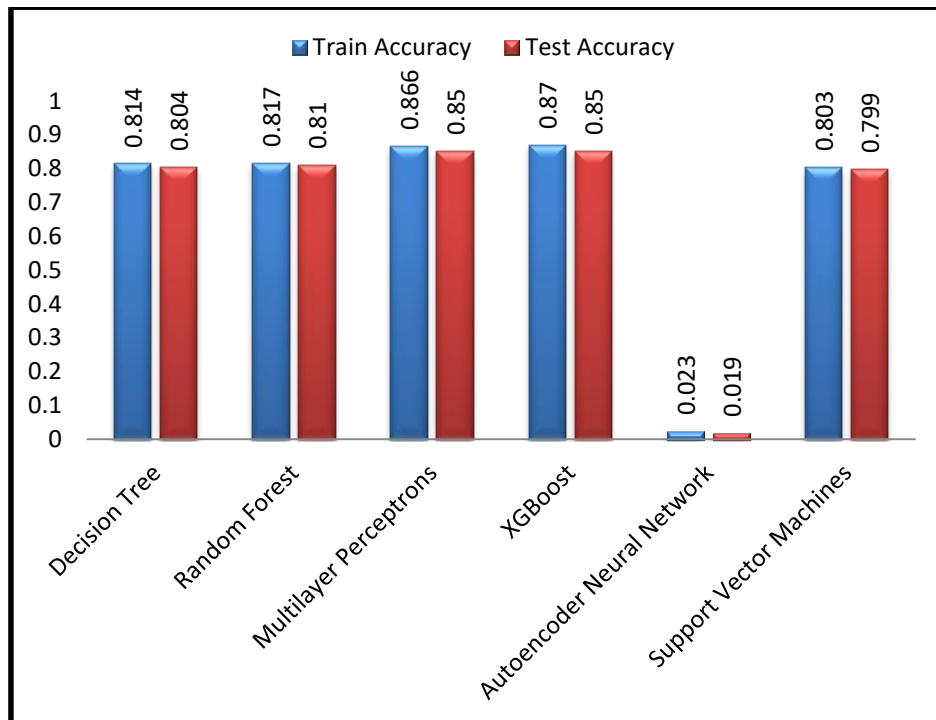
Fig. 7. Comparison of model

References

[1] A.D. Kulkarni, "*Generating classification rules from training samples*",　International Journal of Advanced Computer Science Applications,　vol 9, no. 6,　pp 1-6.

[2] R. P. Lippman, "*An introduction to computing with neural nets*".　IEEE ASSP Magazine, vol. 3 n,o. 4, pp. 4-22, 1987.

[3] Manning, R., & Aaron, G. "*Phishing Activity Trends Report. Anti-Phishing Work Group*",　Tech. Rep. 1st -3rd Quarter.(2015).

[4] Lungu, I., & Tabusca, A. "*Optimizing anti-phishing solutions based on user awareness, education and the use of the latest web security solutions*". Informatica Economica, 14(2), 27. (2010).

[5] Adnan, M.N., & Islam, "*M.Z. Forest PA: Constructing a decision forest by penalizing attributes used in previous trees*",Expert Systems with Applications, 89, 389-403. (2017).

[6] Ahmed, N.S.S., Acharjya, D., & Sanyal, S. "*A framework for phishing attack identification using rough set and formal concept analysis*". International Journal of Communication Networks and Distributed Systems, 18(2), 186-212. (2017).

A Review on Stress Management Intervention on Women's Working Sector with Multimodal Approach

Ms. G.Mahalakshmi[1], Dr. G.Sujatha[2]

PG and Research Department of Computer Science,

Sri Meenakshi Government Arts College for Women, Madurai, India.

jeyzhere@gmail.com

Abstract

Stress management is a pre-planned strategy for dealing with psychological or emotional upheaval. Women in society today have demonstrated their expertise, proficiency, and talents in a variety of fields. Women's roles in our society have shifted dramatically over the last three decades. Women and girls have more opportunities and face a variety of challenges. Women of great prominence in government, business, sports, and the media have become commonplace. The traditional role of women is to be responsible for the efficient operation of her home; this is something we should be aware of as an additional expectation placed on her. Working women have dual roles to play in their families and at work. As a result, working women must prepare for the upcoming challenges in which they must justify their roles in both places, which causes them a great deal of stress.In order to live a happy and successful life, every working woman should learn stress management skills. As a result, the purpose of this research is to provide an overview of the current research in this topic and to explore some of them in further depth. This paper can assist aspiring scientists in identifying a research problem in this field.

Keywords: Stress management, expertise, proficiency, challenges, government, business, sports, media, families, work, successful life.

1. Introduction

Word Stress in Cameroon English, Stress is a general term applied to various psychological and physiologic pressures experienced or felt by people throughout their lives. Stress is a wide phenomenon as it seems to have universal spread. Stress exists in human beings, animals and even in metals. Stress is also related with providing quality services to human beings depend upon their stress levels. However, this study is about women stress with special focus on the work stress in hospitals. Many people think they understand stress, but in reality stress is complex and often misunderstood.

1.1 Stress Management

Stress management consists of making changes to life if women are in a constant stressful situation preventing stress by relaxation. Stress is actually a survival response when our body thinks that it is in danger. Every woman has a different journey when getting to the root cause of stress. Symptoms of stress may include an increased heart rate, sleep disturbances, blood pressure, irritability, headaches, or muscle tension etc. Today's life is so busy and complex, Stress can manifest itself in both a positive and a negative way.

Positive Stress

Stress is said to be positive when the situation offers an opportunity for one to gain something. Eustress means beneficial stress-either psychological, physical or biochemical, radiological. Eustress is often viewed as a motivator since in its absence the individual lacks the edge necessary for peak performance. Positive stress gives the sprit to achieve and induces us to succeed are attain the next level of career opportunities and to get more financial benefits, these positive stress helps women.

Negative Stress

Stress affects individual physiologically, emotionally and behaviorally and it is linked to several health problems, the process of physiological stress response starts from the moment the body realizes the presence of the stressor, followed by the sending of signals to the brain, and to the specific sympathetic and hormonal responses to eliminate, reduce or cope with the stress Sympathetic Response surely likes things to go very quickly when you are faced with stress that The neurotransmitter is released by the nerve endings and is sent to the SNS Enhance the strength of your skeletal muscles.

During stressful moments, your heart beats faster than it usually does so that the parts of your body which are needed to cope up with the stress would be supplied by enough oxygenated blood to remain functional until the stressful situation subsides. Shoot up sugar and fat levels. We all know that sugar and fat provides our body with energy. During stressful situations, we need more energy to cope up, and so the SNS assists us to have more energy

Emotional effects of stress can range from emotional overeating to a feeling of being overwhelmed and pressured. It can increase the incidence of frustration, feelings of a lack of purpose in life, or depression and anxiety. People often complain of poorer concentration, forgetfulness, or have difficulty making decisions when stressed. The stress that you are feeling can be bottled up waiting to express itself through increased anger and frustration. If you respond to stress through angry outbursts then these anger management tips could be beneficial to you.

1.2 Structure of the Paper

The section 2 discusses about the need and significance of the study. The datasets utilised in this investigation are discussed in Section 3. The Section 4 deals with the Health problems due to stress and section 5 deals with the various stress assessment techniques utilized by the other authors using machine learning section 6 deals with the stress assessment techniques using image and video processing section 7 deals with the stress assessment techniques using deep learning and section 8 deals with the Findings of the study and finally section 9 concludes the paper.

2. Need and Significance of the Study

In the wake of rapid social change in various aspects the role and position of women artundergoing changes at a rapid pace. Many women are employed and manage both marital life and career. They take up nontraditional roles and have developed a new outlook of life. These days, women have become more conscious of their own identity and status. Modem women know their self-worth and they wish to develop self-reliance, independence and self-respect by taking up jobs in various disciplines. Many researchers over the past years have made comparative studies of employed and unemployed women on various psychological variables and the

relationships between them. Women Empowerment refers to increasing the spiritual, political, social.   most important areas in to which women have to cope up. As a summary

- To find out root causes of stress in your personal life and professional life.
- To examine the soundness of the methods followed in reducing the stress.
- Learn healthy ways to relieve stress and reduce its harmful effects.
- Examine their influence on the existing pattern of stress management.
- Who are the more effecting person in the family?
- How they are balancing family and professional as women.

2.1 Statement of the Problem

Stress Management: Stress management can be defined as interventions designed to reduce the impact of stressors that an individual faces in his/ her life. These can have an individual focus, aimed at increasing an individual's ability to cope with stressors. In the present study, stress management refers to the various stressful situations that come across in the life of a woman, and the approach in which she is able to overcome them even in the midst of multiple responsibilities.

2.2 Objectives of the Study

- To study the factors affecting stress on women.
- To know the level of stress on women
- To know the health issues due to stress.

3. Datasets

[1] proposed a DBN-based stress classification model that uses stress-related physical activity and lifestyle data obtained from the 2013–2015 Korea National Health and Nutrition Examination Survey (KNHANES VI) database.

[2] considered OSMI Mental Health in Tech 2017 survey as the dataset, using which they trained different machine learning models in order to analyze the patterns of stress and mental health disorders among tech professionals and to determine the most influential factors that contribute to the same.

[3] considered the standard diagnosis criterion for depression is the Patient Health Questionnaire (PHQ), whilst the Depression, Anxiety and Stress Scale (DASS 21), which has 21 questions, is used for screening the symptoms related to these mental illnesses.

The present study [4] is a questionnaire consists of multiple choices and it has distributed to employees of all working women's in area of Secunderabad, Telangana State. The sample size is 100. This has an important bearing on the collection of reliable and accurate information as well as on the outcome of the study. The collection of opinions of employees constituted the survey method in the study.

In this paper [5], the researcher were calculating the mental stress of students one week before the exam and during the usage of the internet. The objective is to analyze stress in the college students at different points in his life. The effect that exam pressure or recruitments stress has on the student which often goes unnoticed. The researcher performs an analysis on how these factors affect the mind of a student and will also correlate this stress with the time spent on the internet. The dataset was taken from Jaypee Institute of Information

Technology, and it consisted of 206 student's data.

4. Health Problem Due to Stress

32.5% of women have health problem due to stress and they strongly agree 32.5% of women have said agree health issue due to stress, 15% of them reported neutral, 15% of respondents reported disagree ,5% of have health problem said strongly disagree, Major respondents of 65% of women have health issues caused by stress .

According to the National Women's Health Information Center, the effects of stress on women's physical and emotional health can range from headaches to irritable bowel syndrome. Specific stress effects include:
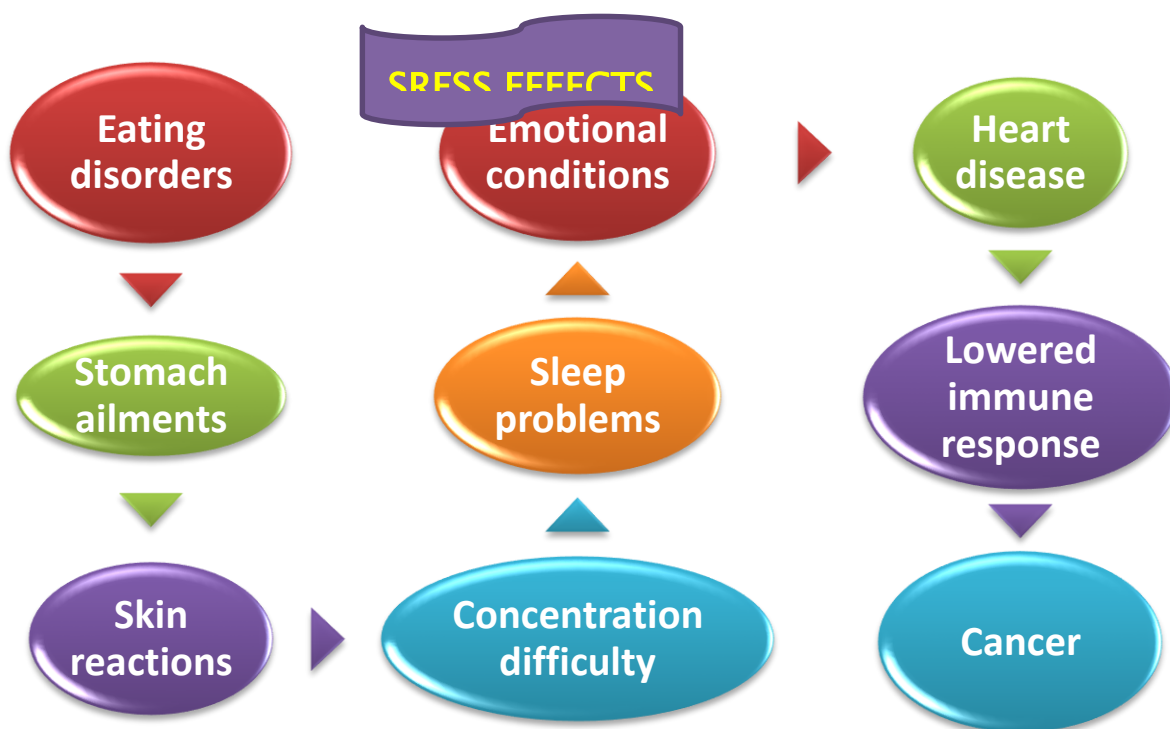
- **Eating disorders:** This may have something to do with stress levels. Like depression, this illness has been linked to low levels of serotonin and is often treated with serotonin-boosting antidepressant drugs.
- **Stomach ailments**: Stress can make you reach for junk or comfort foods, or upset your stomach to the point that you feel like you can't eat. Common stress-related stomach troubles include cramps, bloating, heartburn, and even irritable bowel syndrome. Depending on how you respond, these can lead to weight loss or weight gain.
- **Skin reactions**: Stress can lead to breakouts and even itchy rashes and hives in some people.
- **Emotional conditions:** From being in a blue or irritable mood to more serious mental issues like depression, your emotional health suffers when there's stress in your life. Women are better than men at hiding some emotions like anger and aggressiveness because the parts of their brains responsible for these emotions are larger than men's, but depression strikes women twice as often as

men, says Rosch, adding, "The emotional effects of stress on women can range from postpartum depression after pregnancy to depression after menopause."

- **Sleep problems:** Trouble falling or staying asleep is common in women affected by stress, and this is particularly counterproductive since a good night's sleep can help ease stress.

- **Concentration difficulty:** Stress makes it hard to focus and be effective in your responsibilities at home or work, and that can compound your problems if your stress comes from your job to begin with.

- **Heart disease:** The stress of competing in today's job market has increased women's heart disease risk, Rosch says. Stress can negatively affect the entire cardiovascular system, and lead to high blood pressure, stroke, and heart attack.

- **Lowered immune response:** One of the more complicated physical reactions to stress is your body's lessened ability to fight off disease, whether it's a cold or a flare-up of a chronic condition.

- **Cancer:** Some studies have suggested a link between stress and the development of breast and ovarian cancer. In one study, researchers found that the risk of breast cancer was increased by 62 percent in women

5. Stress Assessment Techniques Using Machine Learning

5.1 Traditional Methods of Stress/Relaxation Measurement

Stress has been traditionally measured by some indicative parameters such as heartrates, galvanic skin response, pupil diameter and another method are Questionnaires that help to find a person prone to stress, some life events also helps to detect stress. But, these traditional methods require continuous observation or assessments or some expensive sensors and also have to believe that the person is sharing correct answers of his or her mindset. And also not lying to make a better image in front of psychiatrists. There is a relationship between personality and psychological stress of a person

5.1.1   Stress Assessment Techniques Employed

Day to day job demands lead to psychological andphysiological strain among individuals. Every individualexperiences strain over time and it is likely to depend on the perceived and real consequences of how they cope with stress. Many personal and environmental resources are involved in coping with stress [8].Many studies have been conducted in this field. AlirezaBolhariet al. (2012) have studied workplace stress [9]. Jong-Ho Kim et al. (1992) investigated the social life ofcollege students and highlighted the impact of physicalexercises in reducing stress levels [10], [11].

EnriqueGarcia-Ceja et al. (2016) used smartphones as a potential toolto detect behavior that is correlated with stress levels [12]. PanagiotisKostopouloset al. (2016) designed a stressdetection system, StayActive, which uses sleep patterns,physical activities and social interactions to detect stress [13].Mario Salaiet al. (2016) have worked on automatic stressdetection by measuring the heart rate variability (HRV)usinga low cost heart rate sensor and chest belt. They used the galvanic skin response (GSR), electromyography (EMG),skin temperature, electrocardiography (ECG) and skinconductance as the indicators of stress [14], [15]. Vanithet al (2014) used a hierarchical classifier to detect mental stress inhumans using the HRV measured using ECG [16]. Padmaja etal. (2016) worked on human stress using a TreeNet classifierbased on

smartphone interaction data and social features [17]. Gimpelet al. (2015) developed the application, mystress, as astress detection system based on hardware and softwaresensor data collected using an android application. Hong Luet al. (2012) developed an application called, stress sense, ahuman voice based stress detection system using real-lifeconversations on smartphones [18], [19].

Besides, stress has negative impact on public health and itplays a major role in behavioral disorders such as depressionand anxiety [20]. Oscar Martinez Mozos et al. (2017) workedon a machine learning approach for stress detection bycombining physiological and sociometric sensors and itaccurately discriminated the stressfuland neutral situations ofan individual [21]. Martin Gjoreskiet al. (2016) used a machine learning approach for continuous stress detectionusing a wearable wrist device using the objective andsubjective labelingof laboratory and real-life data as well as physiological measurements [22]

Table 1. An overview of the related work on stress detection systems

Related work	Focus group	Data sources	Reference	Year
AlirezaBolhariet al.	Occupational stress	Questionnaire	[9]	2012
Jong-Ho Kim et al.	physical exercise onstress	Interviews and questionnaire	[11]	2014
Enrique Garcia-Ceja et al.	Occupational stress	Smartphone sensor	[12]	2015
Gimpelet al.	Individuals	Questionnaire, sensors	[15]	2015
Martin Gjoreskiet al.	Perceived stress	Questionnaire, sensors	[21]	2015
Martin Gjoreskiet al.	Stress in adults	Questionnaire, wrist device	[22]	2016
Mario Salaiet al	Stress in volunteers	Chest belt, heart rate sensor	[14]	2016
Kostopouloset al.	Stress in professionals	Smartphone sensor data, sleeping pattern	[13]	2016
Buddi et al.	Stress in working professionals	Questionnaire, Fitbit data	[23]	2017

5.2 Identification of Stress and Relaxation from Social Media Content Moving to Research on Social Media

As now we are in the upscaling generation of technology, always try to share our ideas and thoughts on social media by posting a status or comments to the particular or current topic of the day to day activities which helps to judge a person is "Stressed" or "Relaxed".

Tweet level emotion detection reflects the instant emotions expressed by a particular tweet which also reflect mental health disorder such as depression or post-traumatic stress disorder (PTSD).

In paper[24] the author Mike Thelwall implemented a WSD technology as a pre-processing stage and further works on lexicon based stress or relaxation method which improves the accuracy of TensiStrength. In this used the dataset of „1000' tweets with word "Fine" which reflects ambiguous meaning in different situation of sentences. This paper also eliminates redundant words such as preposition, conjunction, interjections and articles from the tweet and then work on remaining words and categories them in -5 to +5 based on dictionary indication

The paper[25] studies the co-relation between user psychological stress states and their social interactions defined by a unified hybrid model that integrates a factor graph model with convolutional neural networks. This is because CNN is capable of learning a unified latent feature from multiple modalities and FGM is good at modelling the co-relations. In this designed a CNN with cross autoencoder (CAE) to generate user-level content attributes from tweet level attributes and a partially labelled factor graph to combine user level social interactions to detect stress. Used the dataset of Twitter and SinaWeibo to compare the result for better accuracy also work on some comparison methods such as logistic regression, SVM, gradient boosted decision tree, deep neural network. So, like this presented a framework for detecting stress psychological stress states from user weekly social media data but the limitation is to find users who are stressed but less connected to social media.

The paper[26] focuses on social media to detect and diagnose depression in an individual, social media has characterising features of a person that may be positive or negative. In this twitter is used for measuring and predicting depression. Firstly they collected data from crowdsourcing to detect user and report they have been diagnosed as in a depression. Crowd workers can also have an option into their Twitter profile, with an agreement that their data could be mined and analysed using the computer program. Here questionnaire as the primary tool to determine the depression levels of the crowd workers. Here compare the behaviour of the depressed user class and standard user class and put some conclusions that individuals with depression show lower social activities, greater negative emotions, high self-attentional focus, etc. Worked on SVM classifier to predict depression of an individual user showing depression signs tend to be active during evening and night.

It is observed that in paper[27] the use of social media is for self-expression in the form of daily life events or updates, information sharing on favourite subject and comments of thoughts on a discussed topic and all these things are replacing face to face communication of real life. By using the conversation on twitter anyone can work on stress detection domain. In this survey paper, various comparisons are stated on current work. Various methods are available to detect stress but very few works on stress prediction. And as on the ongoing work, there is a great hope for sentiment analysis and prediction of stress.

The author Andrew Trask [28] proposed a new model for WSD which uses supervised NLP labelling of data. This uses supervised labelling method instead of unsupervised clustering method which analyse the ambiguous work and assign or label the specific work as per its senses. Because of this accuracy, can be increased and also reduces computational complexity. In the proposed system[29], sentiment analysis is done by applying both lexicon based TensiStrength and machine learning algorithm to detect stress.

5.3 Various Studies Conducted in Different Areas with Different Techniques

In Paper [30], Stress has been a havoc identified in software engineers. The authors have incorporated the OSMI (open sourcing Mental Illness) Survey dataset 2017 from tech industry. This dataset has been subjected to various machine learning strategies and the dataset hold the labels such as gender, age, family background etc. Their findings suggested that 75% of the people who worked in the IT industry are prone to pressure. Few of the techniques which they have worked on include boosting, packing, decision trees etc.

In Paper [31], authors have used decision tree algorithm which is applied on the data of students enrolled for an academic year and their stress levels are recorded at the beginning and at the end of the semester. The result is such that the model identifies students with more stress at the semester end, rather than at the beginning of the semester.

In Paper [32], authors have measured stress using different modes like EEG, GSR, EMG and SpO2.

These were used to record or measure the automated stress detection. The measurement recorded from all the external and internal sensors were verified with the value of index which was initially set as a mark for stress prediction.

In Paper [33] focuses on the occupational stress. A survey was conducted at different sectors to collect the data. It focused on three factors, i.e., psychosocial, environmental, and physical factors. Analysis is done using metric (Support Vector Machine and neural Network) and non- metric approaches. The metric approaches provide good accuracy.

In Paper [34] uses CNN-UDRP algorithm which is Convolution Neural Network based on Unimodal disease risk prediction (CNN-UDRP) which is predicting the risk of heart diseases. The accuracy of CNN-UDRP for heart disease is approximately 60 – 65%. Hence the performance is better for risk prediction of heart diseases. The result of the system which has been shown are in the form of low, medium as well as high risk.

In Paper [35] focuses on the linguistic characteristics of Facebook and Twitter data. It evaluates how models trained on Facebook data perform at predicting stress on the twitter language in a cross-domain setting. Hence the need for transfer learning is determined to apply the stress model on Twitter language which is trained on Facebook language.

In paper [36], the proposed system is divided in three modules – Sensor (hardware circuit) Device connected to microcontroller, Mobile App for showing the results & Cloud Storage where data is being stored. Electroencephalography(EEG) brain waves are considered as an important parameter for detecting the levels of stress. It identifies stress and alert is provided to the user on a mobile application which is used to help the user for alerting the stress level anytime. The system depends machine learning algorithms which are used as well as on the datasets which are used.

Occupational Stress is an important health-related concern. Paper [37] focuses on predicting the occupational stress. The data is collected by conducting a survey among the people working at different sectors of occupation. The survey focuses on three factors, i.e., psychosocial, environmental, and physical factors. In this paper, the analysis is done by using metric (Support Vector Machine and neural Network) and non- metric (Decision tree and Random Forest) approaches. It concludes that the metric approaches provide good accuracy.

In Paper [38] predicts the comorbid risk of the patients by analyzing the longitudinal EHR data from the other patients by using a novel deep learning framework. It performs the prediction of risk using the data of patients with single chronic disease of interest. With the help of these contributions, it determines the heterogeneous characteristics within the EHR data of the patients. Finally, the proposed model is validated quantitatively and qualitatively on the EHR data of the patients.

In paper [39], Chen proposed CNN-based (CNN-MDRP) which is also called as Multi-modal Disease Risk Prediction algorithm which uses SQL as well as no SQL data (in JSON format) from hospitals in China. Chen concluded that the proposed system for disease risk prediction provides an accuracy of about 94.80% which is faster as compared to CNN-UDRP algorithm.

InPaper [40] focuses on predicting the next level of stress which is based on the current stress level, road or environment conditions and the driving actions of the drivers. The results describe that every user handles the stress is different for other users. It is stated that it is easy to provide the level of istress based on the current behavior of that user and the conditions for a single-user, but these results cannot be generalized in case for the cross-user scenario.

In Paper [41] conducts a study of stress, mobile usage as well as data which is captured from sensors. This study is based on different observations. It makes the use of an Android application The Stress Collector (TSC) to collect the stress data. After installing the application on an Android mobile using (.apk) file, it continuously runs in the background to collect the mobile usage and sensor data periodically. This paper describes significant correlations among stress and smartpone data, and it performs better than the back reported levels. It helps to encourage for further investigations on stress prediction using smartphones.

In Paper [42] considers the interrelated features and the possible effects of the future events by predicting future stress level of the teen age students from the usage of micro-blog. The effectiveness of the methodology is verified by the experiments. The paper also helps to encourage the work towards teenagers future stress level trends.

In Paper [43] presents a system for predicting various stress levels from (ECG – electrocardiogram) data. In this, the Galvanic Skin Response (GSR) signals are being recorded at a 31 Hz rate and ECG signals were continuously collected at a rate of 496 Hz. The paper states that using a blend of Heart Rate Variability and electrocardiogram features, prediction is difficult of GSR if it is in the highest percentile or lowest 20 percentile with >97% accuracy. It suggests that by using consumer ECG devices, we can predict whether the user is stressed or not without the requirement of GSR estimations.

6. Implementation of Stress Detection Using Image and Video Processing

Tong Chen et.al.[44] Introduced an HSI-based technique for the location of mental pressure. The solid material-separating capacity of HSI was used in this investigation to separate and measure the measure of blood chromophores (Hb and HbO2) utilizing the Beer-Lambert Law. The way where HSI signals are acquired (caught picture) portrays this strategy as a without contact pressure location method.

Andra Adams et.al.[45] Investigate the impact of head movement from outward appearance in the view of complex feelings in videos. They show that head movements convey enthusiastic data that is integral as opposed to repetitive to the feeling content in outward appearances. At last, they show that passionate expressivity in head movement isn't restricted to gestures and shakes and that extra motions (for example, head slants, raises and general measure of movement) could be helpful to computerized acknowledgment frameworks.

G. Giannakakis et.al.[46] Builds up a system for the location and investigation of stress/tension enthusiastic states through video-recorded facial signals. An exhaustive test convention was set up to incite orderly changeability in full of feeling states (unbiased, loose and focused/anxious) through an assortment of outside and inner stressors. The examination was centered basically around non-deliberate and semiintentional facial signals to gauge the feeling portrayal all the more impartially. Highlights under scrutiny included eye related occasions, mouth movement, head movement boundaries and pulse assessed through camera-based photoplethysmography. A component choice system was utilized to choose the strongest highlights followed by characterization plans separating between pressure/uneasiness and unbiased states regarding a casual state in each experimental phase. In addition, a ranking transformation was proposed utilizing self-reports in order to investigate the correlation of facial parameters with a participant's perceived amount of stress/anxiety. The results indicated that, specific facial cues, derived from eye activity, mouth activity, head movements and camera-based heart activity achieve good accuracy and are suitable as discriminative indicators of stress and anxiety.

7. **Implementation of Stress Detection Using Deep Learning**

Human stress has been an intermittent topic in research papers due to its increased relevance in modern society. [47] uses Theano, a python framework where stress is detected by the position of the eyebrow from its mean position. CNN's can be used for image classification using images from handwritten MNIST dataset [48]. Face recognition is an interesting technique which extracts meaningful features from an image. OpenCV library's Haar cascade classifier performs face detection with high efficiency [49]. Many studies have been conducted to identify the precise facial characteristics that are associated to depression. A depressed/stressed face has the same characteristics of a sad face which can be classified by SVM classifier [54]. [55] gives a brief introduction about convolutional neural networks developing brilliantly image recognition models.

In Paper [56] A deep learning CNN based on Keras and tensor flow where four different structures of CNN on CPU system are compared with each other. Heart rate variability refers to the beat alterations in the heart rate. Based on heart rate we can predict whether a person is in stressed or not. An IoT device named Remote Stress Detector can be used to identify the stress level of any person using his/her heart-rate [50]. Stress can also be detected by measuring body temperature and skin conductance using sensors, and which analyzes the stress level [53]. With the help of different wearable sensors signals like ECG and breathing rate are measured [52]. [51] This paper presents an emotion detection system for smart phones which can be used as a smart keyboard. The smart keyboard detects a person's emotional state using ML techniques.

Many researchers have applied different machine learning algorithms for the psychological disorders prediction and the performances and results of different algorithms have been found to vary with varying datasets used by them. On this scenario, no fixed algorithm has been determined as most out-performing in all cases. Still the researches are ongoing in this field to explore the appropriate algorithm which finds all the types of stress levels which can be applied to all the fields.

8. **Findings**

- Major factors that affect the stress on women are Work pressure , job co-ordination ,job time at hospital(Night shift),Lack of family support

- Level of stress is high compared to men and married women are more stressed than unmarried women

- Health issues for women are more due to stress like Heart attack ,Increase blood pressure , sleeping disturbance , head ache , muscle tension and hormone imbalance result in Infertility

8.1 Limitations of Machine learning Algorithms

The limitations of Machine learning Algorithms are considered on reviewing the above listed articles. The following are the limitations:

False labelling

1. Labels are incorrect because of some random error. Deep learning algorithms are quite robust to random errors in the training set.

2. In learning is incorrect learning for obvious reasons and will lead to bad performance on data set and test set. Hence, systematic errors should be dealt by either removing or correcting such samples.

3. Mislabelled labels

4. Lack of detecting stress level of a person in different environment. (i.e same person have a different level of stress at different environment and also lack of mining of stress level of another person in same environment).

5. Lack of accuracy compared with DL

8.2 Question Formulation

On reviewing the above articles, the author has formulated the following questions to be answered in the next stage of this process:

1. What are the circumstances a person experiences stress?
2. How to create a uniform stress scale to measure a stress level for an individual?
3. What are all the factors are going to include assessing the stress level?
4. Whether the machine learning algorithms used by the other researchers had predicted the stress levels accurately?
5. Whether the researches had been conducted on one particular gender or the both?
6. What are the problems faced by the women in cause of stress?
7. Why not we use Deep Learning algorithms to predict the stress levels accurately in women?
8. Can we use Multi modal dataset for stress prediction?
9. Whether the researches had been done in all the fields apart from them listed above?

9. Conclusion

Stress is unavoidable now-a-days it can be handled positively by coordination with workers, proper time management, loving the profession, family support , and creating personal hobbies Improve your diet, By eating well-balanced meals and skipping junk food, you can improve your physical well-being and, in turn, your emotional health. Make time for exercise meditation and yoga that exercise is a phenomenal way of dealing with stress and depression, Research shows that getting active can lift your spirits and increase the release of endorphins, a natural chemical associated with mood. Find fun ways to relax Connect with family and friends and people you enjoy being around. Rediscover favorite hobbies like hearing music, gardening, and day out with family or friends.

References

[1] Se-Hui Song, MS, Dong Keun Kim," Development of a Stress Classification Model Using Deep Belief Networks for Stress Monitoring", Healthc Inform Res. 2017 October;23(4):285-292. https://doi.org/10.4258/hir.2017.23.4.285 , eISSN2093-369X

[2] U Srinivasulu Reddy, AdityaVivekThota, A Dharun," Machine Learning Techniques for Stress Prediction in Working Employees", ResearchGate, December   2018,

[3] AnuPriyaa , ShrutiGarga, NehaPrernaTiggaa," Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms", International ConferenceonComputationalIntelligenceandDataScience(ICCIDS2019),

ProcediaComputerScience167(2020)1258–1267

[4] Mrs. G KusumaKumari, Mrs.Dr.Saradadevi, A Study On Stress Management Of Working Women In Twin Cities, IJSDR April2016Volume 1, Issue 4, ISSN: 2455-2631

[5] RavinderAhujaa , Alisha Banga, Mental Stress Detection in University Students using Machine Learning Algorithms, International Conference on Pervasive Computing Advances and Applications – PerCAA 2019, ScienceDirect ,Procedia Computer Science 152 (2019) 349–353

[6] Das Sanjeeb Kumar," Stress Management In Women: A Case Study In Temple City of Bhubaneswar", International Journal of Scientific Research and Reviews, 2018, 7(4), 11-18, ISSN: 2279–0543

[7] K. SUGANTHI AND S.P. VAANMATHI," a study on stress management among women employees in hospital with reference to thanjavur district", Indian J.Sci.Res. 14 (1): 159-162, 2017, ISSN: 2250-0138 (Online)

[8] F. J. McGuigan et al., "Coping with occupational stress among professional engineers,"StressandTensionControl3StressManagement,vol.3,pp. 155-163, Plenum Press, New York. 1989.

[9] A.Bolhari,A.Rezaeian,J.Bolhari,andS.Bairamzadeh,Professionalsin Iran,"International Journal of Information and Electronics Engineering, vol. 2, issue 5, September 2012

[10] R. Norris, D. Carroll, andR. Cochrane, "The effects of physical activity and exercise training on psychological stress and well-being in an adolescent population,"Journal of Psychosomatic Research, pp. 55-65, 1992

[11] J.-H. Kim et al., "The Impacts of physical exercise on stress coping and well- being in university students in the context of leisure,"Journal on Health,vol.6, 2014.

[12] G.-C. Enrique, V.Osmani, and O. Mayora, "Automatic stress Detection in workingenvironmentsfromsmartphones'accelerometerdata:Afirst step,"IEEEJournalofBiomedicalandHealthInformatics,vol.20,issue4, pp. 1053-1060, 2016.

[13] P. Kostopoulos, A. I.Kyritsis, M. Deriaz, and D. Konstantas, "Stress detection using smart phone data," eHealth 3600 International SummitoneHealth,Budapest, Hungary, June 14-16, pp. 340-351, 2016.

[14] Mario Salai, IstvánVassányi, and IstvánKósa, "Stress detection usinglowcost heart rate sensors,"Journal of Healthcare Engineering,pp.1-13, Hindawi Publishing Corporation, 2016.

[15] Gimpeletal.,"myStress:Unobtrusivesmartphone-basedStress detection,"inProc. the 23th European Conference on Information Systems(ECIS 2015), Münster, Germany, May2015.

[16] L.VanithaandG.RSuresh,"HierarchicalSVMtodetectmentalstressin humanbeingsusingheart rate variability,"inProc.2ndInternational Conference on Devices, Circuits and Systems (ICDCS),2014,pp. 1-5.

[17] B Padmajaet al., "TreeNet analysisof human stressbehaviour usingsocio- mobile data,"Journal of big data,vol.3, issue 1, pp. 1-24, 2016.

[18] H. Lu et al., "StressSense: Detecting stress in unconstrained acousticenvironments using smartphones,"in Proc. ACM Conference on Ubiquitous Computing,pp. 351-360, 2012.

[19] S.Cohen,R.C.Kessler,andL. U.Gordon,"Measuringstress:Aguidefor health and social

scientists,"Oxford University Press, USA, 1997.

[20] O. M. Mozos et al., "stress detection using wearable physiological andsociometric sensors," International Journal of Neural Systems, vol. 27, issue 2, 2017

[21] M. Gjoreski et al.,"Continuous stress detection using a wrist device –In laboratory andreal life,"inProc.the2016 ACMInternationalJointConference on Pervasive and Ubiquitous Computing: Adjunct,2016, pp. 1185-1193.

[22] M.Gjoreskiet al.,"Automaticdetection ofperceivedstress in campus students using smartphones,"inProc. International Conference on Intelligent Environments,2015

[23] Buddi, Padmaja& Prasad, V.V. &Sunitha, K.V.N.. (2018). Machine Learning Approach for Stress Detection using Wireless Physical Activity Tracker. International Journal of Machine Learning and Computing. 8. 33-38. 10.18178/ijmlc.2018.8.1.659.

[24] ReshmiGopalakrishnaPillai,MikeThelwall,ConstantinOrasan,"DetectionofStressand Relaxation Magnitudes for Tweets", International World Wide Web Conference Committee ACM, 2018.

[25] Huijie Lin, JiaJia, JiezhonQiu, " Detecting stress based on social interactions in social networks", IEEE Transactions on Knowledge and Data Engineering, 2017.

[26] Munmun De Choudhury, Michael Gamon, Scott Counts, Eric Horvitz, "Predicting DepressionviaSocialMedia",SeventhInternationalAAAI ConferenceonWeblogsand Social Media, 2013.

[27] Dr. G V Garje, ApoorvaInamdar, HarshaMahajan, "Stress Detection And Sentiment Prediction: A Survey", International Journal of Engineering Applied Sciences and Technology, 2016.

[28] Andrew Trask, Phil Michalak, John Liu, "Sense2vec - A Fast And Accurate Method for Word Sense Disambiguation In Neural Word Embeddings", Under review as a conference paper at ICLR 2016.

[29] ReshmaRadheshamjeeBaheti, SupriyaKinariwala," Detection and Analysis of Stress using Machine Learning Techniques", International Journal of Engineering and Advanced Technology (IJEAT) ISSN:2249 – 8958, Volume-9 Issue-1, October 2019

[30] U Srinivasulu Reddy, AdityavivekThota, Adharun, Machine learningTechniques for Stress prediction in working employee 2018 IEEE International Conference on computational intelligence and computing research.

[31] Norizam,Sulaiman. Determination and classification of human stress index using nonparametric analysis of EEG signals. Diss.UniversitiTeknologi MARA, 2015

[32] Xu, Q., Nwe, T.L., Guan, C.. Cluster-based analysis for personalized stress evaluation using physiological signals.

[33]   IEEEJournalofbiomedical andhealthinformatics2015;19(1):275281.

[34] SayaliAmbekar, RashmiPhalnikar, Disease Risk Pre-diction by Using Convolutional Neural Network, IEEE 2018.

[35] Sharath Chandra Guntuku, AnnekeBuffone, KokilJaidka, Johannes C.Eichstaedt, Lyle H. Ungar, Understanding and Measuring Psychological Stress Using Social Media, ICWSM 2019.

[36] ChrandrasekarVuppalapati,MohamadSKhan,NishaRaghu,PriyankaVeleru, Suma Khursheed, A System To Detect Mental Stress Using Machine Learning And Mobile Development, 2019

[37] S.K. Yadav, ArshadHashmi, An Investigation of Occu-pational stress Classification by using Machine Learning Techniques, Vol.-6, Issue-6, 2018.

[38] Jinghe Zhang, Jiaqi Gong, Laura Barnes, HCNN: Hetero-geneous Convolution Neural Networks for Comorbid Risk Prediction with Electronic Health Records, IEEE 2017.

[39] Min Chen, YixueHao, Kai Hwang, Fellow, IEEE, Lu Wang, and Lin Wang, Disease Prediction by Machine Learning over Big Data from Healthcare Communities, IEEE 2017.

[40] MarioMunoz-Organero,VictorCorcoba-Magana,Pre-dictingUpcomingValues of Stress While Driving, IEEE 2016.

[41] ThomasStutz1,ThomasKowar,MichaelKager,MartinTiefengrabner,Markus Stuppner, Jens Blechert, Frank H. Wilhelm, Simon Ginzinger,Smartphone Based Stress Prediction, UMAP 2015, LNCS 9146, pp. 240251,2015.

[42] Yiping Li1, Jing Huang, Hao Wang, Ling Feng, Pre-dicting Teenagers Future Stress Level from Micro-blog, IEEE 2015.

[43] DavidLiu,MarkUlric,ListentoYourHeart:StressPredictionUsing Consumer Heart Rate Sensors, Stanford CS 229: Machine Learning, Autumn2013-2014.

[44] Peter Yuen, Mark Richardson, Guangyuan Liu, and Zhishun She Tong Chen, "DetectionofPsychologicalStressUsingaHyperspectral ImagingTechnique,"inIEEE Transactions on Affective Computing, VOL. 5, NO. 4 , 2014

[45] M.Mahmoud,T.Baltrusaitis,P.RobinsonA.Adams,"Decouplingfacialexpressionsand headmotionsincomplexemotions,"inAffectiveComputingandIntelligentInteraction (ACII) , 2015.

[46] M. Pediaditis , D. Manousos , E. Kazantzaki , F. Chiarugi , P.G. Simos , K.Marias , M. Tsiknakis G. Giannakakis, "Stress and anxiety detection using facial cues from videos," in Biomedical Signal Processing and Control , 2017, pp. 89- 101.

[47] NishaRaichur,NidhiLonakadiandPriyankaMural,"DetectionofStress Using Image Processing and Machine Learning Techniques", IJET, Vol. 9 No. 3, July 2017. [

[48] M.VijayAnand and MuthukrishnanRamprasath, "Image Classification using Convolutional Neural Networks", International Journal of Pure and Applied Mathematics, 2018.

[49] KritikaGarg and VarunGarg, "Face Recognition Using Haar CascadeClassifier", Volume 3, Issue 12, December 2016.

[50] PurnenduShekharPandey, "Machine Learning and IoT for Prediction and Detection of Stress", IEEE 2017.

[51] Ahmed Hesham, Youssef Elkhorazaty and FadiAloul, "Emotion Recognition Using Mobile Phones", IEEE 2016.

[52] SoniyaLakudzode and S. M. Rajbhoj, "Review on human stress monitoring system using wearable sensors", IRJET, April 2016.

[53] M. Tarun Kumar, R. Sandeep Kumar et.al, "Health Monitoring and Stress Detection System", IRJET, March 2019.

[54] D.VenkataramananandNamboodiriSandhyaParameswaran,"Extractionof Facial Features for Depression Detection among Students", International Journal of Pure and Applied Mathematics, 2018.

[55] Keiron O'Shea and Ryan Nash, "An Introduction to Convolutional Neural Networks", [cs.NE], 2015.

[56]   KaranChauhanandShrwanRam,"ImageClassificationwithDeepLearning   and   Comparison
between Different Convolutional Neural Network Structures using Tensor flow and Keras", IJAERD 2018.

Hybrid Frequent Itemset Mining for Time Efficiency and Decision Making in Medical Field using Weight Based Frequent Item set Modified AnyFI Mining Algorithm

[1]Mrs. Sheeba Sugantharani E.* & [2]Dr. Subramani A.

1Research Scholar, Assistant Professor

1Department of Computer Science, Mother Teresa University, Kodaikanal,

Tamil Nadu, India- 625002

sheebasugantharani@ldc.edu.in,

2Department of Computer Science, M.V. Muthiah Govt. Arts College for Women,

Dindigul, Tamil Nadu, India - 62

Abstract

Data mining is a great concept with unlimited potential to predict future trends and behavior. It states that extraction of meaningful information from large datasets using methods like statistical analysis, machine learning, clustering, and neural networks. Furthermore, the accuracy of the data mining methods are improved by hybrid multiple pre-existing methods. The hybrid algorithm proposed in this paper uses the model of AnyFI (Anytime Frequent Itemset), which is capable of handling transactions with varying inter arrival rate and WD-FIM algorithm to fine tune weighted frequent itemsets searching space and enhance the efficiency of time. When the proposed approach is tested on gene datasets, the results obtained show improved accuracy and minimum processing time in most cases.

Keywords: clustering, AnyFI, WD_FIM, genomes.

1. Introduction

FIM (Frequent Itemset Mining) along with Intelligent decision making, is a hot research topic in data mining. Using these techniques association rules can be generated in datasets (R. Ishita and A. Rathod 2016) (L. Yue 2015) which are greatly applicable in the fields of predictive marketing, analysis and medical diagnosis, individual recommendations. Nowadays there is a rapid development of data acquisition with higher data processing technologies even for complex data, for example uncertain data.

In most of the already obtained studies the items set in the uncertain databases have the same importance. But in actuality, in the case of uncertain databases values of different items and its importance vary a lot based on datasets. Mostly, occurrence frequencies and theprobability of the itemsets are referred for mining. But it is not sufficient to identify meaningful or useful patterns especially in uncertain databases. Also, the existing algorithms miss out on two important characteristics of data streams. First, very often streams do not have constant speed, which varies depending on application domains. The second is that they lack the ability to produce immediate mining results with compromised accuracy, if required. So, when a request for mining results comes from the user, an ideal algorithm should give an immediate approximate result, and improve its quality with an increase in time allowance. The step for mining of FIs on the existing algorithms is quite costly in-terms of execution time and hence cannot give an immediate approximation result, even with compromised accuracy.

In the proposed algorithm, as a solution to these problems various weights were allocated to items in order to notify the importance. Assigning weights can be done by the domain experts with the existing professional knowledge of business risks, profits, costs, etc. The high importance itemsets were identified.

The three properties such as – 1. Inter arrival transaction rate; 2. Generating best results based on time allowance and 3. Improving searing space and time efficiency these are the features of   an anytime mining algorithm for data sets. A few anytime algorithms are proposed for clustering (P. Kranen 2011), (M. Hassani 2011) and classification (Int.J.Mach Learn 2014), (P. Kranen 2012). There is one anytime FI mining algorithm (Shichao Zhang 2002) proposed for multi-user applications which possess the second characteristics. However, the method is static and is not meant for data streams.

In this paper, hybrid mining algorithm of WD-FIM and AnyFI has been implemented. Furthermore, BFI-forest stores buffers at its tree nodes, which aid AnyFI(Anytime Frequent Itemset) to handle variable stream speeds. It can also give out a mining result almost immediately with compromised accuracy and can improve the quality of results. The structure of the paper is as follows :- Section 2 related works. Section 3 BFI-forest and WD-FIM algorithm preliminaries. Section 4 Hybrid algorithm time efficiency and completeness are discussed and analyzed. Finally experimental results and conclusion is drawn . Future work is discussed in Section 5.

2. **Related Work**

2.1 **Survey on WD-FIM Algorithm**

Here is a list of few algorithms used for mining frequent itemsets, they are

Algorithms	Inferences	Working Principles
WAR (Weighted Association Rules) algorithm	Associate weight parameter	Segmenting or identifying degree or volume of data items (W. Wang 2000)
WARM (Weighted Association Rule Mining) algorithm	Weighted download closure property	Discover significant relationships in weighted itemsets (F. Tao 2003)
WFIM (Weighted Frequent Itemset Mining) algorithm	Considering both importance and existential probability	Narrow down the searching space and improve time efficiency (Xueajian Zhao 2018)
WSpan	Weight constraint in sequential pattern	Generating important weighted sequential patterns (U. Yun 2006)
WMFP-SW (Weighted Maximal Frequent Pattern mining over data streams based on Sliding Window model)	One scan over sliding window based data stream environment	Generate weighted maximal frequent patterns over data streams (G. Lee 2014)

MWS (Maximal frequent pattern mining with Weight conditions over data Streams)	Prevents overheads of pattern extractions	Results in enhanced reliability of mining by applying weight conditions (U. Yun 2014)
WEP (Weighted Erasable Patterns) mining algorithm	Employee tree and table structures	Erase unnecessary operations which deals with invalid patterns in datastreams (YoonjiBaek, Unil Yun 2020)
U-WFI (Uncertain Mining of Weighted Frequent Itemsets) algorithm	Mine indeterminate itemsets from an uncertain datasets	discover significant itemsets, existential probabilities (Jerry Chun-Wei Lin1 2015)
HEWI-U apriori (High Expected Weighted Itemset) algorithm	Prune search space	mine high predictable weighted itemsets (Jerry Chun-Wei Lin1 2015)

Table 1. Frequent itemsets mining algorithms

Subsequently, more research should be shown to improve the efficiency of mining frequent itemset in uncertain databases.

2.2 **Survey on AnyFI Algorithm**

Researchers have proposed various FI mining algorithms for data streams using stream models such as landmark window (LW) and sliding window (SW). Sticky-Sampling and Lossy Counting (G. S. Manku 2002) are two algorithms that are based on Apriori and follow the LW model. They produce less accurate results with an error bound. FP-Stream (C. Giannella 2003) uses tilted-time window (variation of LW), which takes data in batches of transactions, builds FP-tree for each batch, mines for FIs from it, and inserts them into a pattern tree, which has tilted-time windows stored at its nodes. Since, this algorithm mines every batch separately, it cannot handle variable stream speeds. DSM-FI (H.-F. Li, 2008) is another algorithm that uses LW. It keeps a forest of prefix trees which are similar to FP- trees and inserts each incoming transaction completely into the forest and thus cannot be interrupted anytime.

CPS-Tree (S. K. Tanbeer 2009) is an FP-tree based algorithm that uses SW. It takes in a pane of transactions at a time and inserts them into the CPS-tree. After insertion of a certain fixed number of panes, CPS-tree undergoes restructuring and pruning. It uses FP-growth for mining FIs, whenever the request comes. FP- growth is not capable of delivering an immediate result with compromised accuracy. This is because it enumerates many conditional trees which takes a long time for larger batch sizes. MFI-TransSW (P. Goyal 2017) is another SW based approach that represents the items in the form of a bit vector and uses apriori for mining, which is very slow for lower support thresholds and is also not capable of delivering an anytime result. SWP- Tree (H. Chen 2012) is another SW based algorithm that uses a prefix tree similar to that of FP-tree. SWP-tree additionally uses decay on its support count to give higher weightage to recent transactions. It uses FP-growth for mining FIs, which is again not capable of giving anytime results. VSW (M. Deypir 2012) is another algorithm which adjusts the size of SW on demand. It uses ECLAT (M. J. Zaki 2000) which is a variant of Apriori and is slow in its computations,

and thus not capable of handling high speed streams.

Based on the above discussion, it is clear that none of the existing algorithms fit the bill for an anytime FI mining algorithm for data streams with maximum accuracy with minimum time efficiency.

3. Proposed Methodology

In this section, we propose a novel mining scheme which is a hybrid algorithm of both WD-FIM and AnyFI. Both mining algorithms have their own individual advantages and disadvantages. In which, the algorithm mines the itemsets using the candidate property in an uncertain database which will result in a test paradigm. In the same way, the input dataset has been mined by AnyFI. It uses sampling and addresses the second aspect of an anytime algorithm, i.e., it gives an immediate approximate result and improves it with increase in time allowance.
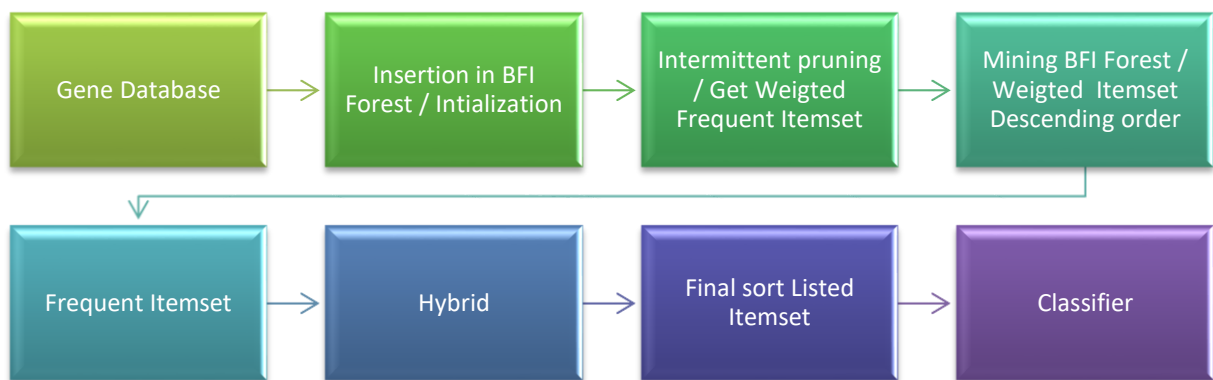
Fig. 1. Flow diagram of hybrid mining technique for gene dataset

Finally, the frequent itemset of two methods are combined for classification. Fig. 1 shows the block diagram representation for the hybrid proposed algorithm. Initially, the dataset defined by its attributes or class labels is taken as input.

3.1 Preliminaries and WD-FIM Algorithm

In this section, the weight judgment downward closure property and the existence property of weighted frequent subsets are first proposed and proved. Moreover, the WD-FIM algorithm is proposed (F. Tao 2003) and described in detail.

```
Algorithm 1: WD-FIM algorithm
Input:
DS, an uncertain transactional dataset;
wtable, a weight table;
ε, a user-specified minimum expected weighted support threshold.
Output:
The set of weighted frequent itemsets WFIS.
/* initialization */
1.  initialize the variables and parameters
/* scan the dataset and get weighted frequent 1-itemset */
2.  for each item I_j in DS do
3.     scan DS and calculate expwSup(I_j)
4.     if expwSup(I_j) ≥ |DS| × ε then
5.        WFIS_1 = WFIS_1 ∪ {I_j}
6.     end if
7.  end for
8.  WFIS = WFIS ∪ WFIS_1
```

```
/* scan the dataset and get weighted frequent k-itemsets */
9.   CWFIS₁ = I
10. let SCWFIS₁ be sorted CWFIS₁ by weight in descending order
11. set k =2
12. while WFISₖ₋₁ ≠ null do
13.     CWFISₖ = Connection(WFISₖ₋₁, CWFIS₁)
14.     NCWFISₖ = wConnection((CWFISₖ₋₁ − WFISₖ₋₁), SCWFIS₁)
15.     RCWFISₖ = CWFISₖ − NCWFISₖ
16.     for each candidate k itemset X in RCWFISₖ do
17.         scan DS and calculate expwSup(X)
18.         if expwSup(X) ≥ |DS| × ε then
19.             WFISₖ = WFISₖ ∪ {X}
20.         end if
21.     end for
22.     WFIS = WFIS ∪ WFISₖ
23. end while
24. return WFIS
```

Fig.2 Pseudo code of WD-FIM algorithm

3.2 Preliminaries

Definition: (Expected weighted support of an itemset in DS)

The expected weight support of itemsetX in dataset DS is denoted as expwSup(X). It is the product of the expected support of itemset X and the weight of itemsetX.It can be formally defined as: (W. Wang 2000)

$$\exp \exp wSup(X) = w(X) * expSup(X)$$

$$= \frac{\sum_{Ij \in X} w(Ij)}{|K|} * \sum_{X \subseteq Tq \wedge Tq \in DS} \left(\prod_{Ij \in X} p(Ij, Tq) \right)$$

Based on the above definition, the problem of mining weighted frequent itemsets in uncertain databases can be formulated as following: (W. Wang 2000)

The uncertain database to be analyzed is, the user-specified weights of the items in DS are defined in wtable, and the user-specified minimum expected weighted support threshold is ∈ . The problem of mining weighted frequent itemsets in the uncertain database DS is to discover the weighted frequent itemsets considering both the weight and the existential probability constraints. An itemsetX is a weighted frequent itemset if the expected weighted support of an itemset X is greater than or equal to the   minimum expected weighted support, (W. Wang 2000) i.e. expwSup(X) ≥ ∈ × |DS|.

As mentioned above, the downward closure property cannot be applied directly to narrow the searching space of mining weighted frequent itemsets. Consequently, how to improve the time efficiency of mining weighted frequent itemsets is a major issue for urgent solution. (W. Wang 2000)

3.3 WD−FIM Algorithm

First, the WD-FIM algorithm is proposed for mining weighted frequent itemsets in uncertain datasets. Though, U-Apriori can only be used to learn frequent itemsets in uncertain datasets. Second, the origin of the proposed WD-FIM algorithm is the above-mentioned weight judgment downward closure property and existence property of weighted frequent subsets, but the downward closure property is used directly to narrow the searching space of frequent itemsets in the U-Apriori algorithm. (U. Yun 2006)

As shown in Fig.2, the WD-FIM algorithm takes the input as: an uncertain transactional dataset DS; a weight table, wtable ; a user-specified minimum expected weighted support threshold ∈. First of all, the variables like WFIS and WFISₖ are initialized (Line 1). Then it scans the dataset to get the weighted frequent 1-itemsets (Lines

2 to 8). (U. Yun 2006)

During this process, the calculation of $CWFIS_k$ is extremely critical (Lines 13 to 15). Line 13 is performed to guarantee all the weighted frequent k itemsets are included in $CWFIS_k$ according to the existence property of weighted frequent subsets. (U. Yun 2006)

The Connection operation is similar to the Apriori_gen operation in HEWI- U apriori algorithm. Line 14 is implemented to obtain the itemsets which are definitely not weighted frequent k itemsets according to weight judgment downward closure property.

The wConnection operation means that the itemsets in $CWFIS_{k-1} - WFIS_{k-1}$ are connected with the itemsets with smaller weights in $SCWFIS_1$. Line 15 is performed to narrow the searching space of weighted frequent itemsets by deleting itemsets which are definitely not weighted frequent k itemsets from CWFIS.

3.4 AnyFI Algorithm

AnyFI is the proposed anytime algorithm for mining frequent itemsets in a continuous data stream. The algorithm in summary consists of the following steps:

Step 1: Read incoming transactions one by one and order them lexicographically.

Step 2: Insert each transaction into the BFI-forest.

Step 3: Intermittently prune the infrequent itemsets from BFI- forest after a fixed interval of arrival of transactions.

Step 4: Mining of BFI-forest for frequent itemsets on demand depending upon the time allowance.

Insertion in BFI-forest. The insertion of an incoming transaction into the BFI-forest is an anytime algorithm, i.e. it is interruptible. The insertion process of a transaction goes on until it's time allowance expires (a new transaction arrives), after which we proceed to insert the newly arrived transaction. The incoming transactions from the stream are inserted one by one into the BFI-forest. For every transaction, we first compute the suffix projections and insert them into the buffers of the corresponding trees. For example, in Fig.3, for the incoming transaction <acde>, the projections <cde>, <de> and <e> will be inserted into the buffers of the roots of Trees 1, 3 & 4 respectively, and efreq of the roots of Trees 1, 3, 4 &5 will be incremented by 1, after getting decayed using equation 1 with T1 & T2 as l time and current time respectively. For inserting <cde> into the buffer of root of Tree1, we first find the hash value to identify the bucket into which it has to be inserted, (bucket # = c mod |hash_size|). All the projections starting with c will be hashed to this bucket. If a buffer node indexing <cde> already exists in this bucket, we increment its frequency count by 1, after decaying it, as was done before. Otherwise, we create a new buffer node and append it at the end of the bucket. All other projections are inserted in the same way. This operation - taking suffix projections of a transaction and inserting them into a buffer, is an atomic operation (non-interruptible) in our algorithm.

 After this step, we take each of the affected trees (trees that were accessed in the above step – Trees 1, 3, 4 & 5 in Fig. 3) and start refining them in depth first order (DFS) of tree nodes. Depending on available time allowance, we refine these trees one after the other, i.e., if a new transaction arrives before refining all the trees, we quit in between and proceed with processing of the newly arrived transaction.
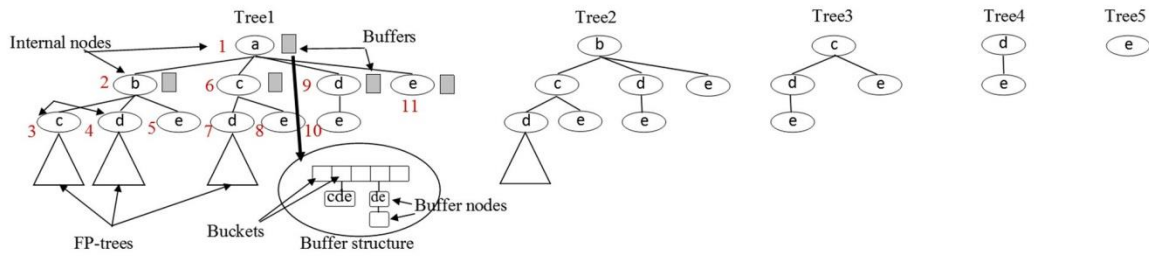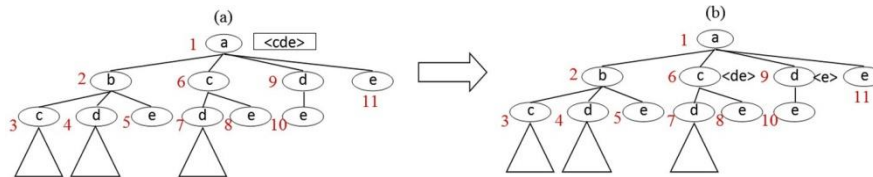
Fig. 3. Structure of BFI-Forest and BFI-Trees

Fig. 4. Insertion in BFI-Tree

Consider Tree1 in Fig. 4a, where <cde> is inserted in its root's buffer in the above step. The refinement starts with root node of this tree (root becomes curr_node). We first remove the first projection (or partially inserted transaction) from a randomly chosen bucket of its buffer. This projection is the oldest and expected to have the highest frequency count among buffer nodes in the bucket. Let's say that <cde> is the partial transaction that was removed. We first update its frequency count as explained before using eqn. 1. If curr_node is an external node, we would insert this projection into the FP-tree beneath it and exit the insert function. Otherwise, we take suffix projections of <cde> and insert them into the buffers of the corresponding children of curr_node as shown in Fig 4b (nodes 6 & 9 get projections into their buffers). The insertion into these buffers is the same as explained before. We also increment the frequency counts of nodes 6 & 9 with respect to the frequency counts of Fig. 3. Structure of BFI-Forest and BFI-Trees Fig. 4. Insertion in B-Tree The partial transactions being inserted into their buffers, after decaying them. If any child into which a projection has to be inserted does not exist, we create that child first, insert it into its buffer and assign it a frequency count. Also, while inserting into the buffer, we take care that the buffers do not overflow. If they exceed a predefined capacity- buffCapacity, the oldest (first) item is removed from the bucket in which we are trying to append the current projection. The oldest is chosen because it has the maximum probability of becoming infrequent as we are decaying the frequency counts with time. After this step, we check if a new transaction has arrived. If yes, we return back to the caller function and move on to process the next transaction; otherwise we continue processing the current transaction. Please note that we check this condition at the beginning of refinement of every node.

After insertion of projections into the sub-trees, buffer pruning is conducted. Buffer pruning prunes away infrequent projections lying in the buffers. So, the buffers of affected children of curr_node (children into which suffix projections were inserted in the previous step – nodes 6 & 9 in Fig. 4b) are pruned before we proceed with further refinement. Buffer pruning is not done every time we visit a given node in our traversal. This is because, there may not be many infrequent projections accumulated every time we visit a node's buffer. So, by delaying it, we let the node accumulate a few more infrequent projections and then remove them in one go. So, buffer pruning is conducted in intervals of some minimum time decided by a parameter y and the height of the node. It can be observed that closer the node to the root, the more filled its buffer will be. So, buffers at lesser depth must

be pruned more often than the buffers at greater depth. So, the pruning interval (PI) for each node is computed using the following formula -PI =]batch_cize $\times y \times$ height(node)]; batch_size is the #of transactions after which we perform intermittent pruning (explained next). Whenever we are visiting a node, we prune its buffer only when - it was last pruned at least PI transactions earlier. To prune a given buffer, we visit every buffer-node (nodes from all the buckets) and check if partial_trans in it is c -frequent or not (after decaying its frequency count). If it is not, then we check if the current affected_child (node for which buffer pruning is being conducted) has a child in its childArr which corresponds to the head of this partial_trans. If it does, then we don't delete this partial transaction, as we might lose a potential FI by removing it. Otherwise, it is safely deleted.

After pruning the buffer, we now select nodes which are to be refined further. We prefer to deliberately delay the refinement of certain nodes to save space and time. We let the nodes accumulate more itemsets in their buffers before they are refined or expanded for insertion into sub trees. This step is critical in making the insertion fast. By doing this, we save time by avoiding repeated insertions and removals of infrequent itemsets. We let many infrequent itemsets be pruned away from buffers itself, rather than getting expanded into a large number of infrequent subtrees. This delayed refinement is achieved by a tuning parameter 8 and this process is known as 8-deferring. So, after inserting the suffix projections at a node and pruning its buffer, every affected child is checked whether it is z-frequent or not. If yes, then we check whether the subtree corresponding to the head of the projection to be inserted into it, is present in its childArr or not. If this is so, we keep this node in a stack for its refinement in subsequent iterations. For example, consider node 6 in Fig 4b. If it is c-frequent and there exists a subtree with root d present in its childArr (node 7), we would want this node to be refined further and thus, push it into the stack. However, if the subtree doesn't exist, then we check if the node is 8- frequent or not. If it is so, only then we would want this node to be refined further and we push it into the stack. Else, we don't refine this node and let it accumulate more transactions in its buffer. This reduces creation of infrequent subtrees in the forest. After this we proceed for the next iteration, in which the nodes accumulated in the stack are refined.

Intermittent Pruning. After a fixed batch (batch_size) of transactions are received, we conduct intermittent pruning. This is for timely pruning of infrequent subtrees in the forest, which makes insertion fast & efficient. In this step, we visit each node of every BFI-Tree in DFS order, and delete it along with all its child subtrees, if it is infrequent. Removing the subtrees beneath infrequent nodes doesn't affect the accuracy of the algorithm because all itemsets in those subtrees can never be frequent. If however, the node is c-frequent, we do not delete it. If the node we visit in the traversal is a leaf, we prune the FP-tree beneath it, as explained in (M. Hassani 2011).

In principle, deletion of a sub-tree in a BFI-tree does not affect the other branches of the tree (and other trees as well). Unlike other algorithms, e.g. DSM-FI (P. Kranen 2011), we do not need to update or remove items from other branches. This is because AnyFI enumerates all possible suffix projections of the incoming transactions, which get stored in independent branches. The frequency counts of itemsets represented by the other branches have no connection with the branch being deleted. Hence, our pruning step is accurate.

Mining BFI-forest. Mining for FIs in BFI-forest is very simple and straightforward. In the insertion step, we have inserted suffix projections of all incoming transactions in the trees, and within each tree we have enumerated suffix projections of those partial transactions and inserted them either into the tree as subtrees or into the buffers. This makes mining tasks easy, without the need of enumerating candidate itemsets like in apriori or generating conditional trees like in FP-growth. This makes AnyFI mine faster than the existing algorithms. All that is needed is to empty all the buffers in the trees (traversing in DFS or BFS order) and then traverse the tree again in DFS

order accumulating the itemsets whose frequency counts are greater than o. So, as we traverse down, at every node we check if it is o-frequent. If yes, we store the itemset represented by that node in FIset, which is a list of FIs. Note that each node in the BFI-tree represents an itemset with items in the path from root to this node. Whenever an external node is encountered, we simply mine the FP-tree beneath and concatenate all the FIs that come from FP-tree with the itemset represented by the current node, and add all of them to FISet. Please note that when we mine a BFI-tree with root node having an item say a, all the FIs that start with are mined from this tree only. This is because, while inserting a given transaction, say<abcd>, we insert all its suffix projections into their respective trees. So, <bcd> is inserted into a tree with root as b and <cd> is inserted into a tree with root c. The contribution of transaction <abcd> to the frequency count of itemset {bcd} is taken care of by this process. So, to mine for {bcd}, it is not required to traverse the tree with root a.

We have seen that mining for FIs has two steps – 1) emptying of buffers and 2) traversing to accumulate FIs. We had empirically observed that emptying buffers is the step that takes maximum time. Traversal for FIs takes time in order of milliseconds. So, we have made emptying of buffers step anytime, i.e. it proceeds as long as the time is available and once the quantum of time allotted by the user expires, the algorithm exits from this step and quickly traverses the forest for accumulating FIs. In this traversal, the residual partial transactions in the buffer, if any, are ignored. Consequently, we get a less accurate result when time allowance is less.

4   Results and Discussion

MATLAB applies the proposed algorithm and other comparative algorithms. Experiments was performed on a computer running the 64-bit operating system Microsoft Windows 7, using the Intel Core i7-4510U 2.6 GHz processor and 8 GB RAM. Three classes of experiments have been performed with the goal of evaluating the efficiency and effectiveness of the proposed algorithm, in terms of runtime, amount of patterns and memory usage.

4.1 Performance of runtime

The runtime of the algorithm proposed and the associated algorithms are evaluated in this chapter first. The sizes of the datasets are calculated in this category of experiments. Nonetheless, to evaluate subsequent running time shifts the minimum estimated weighted assistance level is modified. Remember that the Uapriori algorithm can be seen as a FIM algorithm in which all things are weighted to 1. In fact, the runtime requires both the time to search the sample and the duration.

Last but not least, we compare FP-growth, DSM-FI and proposed Gene data set approaches. We insert the entire data collection without pruning into the description structures for all three algorithms and then into the FIs for fair comparison. Mining is done in full mode in the proposed method, i.e. without interruptible features. In table 1, the results indicate that the procedure suggested is conducted more rapidly. It should be noted that SWP tree and FP-Stream use FP production.

Algorithm	Mining Time
FP-Growth (P. Goyal 2017)	5.8Sec
DSM -FI (P. Goyal 2017)	3.3Sec
Proposed Method	1.1Sec

Table 2 Performance Comparison in Time

It is clear from Fig. 5, all of which slowly drown with an improvement in the minimum weighted support threshold at run-times for Uapriori algorithm, WD-FIM algorithm and proposed algorithms. That's because with the rise of the minimum weighted approval level the number of candidates who weighed the regular things to be checked declined. Therefore, no more time needs to be spent searching the dataset by all three algorithms.

From Fig. 5 It can also be found that, if the minimum predicted weighted support level is relatively high, the proposed algorithm is faster than the Uapriori HEWI-Uapriori and WD-FIM algorithms. This is because the current hybrid approach with the weight assessment of the downward closing property will significantly reduce the number of applicants. When the weighted minimum support requirement is however that, further candidates in CWFIS$_k$ will be available. Therefore, it is very time consuming to run wConnection and to delete itemsets which are undoubtedly often un-weighted k-itemsets from CWFIS k. The runtime of the algorithm proposed is therefore steadily increasing. Fig.5 shows that if the estimated minimum weighted support level ranges from 0.004 to 0.01 the runtime for the proposed algorithm is much shorter. Nonetheless, the algorithm suggested is much less than Uapriori, WD FIM and HEWI-Uapriori when it is set to at the minimum anticipated weighted assist point of 0.002.

In the Gene dataset similar genes will be in a similar pattern. Most repeated patterns are relevant to each other. According to the highest repetition, genes are clustered and classified.

Number of combinations determines the weightage of an item set. According to the repetition or frequency of occurrence of the same pattern the Weightage is determined. Weightage is the additional novelty given to the already existing BFI Forest data structures. Highest weightage FP is generated.In BFI no repeated pattern is generated. For example if ACT is already exists then when it comes again for insertion with ACT combination , then it will be added at the already existing node, no new rule will be generated at this point , thus avoiding redundancy
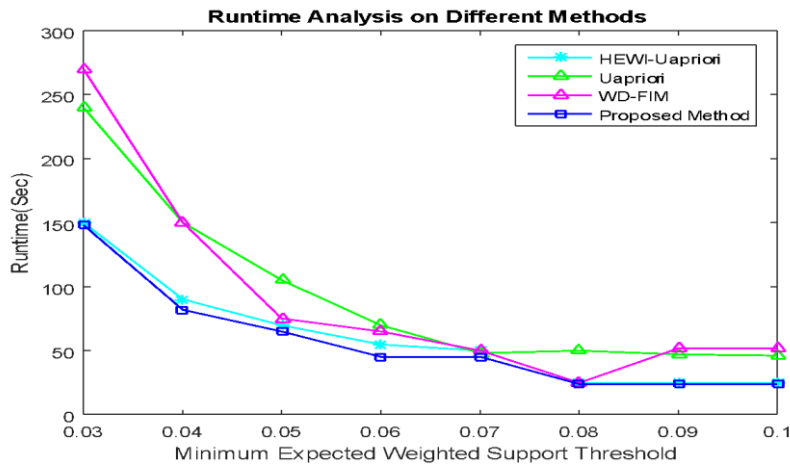
Fig. 5. Run time Analysis on different Methods
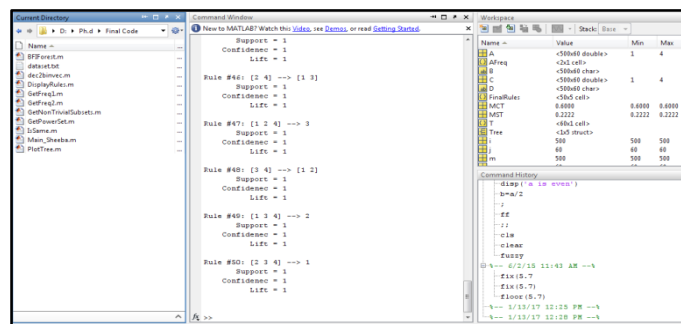
Fig. 6. Association Rule Generation

As shown in the Fig 6 only for new combinations association rules are generated.Final rules generated in the proposed algorithm will be non-repeated highest weightage combination of association rule.

4.2 Pattern Analysis

In this section, the maximum and minimum predicted weighted support threshold values are evaluated for the numbers of patterns observed with Uapriori, the HEWI-Uapriori algorithm, the WD-FIM algorithm and the proposed algorithm. The patterns found in Uapriori algorithms are assumed to be provided by frequent articles (EFIs), strongly predicted weighted items set (HEWIs) by the patterns discovered by HEWI-Uapriori algorithms and weighted frequent articles (WFIs) by the patterns discovered by the proposed algorithm. Therefore, when a weighted minimum assistance level is established, an itemset is more likely to be an EFI In BFI no repeated pattern is generated. For example if ACT already exists then when it comes again for insertion with ACT combination, then it will be added at the already existing node, no new rule will be generated at this point thus avoiding redundancy. Final rules generated in the proposed algorithm will be non-repeated highest weightage combinations of association rules.

From Fig. 7, with the increases of minimum predicted weighted support level, the number of patterns discovered in all four algorithms is decreasing steadily. At the onset, the downward trend is important, but slowly falls late. It is clear that the Uapriori algorithm, relative to the HEWI-Uapriori, the WD FIM and the proposed algorithm, can always uncover more trends when the minimum required weighted threshold is calculated. The explanation is that all things can be seen as 1 as above in the Uapriori algorithm.
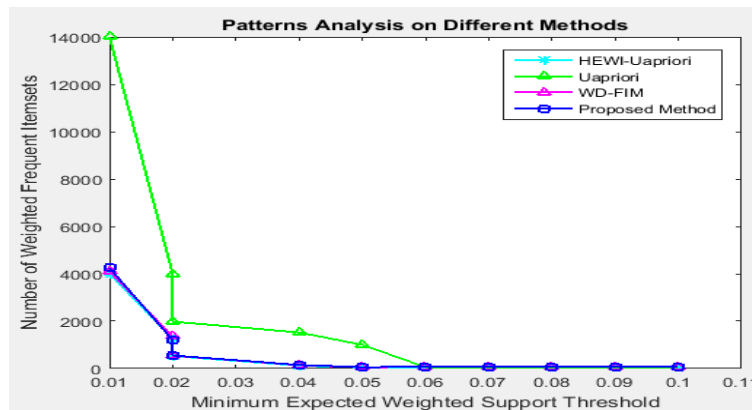
Fig. 7. Pattern Analysis on Different Methods

In comparison, the various models found with the HEWI-Uapriori-WD-FIM algorithm and the algorithm proposed are similar. It is significant. The HEWI-Uapriori-WD-FIM algorithm can detect weighted or frequent items, but the proposed algorithm found patterns of both weighted frequent articles (WFIs) which are commonly accessible objects in the data set.

4.3 Performance of Memory Consumption

Further experiments have been done in this segment to examine Uapriori algorithm memory usage, HEWI-Uapriori algorithm and the WD-FIM algorithm proposed. From Fig. 8, the Uapriori algorithm always needs less memory than the HEWI -Uapriori, WDFIM algorithm and the algorithm proposed. This is for a number of reasons. The upward-closure property to explicitly pick up some potential EFI candidates. This ensures that less candidates are produced. The suggested method of downside closure to prune the candidates is used by the HEWI-Uapriori, WD-FIM algorithm. The candidates are still kept in the main memory after the first dataset

search. The weight evaluation downward closure proprietors of the proposed algorithm are used to narrow the search area for regularly weighted products. The NCWFIS k is determined and also stored in the main memory during this process

4.4 Performance Analysis

Mainly two factors determine the performance analysis of an algorithm, amount of runtime consumed and the amount of memory used during the process. When comparing HEWI-Uapriori-WD-FIM algorithms, the proposed algorithm generates non-repeated highest weightage combinations of association rules. This is accomplished by finding all frequent items with normalized greater % than the thresholds' and the result set will be the Non-Trivial Subsets.That ensures less candidate production which will narrow down the search area thus resulting in minimum runtime and less memory consumption.
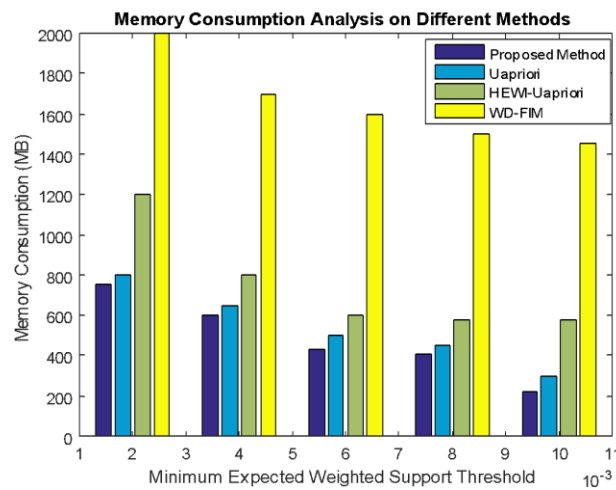
Fig. 8. Memory consumption analysis on Different Methods

5   Conclusion

This paper proposes an AnyFi together with weight judgment based on frequent item collection mining algorithm, to narrow the search field for weighted frequent items, and to improve time efficiency to incorporate smart medical system decision making.

The first ever FI mining algorithm for data streaming is AnyFI in this article. AnyFI has the capability to manage varying stream speeds I to provide immediate mining results and to boost its precision with that timescale. In addition, the weight judgment downward closing property is added and proven first in weighted frequent subsets and the presence of weighted frequent subsets. The provided experimental analysis shows that the system proposed can handle higher concentrations with less volume.

References

[1] PatrykOrzechowski, Krzysztof Boryczko "Hybrid Bi-clustering Algorithms for Data Mining" Springer International Publishing, Switzerland 2016.

[2] R. Ishita and A. Rathod, "Frequent Itemset Mining in Data Mining: A Survey," International Journal of Computer Applications, vol. 139, no. 9, pp.15-18, April 2016.

[3] L. Yue, "Review of Algorithms for Mining Frequent Patterns," International Journal of Computer Science and Network Security, vol. 15, no.6, pp.17-21, June 2015.

[4] P. Kranen, I. Assent, C. Baldauf, and T. Seidl, "The ClusTree: Indexing micro-clusters for anytime stream mining," Knowledge and Information Systems, vol. 29, no. 2, pp. 249–272, 2011.

[5] M. Hassani, P. Kranen, and T. Seidl, "Precise anytime clustering of noisy sensor data with logarithmic complexity," in Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data - SensorKDD 'll, 2011, pp. 52–60.

[6]Int.J.Mach Learn, " International   Journal of machine learning and cybernetics ",Springer, 2014.

[7] P. Kranen, M. Hassani, and T. Seidl, "BT* - An advanced algorithm for anytime classification," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012, vol. 7338 LNCS, pp. 298–315.

[8] Shichao Zhang and Chengqi Zhang, "Anytime mining for multi user applications," IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 32, no. 4, pp. 515–521, Jul. 2002.

[9]   W.   Wang,   J.   Yang,   and   P.   S.   Yu,   "Efficient   mining   of   weighted association rules (war)," in Proceedings of the sixth ACM SIGKDDinternational conference on Knowledge discovery and data mining, 2000, pp.270-274.

[10]   F. Tao, F. Murtagh, and M. Farid, "Weighted association rule mining using weighted support and significance framework," in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp 661–666.

[11] U. Yun and J. Leggett, "WFIM: Weighted frequent itemset mining with a weight range and a minimum weight," in Siam International Conference on Data Mining, 2005, pp 636–640.

[12] U. Yun and J. J. Leggett, "WSpan: Weighted Sequential pattern mining in large sequence databases," 2006 3rd International IEEE Conference Intelligent Systems, 2006, pp. 512-517, doi: 10.1109/IS.2006.348472.

[13] G. Lee, U. Yun, and K. H. Ryu, "Sliding window based weighted maximal frequent pattern mining over data streams," Expert Systems with Applications, vol.41, no. 2, pp. 694-708, Feb. 2014.

[14] U. Yun, G. Lee, and K. H. Ryu, "Mining maximal frequent patterns by considering weight conditions over data streams," Knowledge-Based Systems, vol. 55, no. 55, pp. 49-65, Jan. 2014.

[15] G. Lee, U. Yun, and H. Ryang, "Mining maximal frequent patterns by considering weight conditions over data streams," Journal of Intelligent & Fuzzy Systems, vol. 28, no. 3, pp. 1145-1157, May 2015.

[16] G. Lee, U. Yun, and H. Ryang, "An uncertainty-based approach:Frequentitemset mining from uncertain data with   different   item   importance," *Knowledge-Based Systems*, vol. 90, pp. 239-256, Dec. 2015.

[17]   C. W. Lin, W. Gan, P. Fournier-Viger, T. P. Hong, and V. S. Tseng, "Weighted frequent itemset mining over uncertain databases,"*AppliedIntelligence*, vol. 44, no. 1, pp. 232-250, Jan. 2016.

[18] G. S. Manku and R. R. Motwani, "Approximate Frequency Counts over Data Streams," in *VLDB '02 - Proceedings of the 28th VLDB Conference*, 2002, pp. 346–357.

[19] C. Giannella, et al., "Mining Frequent Patterns in Data Streams at Multiple Time Granularities," in *Data Mining: Next Gen. Challenges and Future Directions*, AAAI/MIT Press, 2003, pp. 191–212.

[20] S. K. Tanbeer, et al. , "Sliding window-based frequent pattern mining over data streams," *Inf. Sciences*, vol.179, no.22, pp. 3843–3865, 2009.

[21] H.-F. Li, M.-K.Shan, and S.-Y. Lee, "DSM-FI: an efficient algorithm for mining frequent itemsets in data streams," *Knowledge and Information Systems*, vol. 17, pp. 79–97, 2008.

[22] H. Chen, L. Shu, J. Xia, and Q. Deng, "Mining frequent patterns in a varying-size sliding window of online transactional data streams," *Information Sciences*, vol. 215, pp. 15–36, 2012.

[23] M. Deypir, M. H. Sadreddini, and S. Hashemi, "Towards a variable size sliding window model for frequent itemset mining over data streams," *Computers & Industrial Engineering*, vol. 63, no. 1, pp. 161– 172, 2012.

[24] M. J. Zaki, "Scalable algorithms for association mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 3, pp. 372–390, 2000.

[25] H.-F. Li and S.-Y. Lee, "Mining frequent itemsets over data streams using efficient window sliding techniques," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1466–1477, 2009.

[26]P. Goyal, J. S. Challa, S. Shrivastava and N. Goyal, "AnyFI: An anytime frequent itemset mining algorithm for data streams," *2017 IEEE International Conference on Big Data (Big Data), Boston*, MA, 2017, pp. 942-947.

[27] Xuejian Zhao, Xinhui Zhang, Pan Wang, Pan Wang, Song Chen and Zhixin sun, "A Weighted Frequent Itemset Mining Algorithm for Intelligent Decision in Smart Systems" Published in : IEEE Access, Vol 6, pp. 29271 - 29282, May 2018.

[28] YoonjiBaek, Unil Yun, "Erasable pattern mining based on tree structures with damped window over data streams", Engineering Applications of Artificial Intelligence, Volume 94, 2020, 103735, ISSN 0952-1976.

Recent Exploration of Artificial Intelligence in Stroke Disease Management

Dr.A.Shanthasheela[a],[c,d] ,Mrs.D.Jeyanthi[b,c,e], Mrs.A.P.Nagalakshmi[b,c,f]

[a]Assistant Professor, [b]Guest Lecturer

[c]Department of Computer Science,

M.V.Muthiah Govt. Arts College for Women,Dindigul - 624302, Tamil Nadu,

[d]drashanthasheela@mvmwgacdgl@ac.in,

[e]djeyanthi@mvmwgacdgl@gmail.com,[f]apnagalakshmi@mvmwgacdgl@ac.in

Abstract

Artificial intelligence (AI), making the system learn and rectify the problem automatically, which dramatically steps into all the fields including medicine. AI-based disease prognosis, detection, and further treatment assist the doctors and improve accuracy. Stroke is a hazardous disease that is difficult for the patient if it will not be detected on time. So AI plays an important role in stroke disease management. Deep learning and machine learning are the subdivisions of AI. This review article offers a peek at the recent research works that underwent in AI-based stroke disease management.

Keywords: Artificial Intelligence, Stroke, Deep Learning, Machine Learning

Introduction

Stroke is also known as brain attack is one of the hazardous diseases which leads to the major cause of mortality or long-term disability with no effective treatment. [1] over 80 million people have had a stroke around the world.A stroke occurs when blood flow to a part of the brain is interrupted or reduced, depriving brain tissue of oxygen and nutrients. This blockage or rupture is caused in the common carotid artery (CCA) or the internal carotid artery (ICA), which provides blood to the brain. [14]The formation of plaque along the artery walls causes obstruction or rupture. Plaque is produced between the lumen-intima (LI) and media-adventitia (MA) interfaces and is often constituted of cholesterol, fatty substances, cellular waste products, calcium, and fibrin. Within minutes, brain cells begin to die. One-minute delayed treatment cost is tremendous.1.9 million neurons, 14 billion synapses, and 12 kilometers (7.5 miles) of myelinated fibers are lost per minute[12]. The doctors insisted the public remember these symptoms by the term F.A.S.T which means F: face drooping, A: Arm weakness, S: Speech Difficulty, T: Time to Call. [2] Despite ongoing new evidence on how to best treat stroke patients, the risk of stroke recurrence remains unacceptably high, demonstrating the need for newer therapeutics. [3] found a more than 3-fold increase in the burden of stroke in developing nations, with 4.85 million fatalities and 91.4 million Disability Adjusted Life Years (DALYs) compared to 1.6 million deaths and 21.5 million DALYs in high-income countries. Considering the bad effects of stroke disease, it is significant to predict stroke disease. [25,26,27]Artificial Intelligence plays a vital role in cerebrovascular disease management. [29,38]Artificial intelligence (AI) could help to accelerate this collaboration and usher in a new era of automated stroke triage. This paper reviewed the methods and techniques available to predict stroke disease using AI Techniques. The overall work of this paper is divided into the following sections:

Section 1: It provides an overview of stroke disease with its types and symptoms.

Section 2: This section lists the Challenges in stroke disease Prediction, Detection, and Treatment.

Section 3: This section gives details about the sources of research papers and the inclusion criteria followed for this review. It categorizes the latest research works into three types such as stroke disease prediction, detection, and treatment using data mining, deep learning, and machine learning.

Section 4: Finally, the paper concludes with the research gap and opportunities for further research.

Stroke definition, types, symptoms, and Treatment

Definition: When the blood flow to a specific portion of the brain is diminished or disrupted, a stroke develops.Brain cells cannot function properly without oxygen and nutrients from the blood and eventually die.

Types: There are two major types of strokes named Ischemic Stroke caused by a blocked artery and Hemorrhagic stroke caused by a ruptured artery. Ischemic stroke happens when a blood clot obstructs an artery. The clot forms locally inside the brain's blood vessels in some patients. When fatty deposits in an artery, also known as cholesterol plaques, break, blood clots. In some circumstances, a clot may form elsewhere in the body and move to the brain. This is most common in those who have atrial fibrillation, a cardiac ailment in which the heart does not pump effectively, blood stagnates in the chambers, and blood clotting is easier.The clots may then pass into the bloodstream, get stuck in smaller brain arteries,and block them. On the other hand, Hemorrhagic stroke occurs when an artery leaks or ruptures. This can result from high blood pressure, overuse of the blood-thinners/anticoagulant drugs, or abnormal formations of blood vessels such as aneurysms and AVMs(Arteriovenous Malformation). As a hemorrhage takes place, brain tissues located beyond the site of bleeding are deprived of blood supply. Bleeding also induces contraction of blood vessels, narrowing them and thus further limiting blood flow.

Symptoms: Stroke symptoms may include paralysis of muscles of the face, arms, or legs such as inability to smile, raise an arm or difficulty walking, slurred speech or inability to understand simple speech, sudden and severe headache, vomiting, dizziness, or reduced consciousness.

Treatment: Cerebral stroke is a medical emergency and requires immediate attention. It is essential to determine if a stroke is ischemic or hemorrhagic before attempting treatment. This is because certain drugs used for the treatment of ischemic strokes, such as blood thinners, may critically aggravate bleeding in hemorrhagic strokes. Emergency treatment aims to restore blood flow by removing blood clots for ischemic strokes. The first line of defence is usually medication, such as aspirin and tissue plasminogen activator, or TPA. TPA can be administered intravenously or delivered directly to the brain via a catheter implanted through a groyne artery if the symptoms have progressed. Blood clots can also be mechanically removed using a catheter and a specific instrument.This type of treatment is called Endovascular Treatment. The insertion of catheters and guidewires into brain arteries carries a small but significant risk of bleeding or the spread of blood clots to previously untouched areas of the brain. If the patient is admitted to the hospital after the stroke was delayed, endovascular therapy is not recommended since it increases the risk of complications.On the other hand, emergency therapy for hemorrhagic strokes tries to stop bleeding, lower blood pressure, and prevent vasospasm and seizures. If the bleeding is severe, surgery to drain the blood and lower intracranial pressure may be required. Stroke prevention methods include removing cholesterol plaques from the brain's carotid arteries, expanding restricted carotid arteries with a balloon, and, in some cases, a stent. This is usually done with a catheter inserted at the groin. Various techniques, such as clipping and

embolization, are used to prevent brain aneurysms from rupturing. Removal or embolization of vascular malformations, bypassing the problematic artery.

Challenges in Stroke Disease Prediction

The prevention of stroke disease is done based on the history of the previous patients. So adequate data is needed to find the possibility of disease occurrence. Deep learning and machine learning are some of the AI techniques which entail a vast number of data for processing. Thus data necessity is the main challenge in stroke disease prediction. It is obligatory to collect the stroke patient record to predict future possibilities. Several stroke predictive scores have been followed in many countries such as NIHSS, ISAN, PNA for the sake of stroke prediction. [23]The score did not reliably distinguish between patients at low and high risk of recurrent stroke, nor did it identify patients with serious artery blockage who needed immediate treatment.

Challenges in Stroke Disease Detection

The various clinical tests and CT, MRI scanning are generally used to detect stroke illness. The neurological expects to have to detect the type of stroke and affected area in the image. Accurate detection is very important because the delay or erroneous detection leads to the spread of illness to other parts of the brain. The automation in stroke detection will improve the efficiency of treatment. Similar to prevention, detection also needs cumbersome data when it goes for the AI technique. When symptom-based stroke detection has been applied the variance in the type of stroke has to be identified correctly.

Challenges in Stroke Disease Treatment

The major challenge in stroke disease treatment is on-time admission tothe hospital. Because the deferral of every second causes the seriousness of the illness. [24]Advancements in drug intervention have shown improvement in Acute Ischemic Stroke treatment. Accurate detection of the damaged part is essential to give proper treatment. The severe health problem related to stroke disease is another challenge in treatment like pneumonia, which has to predict on time.

Selection of Sources and Inclusion Criteria

Several studies had done in stroke disease management using AI techniques like deep learning, Machine learning, and Neural Networks. This literature review divides the previous research work into three categories like stroke prediction, stroke detection, and stroke treatment. The search has been done based on the following keywords: stroke disease management, Stroke disease prediction, detection, treatment, stroke disease, and deep learning, Stroke disease, and machine learning. The research articles were found in the PubMed, Scopus, Elsevier, and PMS. Totally 567 articles were brought into being a consideration. Final filtration selects 52 articles which exclude review papers, stroke papers with AI techniques. This review article selects 30 papers among them.
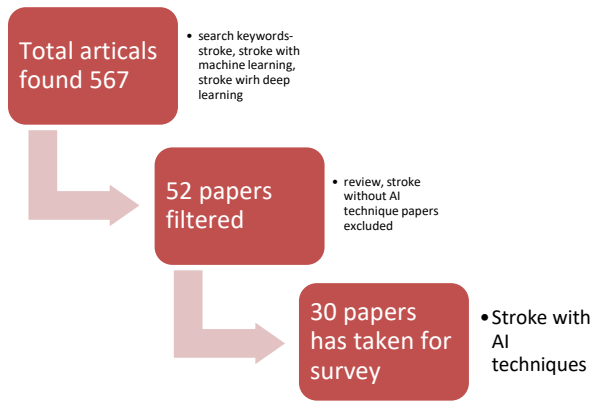
Fig. 1. Overview of manuscript preparation

Sources	Inclusion	exclusion
Dataset	Illness includes AIS and Hemorrhagic Includes demographic and clinical data Includes MRI and CT scan	Excludes the study which involves other than a human being
Techniques	Deep Learning, Machine learning	Excludes the study which does not involve AI techniques

Table 1. Selection and inclusion criteria of reviewed articles

Fig. 2. Year-wise representation of reviewed articles
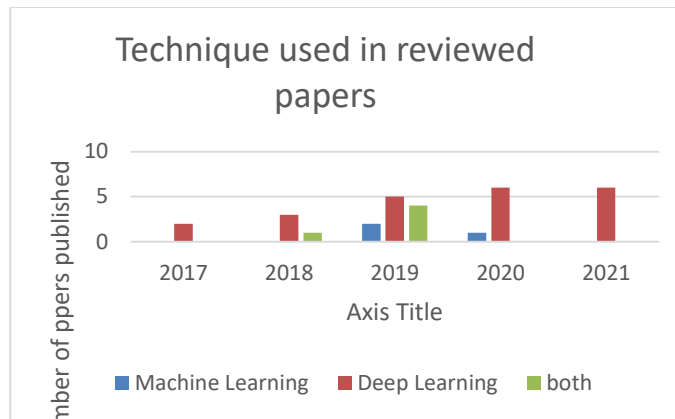
Fig. 3. Technologies used in the reviewed articles

Stroke Prediction

[4]suggested a deep learning system based on U-nets models to directly predict successful reperfusion after endovascular treatment (EVT) and good functional outcomes using CT angiography images. The authors compared the Deep Learning models to Machine Learning models using traditional radiological image biomarkers using Residual Neural Network (ResNet) architectures, Structured Receptive Fields (RFNN), and auto-encoders (AE) for network weight initialization. The models beat standard radiological imaging biomarkers in three of four cross-validation folds for functional outcome (average AUC of 0.71) and all folds for reperfusion, according to the researchers (average AUC of 0.65).

[5] demonstrated the machine learning algorithms such as deep neural network, random forest, and logistic regression to predict the long-term outcome of ischemic stroke patients. The authors highlighted that work is a retrospective study with registered ischemic stroke patients who are admitted within 7 days of the onset of symptoms. The result showed that deep neural networks performed better than other models. The limitation of this study is that was a single-center study and requires validation with data from other sources.

[10]identified 15,099 patients with stroke using a deep neural network that was trained on medical service use and health behaviour data. To extract significant background information from medical data, principal component analysis (PCA) with quantile scaling was utilized, whichwas then used to predict stroke mortality. Five other machine-learning approaches were compared to a scaled PCA/deep neural network [DNN] approach.

[14]presented novel a solution to cIMT (Carotid intima-media thickness) measurements using a Deep learning-based intelligent system. The overall concept was divided into four categories: multiresolution – based pre-processing (phase I), DL-based regional segmentation, Boundary extraction, calibration, and measurement, Performance analysis. In phase I the non-tissue information was removed by cropping. In Phase II the features of the images were extracted using VGG16 which acts as an encoder and FCN (decoder) created segmented images. Phase III further quantified the plaque burden using the g ML method (Polyline Distance Metric (PDM)). The performance analysis was also carried out and concluded that the Deep learning-based cIMT measurement which mainly influences the stroke disease comparatively high accuracy than conventional

methods.

[16] proposed a novel Hybrid Deep Transfer Learning-based Stroke Risk Prediction(HDTL) to address the imbalance in stroke data. This model allows external stroke data such as diabetes, hypertension etc., (which influences the stroke disease) in stroke risk prediction when the unavailability of stroke data in hospitals. The components of HDTL are: the training model using Generative Instance Transfer(GIT) applying GAN which can generate synthetic instances in external databases, Network Weight Transfer (NWT) utilize data from correlated diseases (Hypertension, diabetes etc.,), best-transferred parameters are identified using Bayesian Optimization (BO), the best informative synthetic instances are selected using Active Instance Transfer (AIT) which activate SRP (Stroke Risk Prediction) model.

[18]offered a cost-effective way of stroke disease prediction system using EEG(electroencephalography). Electroencephalography is a wearable electrode is attached to the head and quantity the brain nerve cells activity. The raw EEG data were extracted by using the Fast-Fourier Transform technique. EEG data are time series in nature in sequence values over time. To address these time series values the four deep learning methods (CNN, LSTM, bidirectional LSTM) were used. The performance of the system was analyzed by using precision, accuracy, F1-Score, specificity, sensitivity, FNR, and FPR. The experimental result showed that the Bidirectional LSTM model has the highest predictive accuracy at 89.2%.

[22]developed the machine learning model to predict the incidence of pneumonia after stroke. The authors specified that the ML model better predicts pneumonia than ISAN and PNA scores. These are the measuring scores that are used to predict the AIS in UK and Germany respectively. The Result showed that among six Machine learning models mentioned in the table, the XGBoost model outperformed.

[39]studied the deep learning methods to predict the infarct acute ischemic lesions in magnetic resonance imaging. Without knowing the status of future reperfusion, a deep learning network trained with a collection of acute and follow-up images was used to predict the size and location of infarct lesions 3 to 7 days after baseline. The model performed similarly to existing clinical state-of-the-art approaches in patients with minimal and major reperfusion.

Stroke Detection

[6]collected stroke patients' details from various sources and classified the stroke disease by using machine learning algorithms such as KNN, Naïve Bayes, J48, and Random Forest. The performance of the algorithm was evaluated by using accuracy, precision, recall, and F1-score. The authors concluded that Random –Forest performs well among four.

[12]automatically detected and analyzed the ischemic stroke lesion using deep learning methods.   For this, the authors used four deep learning approaches as Faster R-CNN (VGG-16), Faster R-CNN (ResNet-101), YOLOv3, SSD. They achieved more accurate results in SSD because SSD combined the regression idea from

YOLO and the anchor mechanism from Faster R-CNN.

[17] presented an automatic ischemic stroke detection system using CNN. This method starts with pre-processing the image such as, skull removal, CSF(cerebral Spinal Fluid) removal, creating fixed size patch images to avoid error during training, data augmentation and analyzing the images using FFNN(Feed-Forward Neural Network). The experimental results showed that the training stage had 97.65 and the testing stage had 92.96%.

[19]proposed an algorithm that extracts the main features of a stroke lesion from clinical volumes and combines them with normal DW volumes to construct realistic synthetic volumes of stroke lesions on diffusion-weighted (DW) MRI. By using combinatorial possibilities, this algorithm generated a large number of training datasets, and because the lesions are generated artificially, the exact position of the ground truth lesion is known. The lesion detection sensitivity was exceeded by the 3D U-Net, which was trained on a library of human-labeled clinical stroke volumes and then boosted with 40 000 synthetic stroke volumes.

[20]developed an automatic method to detect the existence of stroke using (Lenet) Convolution Neural Network. The technique also delineates the abnormal regions using Segnet architecture. The accuracy of the method is significant (97%) but the number of images used for detection is very less.

[30]developed deep symmetric CNN to identify the acute ischemic stroke using DWI and ACD images. The internal layers of deep-symmCNN such as convolutional layers, two paths of Inception Modules, fully connected layers, and subtraction layer with L-2 normalization outperformed well with conventional models.[31]CNN algorithms were developed Convolutional neural networks (CNNs) were used to construct algorithms for automatic segmentation of acute ischemia lesions on diffusion-weighted imaging (DWI), and they were compared to traditional techniques, such as thresholding-based segmentation. The limitation of this method is a manual interruption.[32]proposed a new framework for segmenting stroke lesions in DWI automatically. The framework is made up of two convolutional neural networks (CNNs): the EDD Net is an ensemble of two DeconvNets, and the MUSCLE Net is a multi-scale convolutional label evaluation net (MUSCLE Net) that evaluates the lesions discovered by the EDD Net to eliminate probable false positives.[33] On diffusion-weighted imaging (DWI), automatic segmentation of cerebral infarction is typically done using a fixed apparent diffusion coefficient (ADC) threshold.U-net, an encoder-decoder convolutional neural network, was used to train segmentation models. Two U-net models were created: a U-net (DWI+ADC) model that was trained on both DWI and ADC data, and a U-net (DWI) model that was exclusively trained on DWI data. There were a total of 296 subjects employed in the training and 134 in the external validation. [34]For the accurate automatic segmentation of acute ischemic stroke presented a novel 3D fully convolutional and densely connected convolutional network (3D FC-DenseNet).For semantic segmentation, 3D fully convolutional network (FCN) was used.The network was built on the concept of densely linked convolutional networks (DenseNets), which allow each layer to take as input all of the feature maps. [35]An ensemble of CNN trained on multiparametric DWI maps was compared to single networks trained on solo DWI parametric maps. CNN was trained on 116 subjects' DWI, ADC, and low b-value-weighted images

using a combination of DWI, ADC, and low b-value-weighted data. The network's performance was compared to that of other networks and ensembles of five networks.

[36]Using picture data from both ACS and PCS patients, researchers assessed ways for developing a convolutional neural network (CNN) for PCS lesion segmentation. Follow-up non-contrast computed CT scans of 1018 patients with ACS and 107 individuals with PCS were included in the study.Finally, the ACS-CNN was fine-tuned utilising PCS patients to assess the performance of transfer learning.

[37]suggested a deep learning system based on U-nets to automatically detect and segment hemorrhagic strokes in CT brain images. By modifying the network's input, adjusting the structure of different models, and experimenting with alternative training techniques, different U-net based CNN architectures were built.

Stroke Treatment

[7]compared several classical machine learning methods with deep neural network methods to predict pneumonia after stroke. Though the study predicts pneumonia after stroke, this can be considered as stroke treatment category. Because the major features used in this study are treatment parameters. The features used for this study were demographic information, vascular risk factors, assessment of nursing levels, lab test information, and medication information. The authors differentiate their work from previous studies by building prediction models for pneumonia patients in terms of specific time windows after the onset of stroke. [8]established and calculated a deep learning (DL)–based algorithm for assisting the selection of suitable patients with acute ischemic stroke for endovascular treatment based on 3-dimensional pseudo-continuous arterial spin labeling (pCASL). The authors mentioned that the study used the Arterial spin labeling (ASL) MRI technique to estimate the penumbral tissue for a further selection of patients for the treatment because this MRI technique does not have side effects, meanwhile, the contrast agent used in dynamic susceptibility contrast (DSC) perfusion magnetic resonance imaging (MRI) and CT techniques have side effects in patients.

[9]In a diverse sample of stroke patients, researchers used a deep learning approach based on convolutional neural networks (CNNs) to predict the severity of language dysfunction using 3D lesion pictures from magnetic resonance imaging (MRI). The performance of CNN was compared to that of conventional(shallow) machine learning approaches, such as ridge regression (RR) on the principal components of the pictures and support vector regression. A hybrid strategy based on reusing CNN's high-level characteristics as additional input to the RR model was also proposed.

[11]assessed brain MRI text using NLP and ML algorithms to predict a 3-month functional outcome in AIS patients. The NLP approach was implemented in three ways: word level, sentence level, and document level in the MRI text data. Various machine learning algorithms described in the table were used at different levels of the NLP process. And they concluded that document level vectorization in the radiology text report showed better performance in the prognosis of AIS. The authors also pointed out the limitation of this method that external validation is needed to demonstrate the performance in predicting the output.

[13] Foretold the functional outcome of Intravenous thrombolysis treatment by using the combination of ANN and CNN. The study predicted the improvement in NIHSS score at 24 hours, Modified Rankin scale at 90days. The conclusion stated that in both cases(NIHSS score prediction &mRS scale prediction) the combination of ANN and CNN is performed well.

[15] predicted the Hemorrhagictransformation(a serious problem that occurred after the EVT treatment) of stroke patients after Endovascular Thrombectomy (EVT) treatment by using multiparameter on DWI (Diffusion-Weighted Imaging) and Perfusion-Weighted Imaging (PWI). The experiment was done by using a single parameter(DWI, MTT, TTP, and clinical) and a multiparameter model. And the result showed that the multiparameter model was better than the single parameter model.

[24] designed and verified a predictive model capable of detecting and integrating acute ischemic stroke imaging data to properly predict eventual lesion volume for the . For this purpose, CNN was used. CNNdeep outperforms the generalised linear model in predicting the final outcome (AUC=0.880.12) and outperforms CNNshallow significantly.

Stroke Classification

[28] classified ASPECTS (Alberta Stroke Program Early Computed Tomographic Score) of acute ischemic stroke patients using DWI images. The classification algorithm is done by a recurrent residual convolutional neural network (RRCNN). The performance of RCNN was compared with Inception V3, VGG16, , and a 3D convolutional neural network (3DCNN). The conclusion stated that RCNN classifier performed well and could assist the physician in decision making.

Reference number, Authors & Year	Stroke type	Objective	methodology	Methodology type	samples	Features	Feature Description	Accuracy
[4]Hilbert A et al.,[2019]	AIS	Predict reperfusion after Treatment (endovascular)	DL and ML	(ResNet) (RFNN) (AE)	MR CLEAN Registry dataset with 1301 patients	CT Image Data	Image biomarkers	Not mentioned
[5]JoonNyung Heo et al.,[2019]	AIS	Predict long-term outcomes	ML	DNN RF LR	2604 patients data	Demographic and Clinical data	NIHSS score mRS score etc.,	78%
[6]Tasfia Ismail Shoilyetal[2019]		Detecting stroke disease using machine learning algorithms	ML	Naive Bayes, k-NN, J48, RF	1058 stroke patients data	Demographic and Clinical data	23 features including MRI, CT, ECG data	Not mentioned
[7]Yanqiu Ge et al.,[2019]	AIS	Post-stroke pneumonia	Machine learning and	ML: LR	13,930 EHR (Electronic	Demographic and	Lab test, Medications	AUC -0.928

		prediction	deep learning methods	SVM XGBoost DNN methods: MLP GRU	Health Record) data set	Clinical data	etc.,	
[8]Kai Wang et al.,[2019]	AIS	Treatment (Endovascular)	DL and ML	ML: LRC, RRC, KRRC, NN, SVM with radial basis function, and RRC DL: Highres3Dnet	167 image set	MRI	3-dimensional pseudo-continuous arterial spin labeling (pCASL), d dynamic susceptibility contrast	92%
[9]Sucheta Chauhan et al.,[2019]		Treatment	DL and ML	ML methods: 1. RR 2. SVM	132 stroke patients	MRI	Image Feature	Not mentioned
[10]Songhee Cheon et al.,[2018]	AIS and Hemorrhagic	Prediction	DL model featuring scaled PCA	FFNN	15099 patients health record	medical service use and health behavior data	Demographic information Charlson Comorbidity Index (CCI) score etc.,	84.03
[11]Tak Sung Heo et l.,[2020]	AIS	Treatment	NLP and Multi CNN-DL algorithms	LASSO regression SDT RF, SVM, CNN, MLP LSTM, Bi-LSTM CNN&LSTM CNNmax Multi-CNN Bi-LSTM-bidirectional LSTM;	1840 radiology text report	MRI text reports	Demographic information NIHSS scale, mg/dl Dyslipidemia etc.,	80.5
[12]Shujun Zhang et al.,[2021]	AIS	Detection	DNN	Faster R-CNN, YOLOV3, and SSD	300 patients record	5668 MR images	Image features	89.77
[13]Stephen Bacchi et al.,[2019]	AIS	Treatment	DL	CNN and ANN	204 patients images	CT brain images	Clinical data and CT image features	74
[14]Mainak Biswas et		Prediction of the risk factor using	DL and ML	VGG16 network	204 patients images	396 B-mode ultrasound	1.cIMT 2.LI	AUC – 90%

al.,[2018]		cIMT measurement				images	3.MA	
[15]Liang Jiang et al.,[2021]	AIS	Predict the Hemorrhagic transformation after Endovascular Thrombectomy Treatment	CNN	Inception V3	568 patients	MRI (DWI and PWI images)	1.DWI 2.MTT 3.TTP 4. Clinical 5.CBF 6.CBV	Multi-parameter prediction is better than a single parameter.
[16]Jie Chen et al.,[2021]	General	Stroke risk prediction using hybrid deep transfer learning	DNN	GAN	2, 426 stroke incidents record	Hybrid database(Hypertension, diabetes etc.,)	EHR	Not mentioned
[17]Chiun-Li Chin et al.,[2017]	AIS	Automatic early detection	CNN	Otsu algorithm(preprocessing) Feed-Forward ANN.	256 patch images	MR images	Image features	92.96%
[18]Yoon-A Choi et al.,[2021]		Stroke disease prediction system using Real-time bio signals	CNN	LSTM, Bidirectional LSTM, CNN-LSTM, and CNN-Bidirectional LSTM	273 patients	Bio-signal data	non-invasive scalp EEG EEG	94%
[19]Christian Federau et al.,[2020]	AIS	Segmentation and detection of DWI images	DL	3-D U-Net	962 stroke images + 2027 Normal images	MRI	Image feature	96%
[20]Bhagyashree RajendraGaidhani et al.,[2019]	AIS	Stroke detection using CNN and deep learning	CNN and DL models	Classification - LeNet Segmentation-SegNet Prepocessing: Noise removal – Median filter	406 images	MRI	Image feature	Classification- 96% - 97% Segmentation – 85-87%
[22]X.Li et al.,[2020]	AIS	Learning	ML	LR with regulation, SVM, RFC, extreme gradient boosting and fully-connected DNN	3160 AIS patients	Demographic and Clinical Data including NIHSS,	NIHSS, ISAN, PAN	90%

[23]Anis FitriNurMasruriyah,[2020]		Predictive Analytics using ANN	DL	ANN	18425 patients data	Demographic and Clinical Data	10 features including CT scan result, the deficit of dysphasia etc.,	95.15%
[24]Anne Nielsen et al.[2018]	AIS	Functional outcome of AIS treatment	DL	CNN	222 patients record	MRI	MRI image data	Not mentioned
[28]Luu-Ngoc Do et l.[2020]	AIS	Classification of DWI imaging	DL	RRCNN	319 datasets	CT	CT–DWI images	87.3%
[30]Liyuan Cui et al.[2021]	AIS	Identifying AIS via DWI images	DL	DeepSym3D-CNN	190 dataset	MRI	DWI and ADC images	AUC=0.86 %
[31]Ilsang Woo et al.[2019]	AIS	Automatic segmentation of AIS lesions on DWI images	DL	CNN	429 patients data	MRI	DWI and ADC images	AUC=0.85 %
[32]Liang Chen et al[2017]	AIS	Segmentation	DL	CNN EDD net MUSCLE net	741 subjects	MRI	DWI	94%
[33]Yoon-Chul Kim et al.[2019]	AIS	Segmentation	DL	Encoder – Decoder CNN	296 subjects	MRI	DWI and ADC images	ICC-95%
[34]Rongzhao Zhang et al.[2018]	AIS	Segmentation	DL	3D-CNN	242 subjects	MRI	DWI and ADC images	92.67%
[35]Stefan Winzecket al.[2019]	AIS	Segmentation	DL	CNN	116 subjects	MRI	DWI and ADC images	Not mentioned
[36]Riaan Zoetmolte et al.[2021]	AIS	Segmentation	DL	CNN	1018-ACS patients 107-PCS patients	MRI	DWI and ADC images	ICC-95%
[37]Lu Li et al.[2020]	Hemorrhagic	Detection and Segmentation	DL	U-net	190 patients	CT	Image features	98.59%
[39]Yannan Yu et al.[2020]	AIS	Predictio	DL	U-net	182 patients	MRI	PWI, DWI - MRI	92%

AIS- Acute Ischemic Stroke;DL – Deep Learning; ML- Machine Learning;LASSO-least absolute shrinkage and selection operator; SDT-single decision tree; RF-random forest; SVM-support vector machine; CNN-convolutional neural network;MLP-multilayer perceptron; LSTM-long short-term memory;VGG – Visual Geometry Group;FCN – Fully Convolutional Network;cIMT- Carotid intima-media thickness;LI - Lumen-Intima;MA - Media-Adventitia;CBF – Cerebral Blood Flow;CBV – Cerebral Blood Volume;MTT – Mean Transit Time;DWI – Diffusion-Weighted

Imaging;TTP – Time to Peak;GAN – Generative Adversarial Network;HER – Electronic Health Record;EEG – ElectroEncephaloGraphy, AUC-Area Under Cover, PCA- Principle Component Analysis, rfc-random forest classifier; ADB, AdaBoost classifier: GNB, Gaussian naive Bayes; KNNC, K-nearest neighbor classifier; SVC, support vector machine; DNN, deep neural network, RRCNN- a recurrent residual convolutional neural network, DWI- Diffusion-weighted Images, ADC- Apparent Diffusion Coefficient,icc- intraclass correlation coefficient

Conclusion: Research gap and Future opportunities

This review article deliberates recent AI-based stroke disease management. It discussed the type of dataset used and the type of methodology involved in the prognosis, detection, and treatment of stroke. Several studies evidenced that deep learning methods achieved good results when compared to machine learning. But a substantial amount of data is needed to achieve a better outcome. Based on the above review several works lack dataset collection

References

[1]   Sirsat MS, Fermé E, Câmara J. Machine learning for brain stroke: a review. Journal of Stroke and Cerebrovascular Diseases. 2020 Oct 1;29(10):105162.Esenwa C, Gutierrez J. Secondary stroke prevention: challenges and solutions. Vascular health and risk management. 2015;11:437.

[2]   Esenwa C, Gutierrez J. Secondary stroke prevention: challenges and solutions. Vascular health and risk management. 2015;11:437.

[3]   Feigin VL, Krishnamurthi RV, Parmar P, Norrving B, Mensah GA, Bennett DA, et al. Update on the global burden of ischemic and hemorrhagic stroke in 1990-2013: the GBD 2013 study. Neuroepidemiology. 2015;45(3):161-76.

[4]   Hilbert A, Ramos LA, van Os HJ, Olabarriaga SD, Tolhuisen ML, Wermer MJ, et al. Data-efficient deep learning of radiological image data for outcome prediction after endovascular treatment of patients with acute ischemic stroke. Computers in biology and medicine. 2019 Dec 1;115:103516.

[5]   Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning–based model for prediction of outcomes in acute stroke. Stroke. 2019 May;50(5):1263-5.

[6]   Shoily TI, Islam T, Jannat S, Tanna SA, Alif TM, Ema RR. Detection of stroke disease using machine learning algorithms. In2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) 2019 Jul 6 (pp. 1-6). IEEE.

[7]   Ge Y, Wang Q, Wang L, Wu H, Peng C, Wang J, Xu Y, Xiong G, Zhang Y, Yi Y. Predicting post-stroke pneumonia using deep neural network approaches. International journal of medical informatics. 2019 Dec 1;132:103986.

[8]   Wang K, Shou Q, Ma SJ, Liebeskind D, Qiao XJ, Saver J, et al. Deep learning detection of penumbral tissue on arterial spin labeling in stroke. Stroke. 2020 Feb;51(2):489-97.

[9]   Chauhan S, Vig L, De Filippo De Grazia M, Corbetta M, Ahmad S, Zorzi M. A comparison of shallow and deep learning methods for predicting cognitive performance of stroke patients from MRI lesion images. Frontiers in neuroinformatics. 2019 Jul 31;13:53.

[10]   Cheon S, Kim J, Lim J. The use of deep learning to predict stroke patient mortality. International journal of environmental research and public health. 2019 Jan;16(11):1876.

[11]   Heo TS, Kim YS, Choi JM, Jeong YS, Seo SY, Lee JH, et al. Prediction of Stroke Outcome Using Natural Language Processing-Based Machine Learning of Radiology Report of Brain MRI. Journal of personalized medicine. 2020 Dec;10(4):286.

[12]   Zhang S, Xu S, Tan L, Wang H, Meng J. Stroke Lesion Detection and Analysis in MRI Images Based on Deep Learning. Journal of Healthcare Engineering. 2021 Apr 10;2021.

[13]   Bacchi S, Zerner T, Oakden-Rayner L, Kleinig T, Patel S, Jannes J. Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: a pilot study. Academic radiology. 2020 Feb 1;27(2):e19-23.

[14]   Biswas M, Kuppili V, Araki T, Edla DR, Godia EC, Saba L, et al. Deep learning strategy for accurate carotid intima-media thickness measurement: an ultrasound study on Japanese diabetic cohort. Computers in biology and medicine. 2018 Jul 1;98:100-17.

[15]   Jiang L, Zhou L, Yong W, Cui J, Geng W, Chen H, et al. A deep learning-based model for prediction of hemorrhagic transformation after stroke. Brain Pathology. 2021 Oct 4:e13023.

[16]   Chen J, Chen Y, Li J, Wang J, Lin Z, Nandi AK. Stroke Risk Prediction with Hybrid Deep Transfer Learning Framework.

[17]   Chin CL, Lin BJ, Wu GR, Weng TC, Yang CS, Su RC, et al. An automated early ischemic stroke detection system using CNN deep learning algorithm. In2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST) 2017 Nov 8 (pp. 368-372). IEEE.

[18]   Choi YA, Park SJ, Jun JA, Pyo CS, Cho KH, Lee HS, et al. Deep Learning-Based Stroke Disease Prediction System Using Real-Time Bio Signals. Sensors. 2021 Jan;21(13):4269.

[19]   Federau C, Christensen S, Scherrer N, Ospel JM, Schulze-Zachau V, Schmidt N, et al. Improved segmentation and detection sensitivity of diffusion-weighted stroke lesions with synthetically enhanced deep learning. Radiology: Artificial Intelligence. 2020 Sep 16;2(5):e190217.

[20]   Gaidhani BR, Rajamenakshi RR, Sonavane S. Brain Stroke Detection Using Convolutional Neural Network and Deep Learning Models. In2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT) 2019 Sep 28 (pp. 242-249). IEEE.

[21]   Cheon S, Kim J, Lim J. The use of deep learning to predict stroke patient mortality. International journal of environmental research and public health. 2019 Jan;16(11):1876.

[22]   Li X, Wu M, Sun C, Zhao Z, Wang F, Zheng X, Ge W, Zhou J, Zou J. Using machine learning to predict stroke-associated pneumonia in Chinese acute ischaemic stroke patients. European journal of neurology. 2020 Aug;27(8):1656-63.

[23]   Nasr DM, Brown RD. The challenges of stroke prediction scores. JAMA neurology. 2016 May 1;73(5):510-1.

[24]   Nielsen A, Hansen MB, Tietze A, Mouridsen K. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. Stroke. 2018 Jun;49(6):1394-401.

[25]   Lee EJ, Kim YH, Kim N, Kang DW. Deep into the brain: artificial intelligence in stroke imaging. Journal of stroke. 2017 Sep;19(3):277.

[26]   Zhu G, Jiang B, Chen H, Tong E, Xie Y, Faizy TD, Heit JJ, Zaharchuk G, Wintermark M. Artificial intelligence and stroke imaging: A west coast perspective. Neuroimaging Clinics. 2020 Nov 1;30(4):479-92.

[27]   Yedavalli VS, Tong E, Martin D, Yeom KW, Forkert ND. Artificial intelligence in stroke imaging: Current and future perspectives. Clinical Imaging. 2021 Jan 1;69:246-54.

[28]   Do LN, Baek BH, Kim SK, Yang HJ, Park I, Yoon W. Automatic Assessment of ASPECTS Using Diffusion-Weighted Imaging in Acute Ischemic Stroke Using Recurrent Residual Convolutional Neural Network. Diagnostics. 2020 Oct;10(10):803.

[29]   Gupta R, Krishnam SP, Schaefer PW, Lev MH, Gonzalez RG. An East Coast perspective on artificial intelligence and machine learning: part 2: ischemic stroke imaging and triage. Neuroimaging Clinics. 2020 Nov 1;30(4):467-78.

[30]   Cui, L., Han, S., Qi, S., Duan, Y., Kang, Y. and Luo, Y., 2021. Deep symmetric three-dimensional convolutional neural networks for identifying acute ischemic stroke via diffusion-weighted images. *Journal of X-Ray Science and Technology*, (Preprint), pp.1-16.

[31]   Woo I, Lee A, Jung SC, Lee H, Kim N, Cho SJ, et al. Fully automatic segmentation of acute ischemic lesions on diffusion-weighted imaging using convolutional neural networks: comparison with conventional algorithms. Korean journal of radiology. 2019 Aug 1;20(8):1275-84.

[32]   Chen L, Bentley P, Rueckert D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. NeuroImage: Clinical. 2017 Jan 1;15:633-43.

[33]   Kim YC, Lee JE, Yu I, Song HN, Baek IY, Seong JK, Jeong HG, Kim BJ, Nam HS, Chung JW, Bang OY. Evaluation of diffusion lesion volume measurements in acute ischemic stroke using encoder-decoder convolutional network. Stroke. 2019 Jun;50(6):1444-51.

[34]   Zhang R, Zhao L, Lou W, Abrigo JM, Mok VC, Chu WC, Wang D, Shi L. Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional DenseNets. IEEE transactions on medical imaging. 2018 Mar 30;37(9):2149-60.

[35]   Winzeck S, Mocking SJ, Bezerra R, Bouts MJ, McIntosh EC, Diwan I, et al. Ensemble of convolutional neural networks improves automated segmentation of acute ischemic lesions using multiparametric diffusion-weighted MRI. American Journal of Neuroradiology. 2019 Jun 1;40(6):938-45.

[36]   Zoetmulder R, Konduri PR, Obdeijn IV, Gavves E, Išgum I, Majoie CB, et al. Automated final lesion segmentation in posterior circulation acute ischemic stroke using deep learning. Diagnostics. 2021 Sep;11(9):1621.

[37]   Li L, Wei M, Liu B, Atchaneeyasakul K, Zhou F, Pan Z, et al. Deep learning for hemorrhagic lesion detection and segmentation on brain ct images. IEEE Journal of Biomedical and Health Informatics. 2020 Oct 1;25(5):1646-59.

[38] Shafaat O, Bernstock JD, Shafaat A, Yedavalli VS, Elsayed G, Gupta S, et al. Leveraging Artificial Intelligence in Ischemic Stroke Imaging. Journal of Neuroradiology. 2021 May 11.

[39] Yu Y, Xie Y, Thamm T, Gong E, Ouyang J, Huang C, et al. Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging. JAMA network open. 2020 Mar 2;3(3):e200772-.

Review On analysis techniques of Consumer Preferences in Online Matrimonial Portals

Dr. Ramah Sivakumar, V.Ramakrishnan

Department of Computer Science, Bishop Heber College, Tiruchirappalli-620 017.

ramahsivakuamr.cs@bhc.edu.in

Abstract

Marriage is a long- standing custom in India, where it is seen as a holy connection of the heart and soul, as well as a merger of two families, the bride and groom. Presently, an online matrimonial portal is the most common platform for weddings since it eliminates geographical restrictions and it is considered more trustworthy than conventional methods. Millions of users share their personal profiles to find their life partner in a single click. Users may read registered profiles which contain information such as name, religion, caste, legal status, current salary, occupation, and so on. Photo and video sharing is also possible in these sites. The registered members could search profiles based on their preferences by age, religion, caste, and other factors. This paper is a review of consumer preferences analysis techniques in online matrimonial portals such as Percentage Analysis, Questionnaires, Second Order Co-occurrence PMI (SOC-PMI) and so on.

Keywords: Marriage, matrimony, consumer preference

Introduction

Wedding is a greater event in Indian families which intensely evoke the potential social obligation, bonding between two families, traditional values, sentiments and financial resources. There may be some variations across the country with respect to the rituals they perform, as per their religion, caste, and region. And finding a partner principally depends on the social standing of the family, the caste and sub caste, educational qualifications, profession, coloring, physical choices and appearance. Folks say "Matches are made in heaven" on the other hand presently this quote has become "Matches are made in online matrimonial websites". In 1700, once the invention of the trendy newspaper the primary legal status service was created. Brides and grooms trying to hunt out their matches could search the knowledge that typically embraces position, age, gender, photograph, location caste and so on. Some of the features of a successful matrimonial websites include video consultation, booking an appointment; calendar setup, pay online, rating and review for consultation and multilingual apps. In online matrimonial, users could build profiles with their preferences and connect via the web interface. Consumer preferences can be well-defined as the subjective (individual) tastes, as measured by wish and expectations among various bundles of profiles. User satisfaction is one amongst the common ideas that intangibly to be measured. They permit the consumer to rank those profiles according to the level of satisfaction.

The consumer satisfaction might need a control on customer's perception towards loyalty and its relationship. Moreover, from the literatures it was found out that a replacement dimension on the standard

of perspective that get utilized in determining the priority in wedding life would begin relationship, loyalty and wedding satisfaction. The rise of loyal customer retention is the key issue for a long-term success of matrimonial portals. For that Consumer preferences should be deeply analyzed to take appropriate decisions for improvements.

Back ground and objectives

On utilizing the information provided by the customer, matrimonial services assist in the pairing process. It assists the user by providing an online list of profiles of eligible grooms and brides, as well as various information about them. Users will be able to access information about their ideal life partner at their leisure from the comfort and privacy. Wedding site services also include a search function to assist users who are looking for a specific set of qualities to make the search easier and more effective. Since internet is a pivot for modern business, internet-based marriage services provide the way for modernization in matrimonial search.

Some of the methodologies adopted for analyzing consumer preferences

1. *Online questionnaire*

An online questionnaire is a set of questions designed to collect information about a certain audience or group of individuals via the internet. Questionnaire analysis includes pilot study to know about the preferences, measure customer satisfaction, and administration of the questionnaires. The primary data might be collected using a pre-structured questionnaire. As random selection is problematic, the respondents could be picked using convenience sampling. Charts and graphs could be used for data analysis.

2. *Percentage analysis*

Percentage analysis is one of the basic statistical tools which is widely used in analysis and interpretation of primary data. Percentage analysis is used in making comparison between two or more series of data. A percentage is used to determine relationship between the series. The frequency in the category is divided by the total number of participants and multiplied by 100 to get the percentage.

Percentage analysis can be calculated using the formula,

Percentage= (No.of Respondents / Total No.of Respondents)*100

The use of percentages to compare samples with varying numbers of observations is a strong tool. Samples can be compared quickly and readily by standardizing measures on a scale of 0 to 100. Any graph of the data, on the other hand, must encompass the entire range of 0 to 100 to avoid creating false impressions.

3. *Conceptual Model*

A conceptual model is a framework that is first used in research to outline the potential actions and to communicate an idea or thinking. Conceptual framework could be used for analytical purpose as it provides a visual representation of theoretical constructs (and variables) of interest. The research process will be more rigorous if a conceptual model is constructed in a logical manner. In consumer preferences analysis the factors such as ease of use, search, match, interaction, convenience, website features, and trust could influence the preference of online matrimonial portals over traditional methods.

4. *Second Order Co-occurrence PMI Method*

Second-order co-occurrence pointwise mutual information is a semantic similarity metric in computational linguistics. To determine the degree of link between two words, it sorts lists of important neighbor words of the two target terms from a huge corpus using pointwise mutual information (PMI). To compute the relative semantic similarity, the approach evaluates the terms that are frequent in both lists and aggregates their PMI values (from the opposite list). Only those words with pointwise mutual information would be considered. Adjectives are the center of our attention since they are good indicators of the traits that a person seeks in a potential spouse. The SOC-PMI approach is used for each pair, and the list is sorted by PMI values. This strategy has benefits in the automated grouping of semantically comparable words (here adjectives). One of the key benefits of SOC-PMI over PMI is semantically comparable terms throughout the whole corpus could be located, even if they don't co-occur inside the given context window frame.

5. *Cronbach's alpha*

Cronbach's alpha is a measure of internal consistency, that is, how closely related a set of items as a group. It is considered to be a measure of scale for reliability. A "high" value for alpha does not imply that the measure is unidimensional. Cronbach's alpha is most commonly used when internal consistency of a questionnaire of multiple Likert-type scales and items (or survey) is to be assessed. These questions measure latent variables- hidden or unobservable variables like: a person's conscientiousness, neurosis or openness which is very difficult to measure in real life. Cronbach's alpha is most valuable for indicating scale reliability in the sense of the equivalence of items within single-construct scales; however the statistics does not offer any indication regarding unidimensional scales (which should be tested by other means). Cronbach's alpha would also assess the designed test is accurately measuring the variable of interest. It is used under the assumption that multiple items measure the same underlying construct.

6. *DG's Model*

The discontinuous Galerkin (DG) is a robust and compact finite element projection method that provides a practical framework for the development of high-order accurate methods using unstructured grids. The method is well suited for large-scale time-dependent computations in which high accuracy is required.

In this method, the integral of the Residual equation, R, multiplied by a weighting function, φ, is forced to be zero and the weighting functions are chosen to be in the same form. Since the methods use completely discontinuous approximations, they produce mass matrices that are block-diagonal. This renders the methods highly parallelizable when applied to hyperbolic problems. Another consequence of the use of discontinuous approximations is that these methods can easily handle irregular meshes with hanging nodes and approximations that have polynomials of different degrees in different elements. They are thus ideal for use with adaptive algorithms. Moreover, the methods are locally conservative (a property highly valued by the computational fluid dynamics community). Even more, when applied to non-linear hyperbolic problems, the discontinuous Galerkin methods are able to capture highly complex solutions presenting discontinuities with high resolution.

Review of Literature

P. Sailasri et.al., [1] had made a study on the preferences of online matrimonial visitors. The primary data was collected using a pre-structured questionnaire. Since random selection is problematic, the respondents were picked using convenience sampling. Charts and graphs were used to analyze the data. Majority of the members (82%) had their first choice of those profiles who shared their interests first. Educational credentials, job description, family background, physical characteristics, and other family assets were all taken into consideration. The purpose of the study was to learn about customer preferences when it came to selecting online matrimonial services. Consumers choose sites that are simple to use and ensure privacy and availability of information. The study had limitations, as it was a preliminary study that requires a bigger sample size and analysis to get precise results.

R.Monisha [2] conducted a study in Coimbatore city to determine why customers choose online matrimonial arrangements over offline matrimonial arrangements and also to learn about the users' perceptions and attitude towards such sites, as well as their issues with them. The investigations were on various sites and services offered by online matrimony agencies. The purpose of this study was to find out users' degree of satisfaction with matrimony websites, and to investigate the issues that the users experience with these websites, as well as their influence on their lives. This also includes users' preferences for such websites and how the general public perceives them. This research was limited to the city of Coimbatore. For analysis, percentage analysis technique had been used. From the findings it was observed that, as people are not comfortable with traditional search of bride or groom, they register in matrimonial websites. Instead of resorting to marriage brokers, they believe it saves time and money. From the results it was found that, comparatively female respondents register more than male respondents, and respondents are happy with online marriage services.

Ayman Bajnaid et.al.,[3] surveyed the Saudis who use the chosen Muslim matrimonial website, the subject of the current study, are geographically dispersed throughout the Kingdom of Saudi Arabia, an online questionnaire embedded within the matrimonial websites is an effective way to locate, survey, and collect data from a large sample of them. The questionnaire was submitted to the representative of the

matrimony service. In April 2015, the agency placed a month-long ad on the website to attract users. The purpose of this advertisement was to notify website visitors about the study and how they might join. The researchers eliminated 25 replies because they were incomplete. There were 302 participants in this phase (134 females and 168 men) ranging in age from 18 to 65. The participants were 55.6 percent male, 53.3 percent Hadari, and ranged in age from 18 to 65, with the majority being between 31 and 40 years old (62.9 percent). The poll included 64.9 percent singles, with 75.4 percent identifying as moderately religious. About half of the people in the sample (50.3 percent) had a bachelor's degree, and the majority had an annual salary of SAR4000 to SAR9999. These demographic parameters are reflective of the target group and align with the demographic features of the website's subscription base.

Neethu Krishnan et.al., [4] examined sample population of 100 potential brides and grooms who have paid to join online marriage portals. To analyse the variables' preferences, an online survey was conducted utilizing a questionnaire. On a 5-point Likert scale, 1 = "Strongly Disagree" to 5 = "Strongly Agree," respondents' preferences for online matrimonial sites were evaluated. The factors influencing online matrimonial portal choices are treated as independent

(X) variables, while customer satisfaction is treated as a dependent (Y) variable. This online survey was done in Kerala to collect user responses in order to evaluate with a structural model as provided by the conceptual model. According to the premise, client happiness is influenced by the characteristics such as search, ease, website features, and trust. Search, convenience, website features, and trust are all shown to have a substantial impact on user satisfaction in structural equation modelling (SEM). According the findings, the suggestions given for the acceptance of online matrimony services and to increase consumer satisfaction, tailored search should be provided and the websites should be user- friendly. The customer Satisfaction Index (FSI) ratings would be greater if the user obtains more prospective profiles on their search page. Since customer satisfaction is strongly correlated with trust, online portals should adopt a privacy policy for their customers regarding the exposure of personal information.

Nazia Tabassum et.al., [5] ensures that some adjectives might help users to figure out the attributes or characteristics that are seeked for a companion. These adjectives are classified or clustered based on co-occurrence of words using the SOC-PMI approach for semantic similarity. The headword, which is one of the permanent qualities, is mapped to these terms. By adding weightage to each fixed attribute, the list of fixed characteristics is prioritized through Weighted Reciprocal Recommendation model. The developed model leverages SOC- PMI to automatically classify adjectives extracted from free-text on matrimonial websites, which helped to construct this Weighted Reciprocal Recommendation model for more efficient suggestions. The model had been presented using SOC-PMI for more efficient recommendations. The experimental evaluation shows promising results and it is proposed to implement the methodology in both symmetric and asymmetric reciprocal domains.

Siti Norashikin Bashirun et.al., [6] ensures to determine the relevance and link between customer satisfaction and marital satisfaction in terms of service quality. The results of a reliability test demonstrate that all variables were far over 0.80. Cronbach's Alpha for the five variables varied from 0.867 (communication) to 0.910 (communication) (reward). The information was gathered via a questionnaire that was randomly given to married couples. This study examines 190 married couples, which gave a

sample response rate of 95% for the population. According to the poll, Malay people accounted for 86.8% of the respondents, followed by Chinese 6.3 percent, Indian 5.7 percent, and other or perhaps foreigners 1.6 percent. It was also discovered that the majority of respondents (41.1%) were SPM (Sijil Pelajaran Malaysia) holders. The major goal of this research is to use a customer service strategy to discover elements that influence marriage satisfaction. According to the data, "reward" (mean: 5.097) has the highest mean among the other factors. This study's findings show that reward, communication, behavior, responsiveness, activities, and production are all linked to total marriage happiness among spouses.

Edoardo Ciscato et.al., [7] focus on concept of infer marital preferences (At,t) from cross- sectional data on couples (Xt, Yt) for a given year t, and then compute the equilibrium matching P (s, s;t) for population data (Xs, Y s) under the same marital p references. Counter factual distribution of couples' attributes was computed using P (s, s;t) and data (Xs, Y s). The aim of the study was to assess   changes in marriage preferences   based on the change in the distribution of household characteristics. In the United States, matching patterns from 1964 to 2017 give a thorough picture of the evolution of marital preferences. Positive assortative mating on schooling has been stronger through time, while household specialization has weakened as age and wealth disparity have increased. On the marriage market, the importance of socio-economic observable features has reduced. Since 1970s, racial homogamy appears to have grown in popularity. Whites and Hispanics are becoming less and less attracted to one another, and spouses' education is decreasing as they become older.

Discussion

It is inferred from the reviewed papers that four techniques such as online questionnaire, DG Model, Conceptual model and SOC PMI model are considered significant in consumer preference analysis. Questionnaires are popular research method as they offer a fast, efficient and inexpensive means of gathering large amounts of information from sizeable sample volumes. Online questionnaires reduce the research costs. Responses could be processed automatically and the results are accessible at any time. Percentage analysis refers to a special kind of rates, percentage used in making comparison between two or more series of data. A percentage is used to determine relationship between the series which is simple to implement and easy to understand. It also helps in comparing the numbers of visits between different time periods. It helps in understanding the percentage or share   of the individual items, and the structural composition of components. The discontinuous Galerkin (DG) method is a robust and compact finite element projection method that provides a practical framework for the development of high-order accurate methods using unstructured grids. The method is well suited for large-scale time-dependent computations in which high accuracy is required. For dynamic mechanical problems the Galerkin method is used as a global/spectral method to help in reducing the Partial Differential Equation (PDE) into ordinary differential equations. A conceptual model is a framework that is initially used in research to outline the possible courses of action or to present an idea or thought. When a conceptual model is developed in a logical manner, provides a rigor to the research process. Second Order Co-occurrence PMI Method is a statistical method that uses pointwise mutual information to shortlist the important neighbour words of the two target words. Then common words in both lists are considered and their PMI values (from the opposite

list) are aggregated. It is found that Second Order Co-occurrence PMI method provides the best result among all statistical approaches and also comparable to all lexical approaches. It can determine the semantic similarity of two words even though they do not co-occur with the window size. These techniques are particularly effective for measuring subject behavior, preferences, intentions, attitudes and opinions.

Conclusion

This review paper mainly focuses on the analysis techniques of consumer preferences in online matrimonial portals. Some of the methodologies discussed were Percentage analysis, DG Model, Conceptual model, SOC-PMI model, Online questionnaire, and Cronbach's alpha. Each method has its own pros and cons on analyzing the consumer preferences on matrimonial portals. By proper analysis of consumer preferences the portals' could be improved so that it might help user in payment premium, chat, search and filter data, image transfer, time and money saving and so on.

References

[1]   P. Sailasri, "Consumer Preferences Towards A Online Matrimony Portals Â A Study On Matrimony Portals In Hyderabad ", Business Sciences International Research Journal, ISSN :2321 – 3191, Volume 8, Received: Jan. 2020 Accepted: Feb. 2020 Published: Feb. 2020.

[2]   R.Monisha, "A Study On Users' Preference Towards Matrimonial Sites In Coimbatore City", international journal of research in commerce, it & management , ISSN: 2231-5756. volume no. 5 (2015), ISSUE NO. 12 (December)

[3]   Ayman Bajnaid, & Giuseppe Alessandro Veltri, "A Quantitative Survey of Online Impression Formation and Mate Preferences among Saudi Users of Matrimonial Websites", Published by Canadian Center of Science and Education, ISSN :1913-1844E-ISSN 1913-1852, Vol. 12, No. 11; 2018, Received: August 16, 2018 Accepted: September 20, 2018 Online Published: October 29, 2018.

[4]   Neethu Krishnan, Hemamala k, HariKrishnan.A.V, "Study On The Factors Affecting The Preferences Of Online Matrimonial Portals Over Traditional Methods". International Journal of Pure and Applied Mathematics, ISSN: 1314-3395 vol. Volume 118 No. 22 2018. pp 20 march 2021.

[5]   Nazia Tabassum, & Tanvir Ahmad. "Extracting Users' Explicit Preferences from Free-text using Second Order Co-occurrence PMI in Indian Matrimony". International Conference on Computational Intelligence and Data Science (ICCIDS 2019), The Authors. Published 2020.

[6]   Siti Norashikin Bashirun, & Nur Idayu Badrolhisam. "Customer Service Approach on Determinants of Marriage Satisfaction". Advances in Economics, Business and Management Research, volume 57, (PICEEBA 2018)

[7]   Edoardo Ciscato, & Simon Weber, " The role of evolving marital preferences in growing income

inequality". Journal of Population Economics, Received: 29 June 2017 / Accepted: 5 June 2019 /published online:03 July 2019.

[8]   Sonora Jha & Mara Adelman. "Looking for Love in All the White Places: A Study of Skin Color Preferences on Indian Matrimonial and Mate-Seeking Websites*"*. Studies in South Asian Film and Media Volume 1 Number 1 © 2009 Intellect Ltd.

[9]   David M. Buss. "A Half Century of Mate Preferences: The Cultural Evolution of Values*"*. Journal of Marriage and Family 63 (May 2001): 491–503.

[10]   Sucanyaa J Iyer & Deepak Gupta & Sangeetha Gun asekar, "Online Search Tool: Case of Indian Matrimony Portals". International Conference on Communication and Signal Processing April 6-8, 2016, India, 978-1-5090-0396-9/16/$31.00 ©2016 IEEE Received: 29 June 2017 / Accepted: 5 June 2019 /published online:03 July 2019.

[11]   Natalie Nitsche, "Preferences, Partners, And Parenthood: Linking Early Fertility Desires, Marriage Timing, And Achieved Fertility", Business Sciences International Research Journal, Jan. 2020

[12]   Jiehua chen, "Stable Marriage With Multi-Modal Preferences", Acm Ec'18, June 18–22, 2018, Ithaca, Ny, Usa , Acm Isbn 978-1-4503-5829-3/18/06. .

[13]   Patrick mcdonald, "The Male Marriage Premium: Selection, Productivity, Or Employer Preferences", Journal of Marriage and Family (2020).

[14]   Pal, J. K. "Review On Matrimonial Information Systems And Servicesâan Indian Perspective". International Research Journal of Library, Information and Archival Studies, 1(4), 126-135 (2011).

[15]   Sharda, B. D. "Marriage Markets And Matrimonial: Match Making Among Asian Indians Of The United States". International Journal of Sociology of the Family, 21-29 (1990).

[16]   Seth, N. "Online Matrimonial Sites And The Transformation Of Arranged Marriage In India. In Virtual Communities: Concepts", Methodologies, Tools and Applications (pp. 951-974). IGI Global (2011).

A Review on inventory Management in small scale businesses

Dr. Ramah Sivakumar, G. Rajalaxmi
Department of Computer Science, Bishop Heber College, Tiruchirappalli-620 017.
ramahsivakuamr.cs@bhc.edu.in

Abstract

Inventory management is a finance and production functional sector that deals with the efficient and effective utilisation of raw materials and spare parts used in the manufacture of finished goods in a manufacturing company. Inventory management assures availability of sufficient quantities of commodities at the right time, as well as reduces unnecessary inventory investments. Inventory management encompasses numerous areas of the supply chain which includes regulating and overseeing purchases from suppliers and customers, stock storage, controlling the amount of items for sale, and order fulfilments. A good inventory management and control are two critical mechanisms in a manufacturing organisation. Improper Inventory Management leads to customer dissatisfaction and also slower sales. The decrease of 'excessive' stock has a beneficial effect on the profitability of the business. This study focuses on the management techniques in inventory, which is critical to the organisation. Different Inventory control approaches used in various research papers are reviewed in this paper, which gives an insight of inventory management approaches.

Keywords: Inventory Management, stock control, small scale business.

Introduction

An organisation's main goal is to ensure sufficient items or materials on hand to meet the demand without creating surplus or excess inventory. Inventory management is the process of providing an organisation with the items and services needed to achieve its purchasing, stockpiling, and development. This comprises the storage and processing of raw materials, components, and completed products, as well as the administration of final products. A series of actions has to be adopted to reduce dead stock and material usage. Every business, large or small, must maintain track of inventories which include managing inventories from raw materials to finished goods, avoidance of both gluts and shortages. Hence businesses of any size needs inventory management. Depending on the industry appropriate inventory management strategy varies. Effective inventory management allows a firm to generate more sales, which has a direct impact on the company's performance. The risks of inventory gluts and shortages are especially difficult to balance for companies with complex supply chains and manufacturing procedures. Several methods have been developed by the companies for balancing these types of risks. Inventory must be maintained properly for a variety of reasons which includes smooth business operations, estimation of demand, avoidance of stock-out situation, avoidance of price hikes in the near future and a seamless distribution cycle. As a result, it necessitates systematic inventory management, which is overseen by a team of professionals in the field.

Back ground and objectives

Inventory management software should assist the users in reducing costs, improving cash flow, and increasing profitability. It keeps a constant watch on supply and helps to predict demand forecasts. It also

enhances easy inventory analysis from any device. In retail it should be accessible at the point of sale. It organizes the warehouse and save time for the employees. It provides rapid and easy bar code scanning to speed up intake and also allows managing inventories and tracking across numerous locations or warehouses. Manufacturing shortages could be limited by avoiding all unnecessary costs, excess stock and raw resources. Here are some inventory management approaches used by small businesses:

1. *Forecast Fine tuning:*

The precision of forecasting is crucial. Market trends, predicted growth and the economy, promotions, marketing efforts, and so forth should be considered into the projected sales estimates (if the company sell using square, online dashboard for this information should be checked).

2. *First In First Out (FIFO):*

Items should be sold in the order in which they were purchased or made. This is especially true for perishable items such as food, flowers, and cosmetics. For example, a bar owner must be aware of the items behind the bar and use FIFO practises to increase bar inventory. It is also a good idea for non-perishable goods, as items left out too long may become damaged or out of date, rendering them unsellable. In a storeroom or warehouse, the easiest approach to implement FIFO is to add new goods from the back so that the older products are at the front.

3. *Identification of low-turn stock:*

It's probably time to get rid of inventory that has not sold in the last six to twelve months. As excess inventory wastes both space and money, users might consider other options to get rid of it, such as a special price or promotion.

4. *Stock Audit:*

Stock Audit includes year-end physical inventory checking that counts every single item and frequent spot-checking, which is especially useful for goods that move rapidly or have stocking issues. Usage of cloud-based inventory management soft wares lets the company Stock Audit in real time.

5. *Stock level Tracking:*

Tracking the stock level could be effectively done by appropriate software usage like square that makes sure that the company have a robust system in place to keep track of current stock levels, starting with the most expensive things to cheaper ones. It also helps to save time and money by performing most of the hard lifting for business.

6. *Reduction of equipment repair time:*

The crucial gear is not always in working order, however keeping track of assets is critical. A malfunctioning piece of an equipment might increase the lead time. Monitoring the machinery and its parts is critical for gaining a better knowledge of its life cycle and anticipating issues before they arise.

7. *Quality control:*

The practice of inventory control should be designed to ensure that quality standards are achieved by measuring the quality and quantity of production output. An inventory management system ensures that inventory levels are monitored at every stage to ensure that proper quality standards are maintained. It might be as easy as doing a cursory examination during stock audits, replete with a checklist for damage symptoms

and accurate product labelling.

8. *Hiring Stock controller:*

The quantity of inventory that is on hand at any one moment is referred to as stock control, and it encompasses everything from raw materials to completed items. If company have a lot of inventory, then might need more than one person to manage it. A stock controller is in charge of handling all purchase orders, receiving deliveries, and making sure that everything arrives in order.

9. *ABC inquiry:*

ABC investigation is a stock order technique which comprises in partitioning things into three classes, A, B and C. Here, A being the most profitable things, B being the minimum significant ones and C being the paltry numerous. This strategy plans to draw directors' consideration on the basic few (A-things) and not on the insignificant C-things.

10. *Drop shipping Consideration:*

If any organisation employs drop shipping techniques, one may sell items without actually storing the inventory. A wholesaler or manufacturer is responsible for carrying inventory and shipping things when a consumer purchases from the shop. This way, company won't have to worry about inventory keeping, storage, or fulfilment. Drop shipping is a supply chain fulfilment approach employed by many online store owners, but it may also be used by businesses in a range of sectors.

Review of Literature

Dr.Nagendra sohani et.al.,[1] ensures adequate supply of products to customer, avoiding shortages as for as possible and makes sure of the financial investment in inventories to minimum by adopting a strategy. The study occurred at an associate automobile company that was facing two forms of issues, stock out and over stock often times during this analysis. A crucial part of the strategy is assuring timely action for renewal of short term & long run materials designing. The challenge in managing inventory is to balance the provision of inventory with demand. Poor inventory management interprets directly into strains on the company's income. It is thought about that it results from short management over inventory. The method adopted in this study is by taking three months data of the company with the help of product quantity analysis. First the machining time of each and every product and the cumulative percentage of the product were calculated. The results gave an insight on the quantity of production per month and year. Suggestions were given which include interaction of higher level staff with lower level employees to share their problems and to coordinate with inter departments like design department and production department. During the analysis for inventory and supply chain management, it was suggested for the higher management to run programmes for vendor development.

Varalakshmi G S et.al.,[2] had examined sales control, inventory management, and corporate irregularity management. The focus is on the ability to create new sales, update products and sales information, view existing ones and to speed up operations by documenting and automating manual tasks. According to this paper, inventory management system software is a critical and important tool for all firms that deal with inventory. It manages inventory flow in and out, keeps track of inventory levels for all items and stock, provides sales data and analytics, and helps organisations to define unique safety stock requirements. The primary method employed in this study was shopkeeper interviews. This research is based

on both direct data from store owners and secondary data from journals, books, articles, annual reports, and websites. The methodologies used in this study were safety stock inventory, FIFOand LIFO.

Manjusha Talmale et.al.,[3] had reviewed and redefined the automation's requirements. The proprietor of this organization can discover every one of the suitable elements in one view page design. All the stock section ought to be keeping up in that view, so it can be refreshed in each time as saucy the necessity. It can be changed over into a PDF design and produced the invoice for the client. The objective is to determine the best level of inventory to order at any given time for a certain construction project, as well as how much can be ordered based on the size of the structure or the cost of such a project in the post-construction phase. Research strategy can be characterized as efficient and purposive examination of actualities with a goal deciding the powerful relationship between such certainties and research. Poll study was directed among development experts to distinguish their feeling towards stock administration framework in their association. The obtained data is examined to determine the frequency of reaction for various elements. This paper concludes as an alarm about the information section in the bill which in view of desktop application. It is a secure application in which there was no information spillage from the stockroom. And furthermore gives the one table organization so that comprehensive sales details could be obtained at every month end.

Anjali Mishra et.al., [4] had examined the inventory management process using unstructured interviews, on-site study, and annual report analysis. The basic objective was to study the inventory management techniques used in Linamar India Pvt. Ltd. The study was based on the primary data collected by the finance executive of the Company and secondary data which was collected from the books, journals articles and the annual reports of the company and websites. The techniques used in the study were ABC Analysis, EOQ, Inventory turnover ratios and Safety Stock. EOQ is the inventory management tool for determining optimum order quantity which is the one that reduces the total of its carrying costs and order. Safety stock is the reduced surplus inventory, which serves as a safety margin to meet an unexpected increase in usage resulting from a strangely high demand and an unmanageable late receipt of incoming inventory. Inventory turnover ratio is also called Stock Turnover Ratio. It is the calculated inventory frequency that is converted into the sales. In simple terms, it measures capacity of the firm to generate revenue. It is calculated by dividing net sales by average inventory,

Inventory Turnover Ratio = Net Sales/ Avg. Inventory

This paper concludes that Linamar India Private Limited is managing its inventory very efficiently. EOQ, safety stock analysis, ABC analysis are being undertaken efficiently and effectively. Inventory turnover ratio is also showing an increasing trend which indicates that sales of the organization is increasing every year. The used techniques helped it in continuous flow of its production activities.

Pratap Chandrakumar.R et.al.,[5] had investigated the inventory control in WABCO INDIA which is a major brake manufacturing company. The research demonstrates the inventory analysis of ABC items, SAP, and stock policy implementation. The company lacks appropriate demand forecasting, which has an impact on the manufacturing process. Three kinds of stock such as raw materials, work-in-progress, and completed items are used. ABC inquiry method is used which divides items into three categories: A, B, and C. A represents the most profitable items, B represents the least significant, and C represents the smallest number. A detailed analysis of the company's inventory management is included in this assessment. Finally it was concluded that proper inventory management is effective to company's control and expansion, and

improper inventory management can impair the company's operations.

Surbhi Mishra et.al., [6] aims to examine a supply store of electrical and non-electrical products by evaluating its more frequently demanded and less frequently consumed products on the basis of their consumption and frequency of use. The objective of the study is to apply the XYZ - FSN inventory management technique to such products. The term 'safety stock' is used by logisticians to describe a level of extra stock that is maintained to mitigate risk of stock outs due to uncertainties in supply and demand. The researchers used both primary and secondary data to complete this study as required for both qualitative and quantitative methods has been used. Data was obtained from a structured interview as well as primary sources such as annual report, journal, article and textbook. Researchers used ratio analysis which is the information obtained from annual report to work out the relationship between inventory and quantity. Inventory was revealed to have a lot of imbalance in maintaining its level, therefore it was concluded. Some items were often requested, while others were requested relatively seldom, and yet others were not moving from or into inventory at all. Because they are the only products on which a large expenditure is made, the X category items were identified to be the most important from a management aspect and should be closely watched. The overall number of such significantly invested objects made about 15% of all components. Y class items account for 20% of total components in terms of investment, and as they are less critical, they require less standard controls and periodic usage reviews. Approximately 65% of the total is made up of Z class products.

Prayag Gokhale [7] ensures the work divided into sections in which inventory management ideas are presented. It also discusses the various costs associated with inventory storage and management. The research is based on secondary data and the statistical tool used was MS Excel. Pig Iron's EOQ was 6.24 tonnes in 2014-15 and climbed to 6.93 tonnes in 2015-16. In 2016-17, it grew upto 8.98 tonnes due to lower carrying costs, before dropping to 7.49 tonnes in 2017-18. By effectively implementing the EOZ method for acquiring pig iron, the corporation may save a large amount of money. The results depict a trial-and-error approach in calculating the EOQ of pig iron. In overall, the study shows how inventory management affects the financial statements and operational activities of the company.

SS Islam et.al.,[8] focus on finished goods that are one of the most important things for trading companies to meet customer demands and thus to increase customer satisfaction. One of the important activities in production and procurement is inventory management. The research aims to evaluate factors that affect inventory mismanagement in small medium enterprises. One of methods used to analyse the factors that affect inventory management inefficiency is fishbone diagram. This study finds the PT ABCD Company's efficient inventory level by using Pareto diagram and also finds three material categories (amount, quantity, space) which significantly affect the efficiency of spare part management. Inventory turnover is calculated using   cost of goods sold from stock sales and the   average inventory management during the past 12 months. The quantitative method of analysis is applied to evaluate the company's inventory management as well as its human resources and information systems. Findings suggest that inefficient inventory management can cause delays in production, unsatisfied customers, or working capital curtailment. This research enriches industry by suggesting the importance of information systems and human resources for inventory management. Based on the findings, this study concludes that the inventory cost efficiency is the uninterrupted company's information system and lack of qualified human resources.

Nwafor Chidinma Anulika et.al.,[9] ensures a mechanized stock administration framework is a modernized strategy for controlling deals and item in an organization. This application amplifies proficiency, decrease work and stay away from over expense. With the goal for organizations to have significant and satisfactory data for appropriate administration and dynamism, it is a basic practice for organizations to put resources into stock administration frameworks. These frameworks perform numerous exercises among which is handling information that include exercises, for example, figuring, look ups, arranging, ordering, and summing up information. These exercises compose examination, information control, and collecting data. The planned framework was tried, assessed and prescribed to the administration of the organization. This research work centers on the utilization of Computer (PC) framework regarding automated stock control arrangement of Michael Okpara University of Agriculture Umudike (MOUAU) market. The work was effectively evolved utilizing python flagon structure, SQLite Database and client experience. The establishment of the product would improve general business measures at the organization. The research was focussed on developing a working prototype of an automated inventory management system that would perform business functions such as registration of products, customers, and system users; generating various reports; and maximizing sales through ensuring product availability.

 Zhang Yi et.al.,[10] had analysed the weights of factors, that limited the capability and supply of resources as well as the shortage of experience in e-commerce industry. In order to explore the new inventory management problems of small and medium-sized retailers in the context of New Retail, this paper takes ASY company as an example, and analyses the weights of factors affecting ASY company's inventory management before and after New Retail appears by using the Analytical Hierarchy Process(AHP) method. AHP decomposes the correlation factors associated with the decision-making process into sub factors at different hierarchies, analyses their relevant degree of influence, and then conducts qualitative analysis and quantitative coupling analysis with mathematical tools. AHP method is applicable to problems that have a target system for hierarchical interlaced evaluation indicators and target values which are difficult to make quantitative description. For the factors affecting inventory management of ASY Company mentioned, their indicators are hard to quantify or make quantitative descriptions. Paper studies the effect of new Retail model on the inventory management of small and medium-sized retailers. Influencing factors include management system, personnel capabilities, delivery method, sales model, product categories, and resource limitation.

Discussion

It is inferred from the reviewed papers that three techniques such as "ABC inquiry", First in First Out, and EOQ model are considered significant. "ABC inquiry" is a stock order technique that divides items into three categories: A, B, and C, with A being the most profitable and C being the least significant. This method aims to focus directors' attention on the essential few (A-things) rather than the insignificant number (C-things). First In, First Out (FIFO) is an asset management and valuation strategy in which the first assets created or purchased are sold, utilised, or disposed of. For tax reasons, FIFO assumes that the assets with the oldest expenses are included in the cost of goods sold in the income statement (COGS). This approach enables users to locate the stocks more quickly and utilise it more effectively and The EOQ model assists a corporation in balancing storage and order costs by determining the quantity that should be used to refill inventory in order to reduce both order and holding costs. The method for calculating Economic Order Quantity (EOQ) includes yearly demand, holding cost, and order cost. This method seeks to strike a balance

between how much business sell and how much they spend on inventory management. [4] On behalf of using "ABC inquiry" technique, the organization turns over its inventory ratio from 1.27 to 1.80 and further to 3.27. FIFO gives accuracy in inventory, improve customer service, prevent stock issues and increase profit. The EOQ model aided the company by lowering order costs and calculating the ordering quantity for a certain inventory item.

Conclusion

This review paper mainly focuses on the techniques of inventory management, which helps in avoiding mismatch of stock. Some methodologies such as ABC inquiry, FIFO, EOQ model, XYZ analysis, and AHP method, calculating Machine Time, Fishbone diagram and EOZ method were discussed. Each and every method has its own advantages and disadvantages in maintaining the inventory. From the study it is found out that, apart from above said techniques, researches should be focussed on to propose more efficient and accurate methods in maintaining inventory for the small scale businesses.

References

[1] Dr.Nagendra Sohani, Ajay Pagare, "Study And Analysis Of Inventory Management Practices In Small Scale Industry", International Journal Of Advanced Research In Computer And Communication Engineering Vol. 10, Issue 6, June 2021,DOI:10.17148/IJARCCE.2021.10689, Copyright To IJARCCE- International License ISSN (Online) 2278-1021, ISSN (Print) 2319-5940.

[2] Varalakshmi G S1, Asst Prof. Shivaleela S2, "A Review Of Inventory Management System", International Journal of Advanced Research in Computer and Communication Engineering Vol. 10, Issue 6, June 2021 DOI 10.17148/IJARCCE.2021.10689

[3] Punam Khobragade, Roshni Selokar, Rina Maraskolhe , Prof.Manjusha ,Talmale," Research paper on Inventory management system", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 ,Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072

[4] Anajali Mishra & Harshal Anil Salunkhe, "A Study of Inventory Management System of Linamar India Pvt. Ltd, Pune" ,Amity Journal of Operations Management vol3 (1), (35-41) ©2018 ADMAA

[5] Pratap Chandrakumar. R,Gomathi Shankar , "A Study On Inventory Management And Control", Vol-3 Issue-5 2017 Ijariie-Issn(O)-2395-4396

[6] Surbhi Mishra , Sourabh Tege , Vishnu Agarwal, "Analysis of Inventory Management Performance– A Case Study" ,International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV6IS030032(This work is licensed under a Creative Commons Attribution 4.0 International License.)Published by :www.ijert.org Vol. 6 Issue 03, March-2017

[7] Prof. Prayag P. Gokhale, Megha B Kaloji, "A Study on Inventory Management and Its Impact on Profitability in Foundry Industry at Belagavi, Karnataka", International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS) Volume VII, Issue IX, September 2018 | ISSN 2278-2540

[8] S S Islam, A H Pulungan, A Rochim, "Inventory management efficiency analysis: A case study of an SME company", 4th Annual Applied Science and Engineering Conference Journal of Physics: Conference Series 1402 (2019) 022040 IOP Publishing doi:10.1088/1742 6596/1402/2/022040

[9] Nwafor Chidinma Anulika, Idoko Nnmadi A, Agbo Jonathan Chukwuwike, Ogbene Nnaemeka Emeka "Design and Optimization of An InventoryManagement System for Central Stores", International Journal of Interdisciplinary Innovative Research &Development (IJIIRD) ISSN: 2456-236XVol. 05 Issue 01 | 2020

[10] Zhou Chuanpeng, Zhang Yi, "Study on Inventory Management Optimization of Small and Medium-sized Retail Companies in the Context of New Retail: Based on AHP", Asian Journal of Business and Management (ISSN: 2321 - 2802)   Volume 9 – Issue 5, December 2021

[11] Yu Ping. Discussion on the Problems and Countermeasures in Inventory Management of SMEs [J/OL]. China Journal of Commerce, 2019(08):96-97[2019-05-20]. https://doi.org/10.19699/j.cnki.issn2096-0298.2019.08.096.

[12] Aditya A. Pande, S. Sabihuddin, "Study of Material Management Techniques on Construction Project", International Journal of Informative & Futuristic Research, e-ISSN: 2347-1697, Vol.2 (3), May 2015, ISSN-p: 3479-3486.

[13] Lancioni R A and Howard K 1978 Inventory Management Techniques International Journal of Physical Distribution & Materials Management **8**(8) 385-428

[14] P.G. Matsebatlela, K. M pofu, "Inventory Management Framework to Minimize Supply and Demand Mismatch on a Manufacturing Organization", International Federation of Automatic Control, Vol.3, No.48, Mar 2015, pp-260- 265.

[15] Jose, T., Jayakumar, A., & Sijo, M.T.. (2013). Analysis of Inventory Control Techniques- A Comparative Study. International Journal of Scientific and Research Publications, 3(3), 520–530

[16] Dr. Rakesh Kumar, "Economic order quantity model",   Global journal of finance and economic management, Vol.5 (2016).

[17] Sachin Agarwal, "Economic Order Quantity Model", International Journal of Mechanical, Civil, Automobile and Production Engineering, Vol.5, Issue 12(2014).

[18] S.P. Desselle, and D.P. Zgarrick, Purchasing and Inventory Management, Pharmacy Management: Essentials for All Practice Settings 2nd ed., New York: McGraw-Hill Co., Inc, 2009, p. 383,

[19] Ebrahimpour, M and schoberger, R.J. (1984), "The Japnese just in time/ total quality control production system: potential for developing countries" , International journal of production research, vol. 22 No. 3 pp.421-30.

[20] Golhar , D.Y. stamm, C.L. and smith, W.P. (1990), "JIT implementation in manufacturing firms", Production and inventory. Management Journal, vol. 81 No. 2, pp. 44-48.

Development of Image Encryption Algorithm

Suruthi Vanthana R, Eliahim Jeevaraj P S
Department of Computer Science,
Bishop Heber College (Autonomous), Tiruchirappalli.
eliahimps.cs@bhc.edu.in

Abstract

Digital images have become increasingly significant in modern life since they are simple to comprehend and provide realistic descriptions. Because many digital photos contain secret or private information, it is important to explore how to secure them. The challenges of guaranteeing individuals' privacy are becoming increasingly tough. To secure data and personal privacy, several solutions have been studied and created. Encryption is, without a doubt, the most evident. Image encryption is critical for safeguarding sensitive information from prying eyes. Encryption entails using special mathematical algorithms and keys to convert digital data into cypher code before transmission, and decryption entails using mathematical algorithms and keys to recover the original data from cypher code.

Keywords: Image Encryption, Public Key, AES

1. Introduction

Digital images, and digital video based digital images, have become an important medium for information storage and transmission in the computer network in the civil and military fields in recent times, along with the rapid creation and popularization of network technology and digital communication technology around the world. Nonetheless, network security challenges have long been a major stumbling block to the advancement of network technology. The implementation of data security protection in the computer network is an essential content and direction of investigation in the field of network security and information security, especially in the setting of public and government information coffers. Digital picture and digital video have emerged as key data transfer material among them.

The objectives of this paper are to develop the system to secure the images files, to provide the algorithm for securing image data from unauthorized access and encrypt the image files using AES algorithm.

 Section 1 includes about introduction. Section 2 specifies the literature survey work. Section 3 provides the details about algorithms. Section 4 includes results and discussion. Finally the conclusion is given in Section 5.

2. Related Works

P. Radhadevi, et.al., [1] proposed digital images account for 70% of the information transmission on the internet. Illegal data or image access has become easier and more common in wireless and general communication networks. For safe image transmission over channels, an encryption system has been developed to protect valuable data from undesirable readers.

P.Karthigaikumar et.al., [2] developed substantially two types of cryptographic algorithms symmetric and asymmetric algorithms. NIST has blazoned the Rijndael Algorithm is the stylish in security, performance, effectiveness.

Roshni Padate, et.al., [3] proposed an AES algorithm for encryption and decryption. The National Institute of Standards and Technology (NIST) has initiated a process to develop a Federal Processing Standard (FIPS) for the Advanced Encryption Standard (AES).

Swarali Naik, et.al., [4] developed the data is encrypted at the sender's end, and the data is received in encrypted form by the receiver. The secret key is used at the receiver end, which is shared by both the receiver and the sender.

Priyanka Sharma et.al., [5] proposed an images are inescapable. Image attributes such as redundancy, bulk, and high transmission rate with limited bandwidth are important. Maintaining a high level of security has always been a concern. This paper offers a modified AES Key Expansion is a bitwise encryption method.

Ahmed Bashir Abugharsa, et.al., [6] developed a new picture encryption technique based on the integration of shifted image blocks and basic AES is proposed. Shifted image blocks are made up of a certain amount of pixels, which are shuffled using a shift technique that shifts the original image's rows and columns to produce a shifted image.

Roshni Padate, et.al., [7] proposed that National Institute of Standards and Technology (NIST) has begun work on a Federal Information Processing Standard for the Advanced Encryption Standard (AES). The AES algorithm for encryption and decryption is used in this paper to develop an effective security system for data transfer.

Sneha Ghoradkar, et.al., [8] proposed an AES-based image encryption and decryption algorithm. Design is iterative with a block size of 128 bits and a key size of 256 bits. Image is used as an input to AES Encryption to produce the encrypted image, and encrypted image is used to produce original image.

G. Venkatesha, et.al., [9] proposed it is critical to safeguard confidential picture data from unwanted access. This paper proposes an AES-based image encryption and decryption algorithm. With a block size of 128 bits and a key size of 256 bits, the design is iterative. The complexity and security of cryptography algorithms increase as the secret key grows.

Kaushal Kumar, et.al., [10] developed an image encryption is a method of securing photographs by replacing the initial image with a decryption key. It takes care of things like encryption, validation, and key management. The investigation of cryptography techniques for secure correspondence in the presence of an adversary.

3. Proposed Algorithm

The AES algorithm is also known as the block cipher algorithm. No successful attack has been reported on AES algorithm. Some advantages of AES are easy to apply on 8- bit armature processors and effective

perpetration on 32- bit armature processors. In addition, all operations are simple (e.g, XOR, permutation and negotiation). The encryption is performed in multiple rounds. Each round has four main way including sub-byte, shift row, mix column and add round key. Sub-byte is the negotiation of bytes from a look-up table. Shift row is the stirring of rows per byte length. Mix column is addition over Galois field matrix. Eventually, in the add round crucial step, the affair matrix of blend column is XOR with the round key. The number of rounds used for encryption depends on the crucial size. For a 128- bit crucial, these four way are applied to 9 rounds, where the 10th round doesn't consider the blend column step. Since all way are recursive, decryption is the reverse of encryption.

3.1. Substitute bytes

Sub Bytes transformation, also known as substitute Byte, is a nonlinear byte substitution and a basic table lookup approach that is one of the transformation techniques.

3.2. Shift rows

It moves the bytes in each row by a specific offset in a cyclic manner. The first row is unaffected for AES. The second row's bytes are shifted one to the left. The third and fourth rows are also displaced by two and three offsets, respectively.

3.3. Mix Columns

Each column of the state is multiplied by the known matrix in the Mix column operation. It's a method that accepts 32 bits of data and produces 32 bits of data. We can significantly lower the cost and power consumption of the hardware that implements AES by optimizing its architecture.

3.4. Add round key

A Round Key is added to the State using an XOR operation in the Cipher and Inverse Cipher transformations. The size of the State is equal to the length of a Round Key.

3.5. Image Encryption Algorithm using AES
Phase 1: Encryption

INPUT: Image (I)

OUTPUT: Encrypted Image (Ie)

    Step 1.   Read an Image
    Step 2.   Initialize the Rounds for Encryption
In Each Round (except last round), add Round Key

    Step 3.   Apply the Substitution Bytes
    Step 4.   Apply The Shift rows & Mix Columns Operations
    Step 5.   In Last Round Apply the substitute byte, shift row & Add round key
    Step 6.   Encrypted Image is generated

Phase 2: Decryption

--- --

INPUT: Image (J)

OUTPUT: Decrypted Image (Je)

--- --

  Step 1. Read an Encrypted Image
  Step 2. Initialize the Rounds for Decryption
  Step 3.   Apply Add round Key
  Step 4.   Apply the Inverse Shift rows and Inverse sub bytes
  Step 5. add Add round key
  Step 6. In last round Inverse mix columns, Inverse shift rows, Inverse sub byte
  Step 7. Decrypted Image is generated

-- --
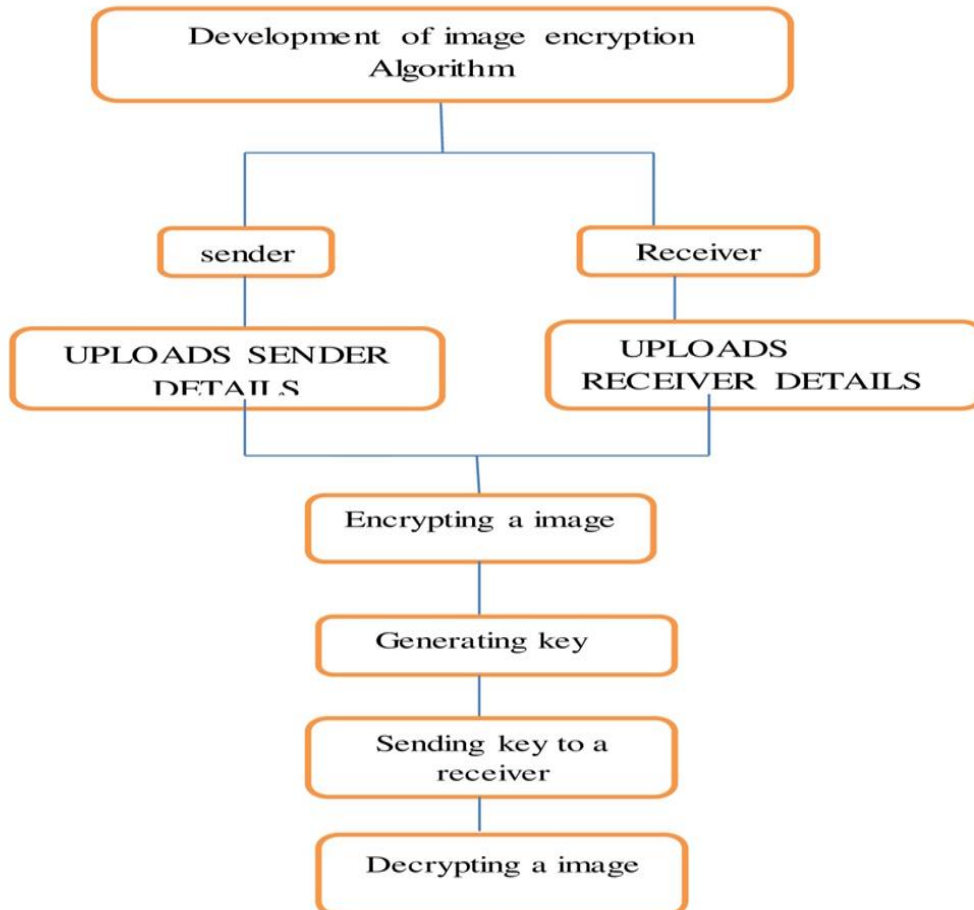

Fig 3.1.Methodology

4.   Results and Discussion

The algorithm has been enforced in JAVA in web terrain with a system configuration of Intel Core i3 processor with 2 GB RAM. The proposed algorithm has been tested with colorful images. A good encryption method should be resistant to cryptanalytic, statistical, and brute-force attacks of all kinds. It's also advantageous if the encrypted image has little or no statistical similarity to the plain image to aid information leaking to an adversary.

(a)Original Tulip image	(b) Encrypted Tulip image	(c) Decrypyted Tulip image
(d)Original Train image	(e)Encrypted Train image	(f)Decrypted Train image
(g)Original Air image	(h)Encrypted Air image	(i)Decrypted Air image
(j)Original Jelly Fish image	(k)Encrypted Jelly Fish image	(l)Decrypted Jelly Fish image

(m)Original Desert image   (n)Encrypted Desert image   (o)Decrypted Desert image

(p)Original Penguin image   (q)Encrypted Penguin image   (r)Decrypted Penguin image

(s)Original Koala image   (t)Encrypted Koala image   (u)Decrypted Koala image

(v)Original   Light   house   (w)Encrypted Light house image   (x)Decrypted Light house image
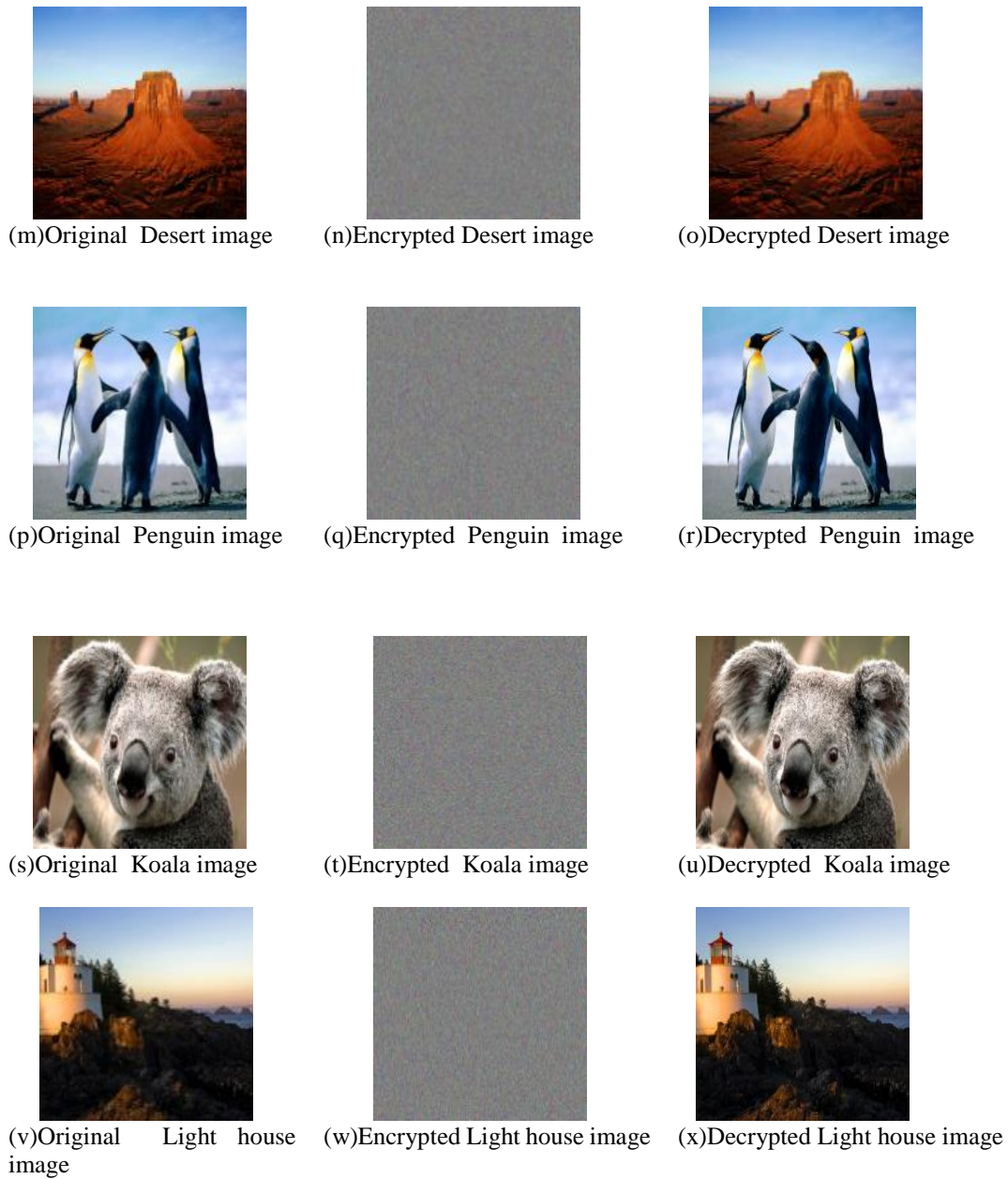image

Fig 4.1. Result Images of Proposed Algorithm

Table 4.1. Entropy values of the Proposed Algorithm

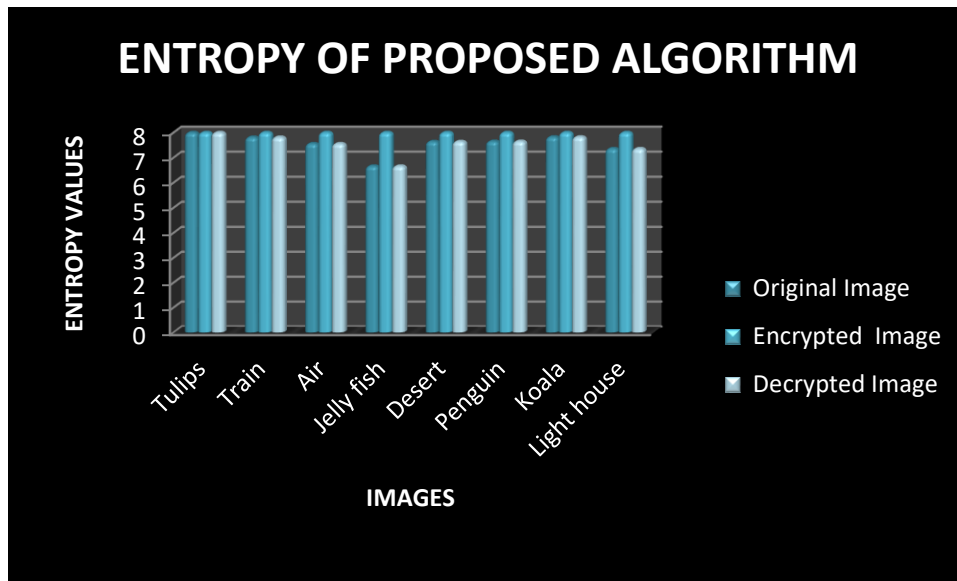| Image	Original image	Encrypted image	Decrypted image
Tulips	7.9999	7.9999	7.9997
Train	7.8139	7.9999	7.8139
Air	7.5560	7.9999	7.5560
Jelly fish	6.6624	7.9997	6.6623
Desert	7.6419	7.9999	7.6419
Penguin	7.6517	7.9999	7.6515
Koala	7.8254	7.9999	7.8253
Light house	7.3582	7.9999	7.3580

Fig. 4.2. Entropy values of Proposed algorithm

5. Conclusions

The basic goal of picture encryption is to send an image securely over a network so that no unauthorized may decode it. Image encryption is used in a variety of industries, including banking, telecommunications, and medical image processing, among others. Because of the rise of the Internet, where moving and entering data across computers requires some level of security, encryption has become increasingly crucial. In this topic, colorful algorithms for ciphering and deciphering images have been presented. The proposed cryptographic methodology has been tested on several types of input photos with changes in image size and AES encryption algorithm keys. The entire time, the secret image is captured in high resolution. The proposed algorithm is assessed by entropy values of the images.

References

[1] P. Radhadevi, P. Kalpana. "Secure Image Encryption using AES". ISSN: 2319 - 1163

[2] P. Karthigaikumar, Soumiya Rasheed. "Simulation of Image Encryption using AES Algorithm". IJCA Special Issue on Computational Science - New Dimensions & Perspectives. NCCSE, 2011

[3] Roshni Padate, Aamna Patel. "Image Encryption and Decryption using AES Algorithm" ISSN 0976 – 6464(Print) ISSN 0976 – 6472(Online)

[4] Swarali Naik, Shrinidhi Patil, Ashna Maurya, Sangeetha Selvan. "Image Encryption Using AES Algorithm" .ISSN :2583-1062 Impact Factor :2.205

[5] Priyanka Sharma. "A New Image Encryption using Modified AES Algorithm and its Comparision with AES" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181

[6] Ahmed Bashir Abugharsa, Abd Samad Bin Hasan Basari, Hamida Almangush. "A New Image Encryption Approach using The Integration of A Shifting Technique and The Aes Algorithm " .International Journal of Computer Applications (0975 – 8887) Volume 42– No.9, March 2012

[7] Roshni Padate, Aamna Patel. "Encryption And Decryption Text Using AES Algorithm" (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 5, May 2014)

[8] Sneha Ghoradkar, Aparna Shinde. "Review on Image Encryption and Decryption using AES Algorithm". International Journal of Computer Applications (0975 – 8887)National Conference on Emerging Trends in Advanced Communication Technologies (NCETACT-2015)

[9] Prof. GVenkatesha "AES Based Algorithm for Image Encryption and Decryption ". Perspectives in Communication, Embedded-Systems and Signal-Processing (PiCES) – An International Journal ISSN: 2566-932X, Vol. 2, Issue 11, February 2019

[10] Kaushal Kumar, Er. Jasdeep Singh Mann. "A Novel Approach in Image Encryption Using AES Algorithm". International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 6 Issue: 5 187 – 192.

Video Summarization Algorithm for Cricket Matches

Arun Prakash V S, Eliahim Jeevaraj P S
Department of Computer Science,
Bishop Heber College (Autonomous), Tiruchirappalli.
eliahimps.cs@bhc.edu.in

Abstract

Video Summarization is a technique that will create a short summary of the content of a large video file by extracting the important information and generate the summary with most informative or interesting part for the users. Nowadays there are large amount of video content are present on the Web, and the number of videos is also increasing in each second and it is uploaded on the social media or on the internet. Nowadays most of the audience did not have much time to watch the cricket matches as they are very long-lasting. For that reason, audience are giving preference to watch the summarized video or the highlights of the match. At present the highlights of the match are done manually. Therefore, in this paper we propose a system that the match summary video is generated automatically with the help of the OpenCV algorithms. The summarized video of that match will have interesting and important parts that are present in the complete cricket video. The quality of the produced summary highly based on users' understandings.

Keywords: Video summarization, Otsu thresholding, Contour based thresholding

1. Introduction

A video summarization is a technique that will generate a short summary of a large content video file by extracting the important key frames or useful information from it and present it to the user. The output summary is usually a collection of a set of video frames or video clips that is extracted from the original video with some editing process.

The main focus of the video summarization is to reduce the duration of the video without degrading the important information of the video content. The produced summary that is joined in sequential order to form a shorter video. By watching the summary video user can get useful information from the video in short period of time.

The objectives of this paper are listed as follows:

- To watch the full video and to extract the features by manually takes more time.

- To recognize the specific frames from the input video.

- To extract frames based on the selected frame.

- To combine the extracted frames and form into output video.

To produce the sports highlights video, it requires best video editing tools in addition it also require more time to edit these types of videos to exact output. Alternatively, video summarization will automatically capture the specific or important frames from the given video based upon the information given to the regarding specific sports. The main focus of the work is to reduce the duration of the video without degrading the important information from the video. This approach is proposed that it will convert the given input video

into short summarized video. Several methodologies have been adopted by different researchers to summarize the video. But the accuracy rate differs based on the features and methodology is used.

The Introduction are given in section 1. Section2 provides detailed review about the proposed method. Section 3 provides methodology used in the system. Result and Discussion in the system are mentioned in section 4. Conclusion was discussed in section 5.

2. Review of Literature

Damnjanovic *et al.* [1] concentrated in an event-based summarization in which intensity each frame in the video calculated by the difference between current frame and the next frame. Then all the difference values are added together. Finally, video summary is produced to the frames using the summarization method.

P. Kalaivani *et al.* [2] proved that event detection algorithm is very much helpful to detect the objects in the video and it is easy to make the summary video with the help of them. Kumar *et al.* [3] used the bootstrap aggregating for improving the frame accuracy for the selected frames.

Thomas *et al.* [4] introduced the Human Visual System (HVS) for the conceptual video summarization which detect the important information frames and eliminate the unnecessary events from video and produce the video summary of the given video.

Cricri *et al.* [5] used magnetometer for the event detection video that captures attractive point in the sports video. This technique captures crystal clear snapshot from the video and performs attachment study process to eliminate the calculation difficulty and false positive.

Andaloussi *et al.* [6] defined a way that using the audio quality, object detection, and the result box will helps in making the summarized video of the given content. Cote *et al.* [7] used the Hidden Markov Model (HMM) to catch the unusual or irregular way of action by the students in the time of writing a test. The result is based on the position of student head and recognize the abnormality.

Feng *et al.* [8] used the same Hidden Markov Model (HMM) that detect the motion of the video that extract the low-level to high-level features based on the object detection after that it will generate a summarized video by choosing the selected frame from the video.

M. Miniakhmetova *et al.* [9] proposed an approach of summarizing certain videos based on objects. It uses detection techniques to organize different scenes from the given video. Then, using a subset of video scenes, a video summary is created. Finally, the video summary is created that only includes shots/scenes with objects.

Rav-Acha *et al.* [10] showed the video summary, which combines all of the video's actions to create a video summary. This goal may be accomplished by recognizing items in motion and then enhancing video performance based on the detected things.

K. Davila *et al.* [11] showed how to choose frames in lecture films based on segmentation by eliminating clashes between removing items, content areas, and fixing up each frame to provide a summary video. K. Grauman *et al.* [12] offered a method for estimating the main of each shot using a sub shot from the original video sequence. Lu *et al.* [13] also discussed how to calculate the relative value of each video clip.

J Varghese *et al.* [14] The Structural Similarity Index (SSI) and the Stroboscopic effect were used to propose a technique of video summarization based on recognizing shot limits and deleting uninteresting frames. Muhammad *et al.* [15] used Convolutional Neural Networks (CNN) to develop a structured video summarizing approach for examining surveillance recordings. Using shot segmentation, this system will learn deep features. Furthermore, using entropy and memorability, this approach generates a video summary. The shot that is used as a keyframe will have a high memorability and entropy score.

Ngo *et al.* [16] described a method based on perceptual quality and content balance of video summarization. As a result, a motion observation model is used to compute the shots, and the clustering technique is adapted to edit/cut the video. The temporal graph is created using both methods for determining the relevance of the cluster. Finally, the scenes from clusters are combined using this temporal graph. Its attention values are utilized to choose relevant sequences from which a video summary is generated.

3. Proposed Algorithm

The video summarization algorithms are developed two different strategies namely mean of the mean of the frames and contour segmentation. All the algorithms are separated as a different frames from the input videos. The statistical feature such as mean and component feature namely contour segmentation are extracted. The frames are clustered based on the mean found and contents segmented. The classified frames are summarized. These algorithms are accessed by the time complexity and the summarized percentage. The first step is to give the input video. Basically, a series of images played in sequence at a specified frame rate, form as video. The given input video is converted into group of images or frames in sequential order and stored it in a folder.

INPUT VIDEO

↓

EXTRACTING IMAGES

↓

FEATURE EXTRACTING
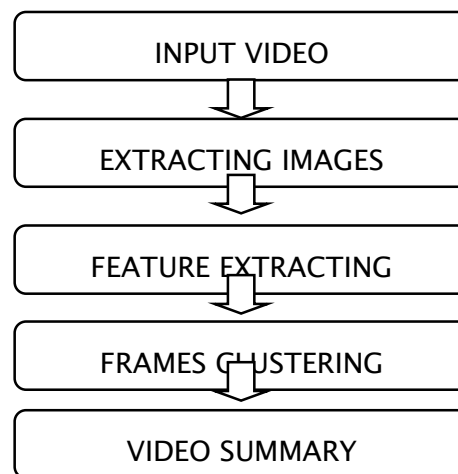
↓

FRAMES CLUSTERING

↓

VIDEO SUMMARY

Figure 3.1 Flowchart of the Project Methodology

3.2 Otsu Thresholding of the Image

In Otsu's method thresholding, uses an arbitrary chosen value as a threshold and it will determine the values automatically. For this process an equation is described as:

$$\sigma_w^2(t) = w1(t)\sigma_1^2(t) + w2(t)\sigma_2^2(t)$$

where w1(t), w2(t) are the probabilities of the two classes divided by a threshold t, which value is within the range from 0 to 255.

3.3 Dilation of the Image

Morphological operation is a process based on shapes in an image. They apply a structuring element to an input image and generate an output image and there are two basic morphological operations are Erosion and Dilation. The pixel element present in the color image is '1' then it will increase the white region in the images.

3.4 Mean based Video Summarization Algorithm

-- --

Input: Original Video.

Output: Frame by frame of the video.

Step 1.   Read video as input and then extract frame by frame of the original video.

Step 2.   Save and convert that color image into gray scale, gaussian blur images.

Step 3.   Apply the Otsu Thresholding and dilation operations to every gray scale images.

Step 4.   Find the mean value of every gray scale images.

Step 5.   Fix the threshold value (T) as average of all frames.

Step 6.   Group the frame-based condition of if ($Mi > $T).

Step 7.   Summarize the grouped frames and the output video is generated.

3.5 Contours Function

A Contour function is specified as a plain curve that joining all the continuous points that having same color or intensity. This contour's function is very useful tool for shape analysis, object detection and recognition. To apply contours, the original color image should be converted into gray scale image format. we should apply threshold or canny edge detection to gray scale images. We can also find contours that are white object that present in the black background. When thresholding is applied to the gray scale images which turns the boundary of the object in the image completely white so that the algorithm can detect the boundary of the image. Once contours have been recognized, we use the drawContours() function to cover the contours in the original color image.

3.6 Contours Segmentation Algorithm

-- --

Input: Original Video.
Output: Summarized video.

-- --

Read video as input and then extract frame by frame of the original video.

> **Step 1.**   Apply the gaussian blur images for enhancing the frames
>
> **Step 2.**   Apply the contour method segmentation with the help of deep neural networks.
>
> **Step 3.**   Based on the labels the frames are grouped.
>
> **Step 4.**   Summarize the grouped frame and generate the output video.

-- --

4. Results and Discussion

Video summarization of Cricket Match is implemented with the OPENCV packages present in Python programming with the system configuration of Dell AMD Ryzen 5 series. This project focus on the video summarization that generate short summary of the large content file. The main purpose is to detect the object and motion of the given input video. The produced video summary is based on the process of contour function and mean of the images.

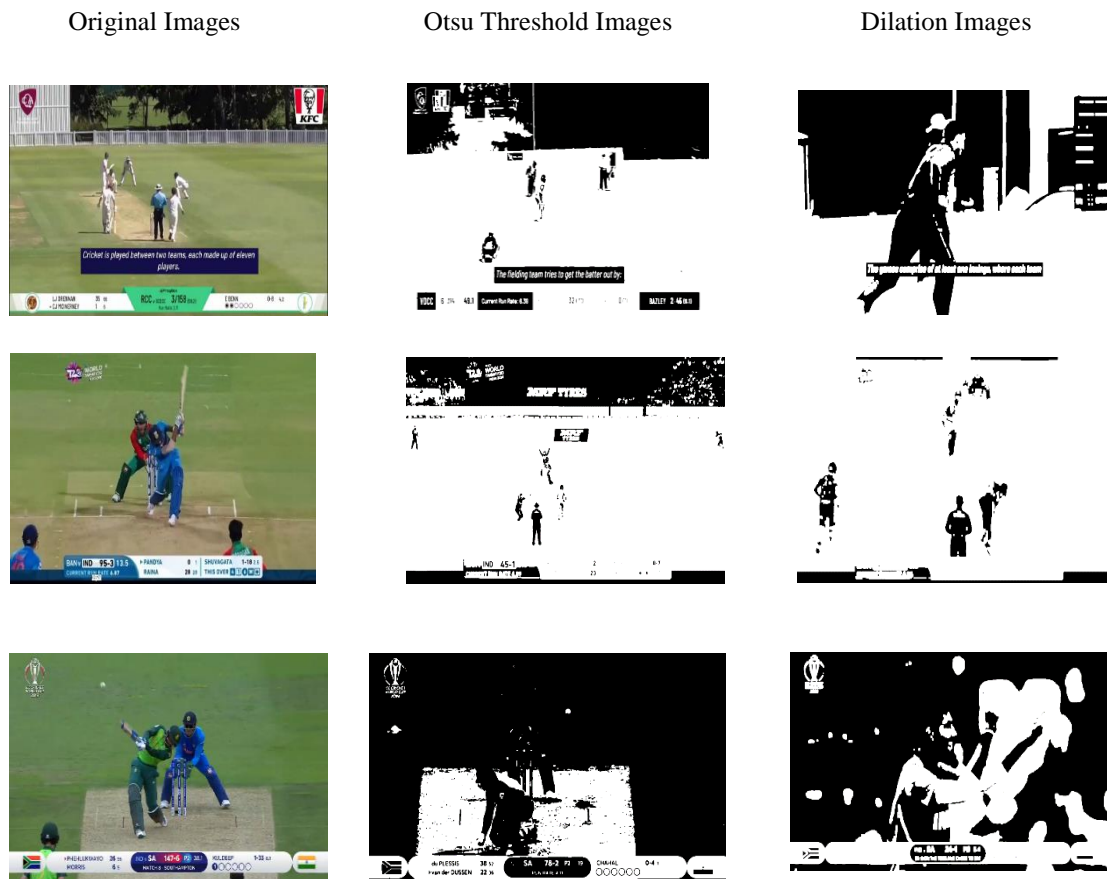| Original Images	Otsu Threshold Images	Dilation Images

Fig. 4.1. Results of Video Summarized based on Mean

Fig 4.1. shows that the processing of frames from the original video are converted into Otsu thresholding images and dilation images.

Original Images   Contour Images   Labeled Images

Fig. 4.2. Video Summarization using Contour Segmentation

Fig. 4.2. shows that the processing of frames from the original video is converted into contour images and labeled images.

Table 4.1. Running Time for Algorithm #1

S. No	Video Duration	Time Taken (In Seconds)
01	1 Minute	282
02	3 Minutes	897
03	5 Minutes	1310
04	10 Minutes	2452

Table 4.1 denotes that time taken to run the program for image segmentation video generation for different video duration.

Table 4.2. Running Time for Algorithm #1

S. No	Video Duration	Time Taken (In Seconds)
01	1 Minute	512
02	3 Minutes	1373
03	5 Minutes	2262
04	10 Minutes	3603
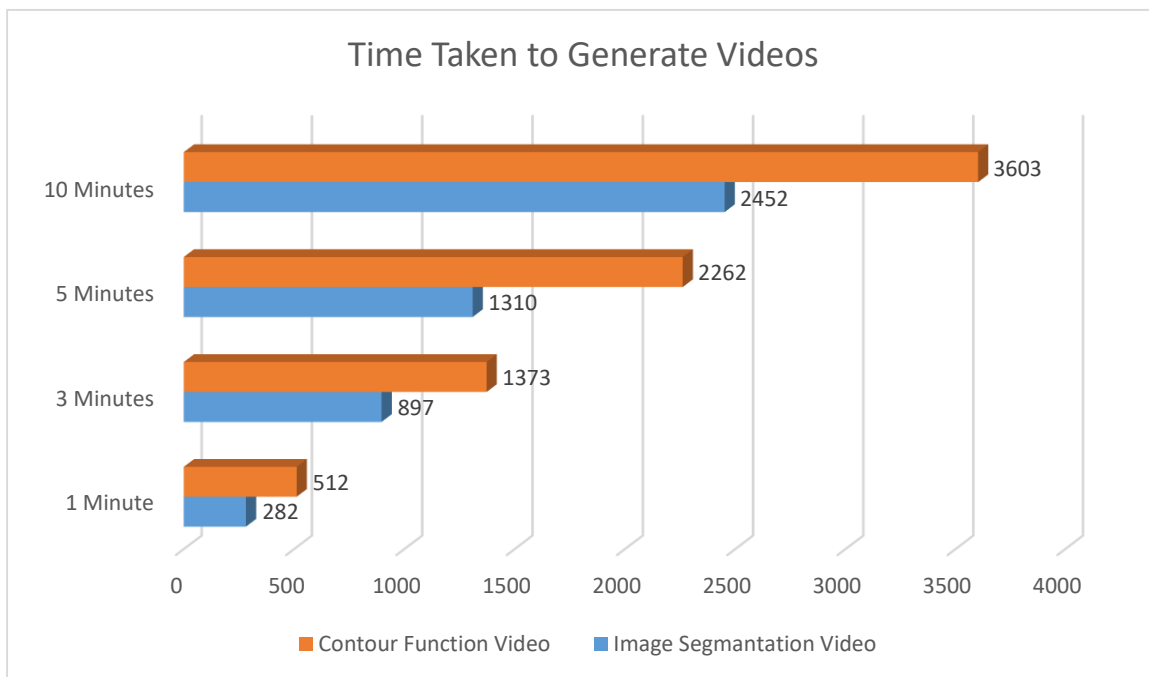
Fig. 4.3. Comparison of Running Time (in Sec.)

Table 4.2. denotes that time taken to run the program for contour function video generation for different video duration. The average running time of two algorithms depicts in Fig. 4.3. In terms of running time, the mean based segmentation is better than the contour segmented based summarization.

Table. 4.3. Percentage of Summarization of Algorithm #2

S. No	Video Duration	Frames in		Percentage of Summarization
		Original Video	Summarized Video	
01	1 Minute	1936	1527	78.87%
02	3 Minutes	4452	3559	79.94%
03	5 Minutes	7697	5379	69.88%
04	10 Minutes	14942	9659	64.64%

Table 4.3 listed the percentage of summarization for their efficiency. The corresponding the graph is plotted in Fig. 4.3. According, the tables and graph, the summarized algorithm based on contour segmentation is better than mean based algorithm.

5. Conclusions

In this paper, the two summarization algorithm are developed using the mean of mean of the frames of the videos and contour based segmentation. These algorithms are evaluated by the Time Complexity and Percentage of summarization. From the results of these algorithm, the contour segmentation based video summarization is better than mean based summarization in terms of percentage of summarization whereas in terms of running time, mean based video summarization.

References

[1].   U. Damnjanovic, V. Fernandez, E. Izquierdo, "Event Detection and Clustering for Surveillance Video Summarization". In: Proceedings of the Ninth International Workshop on Image Analysis for Multimedia Interactive Services. IEEE Computer Society, Washington, USA (2008).

[2].   P. Kalaivani, S. M. M. Roomi, Towards Comprehensive Understanding of Event Detection and Video Summarization Approaches". In Recent Trends and Challenges in Computational Models (ICRTCCM), 2017 Second International Conference, IEEE (2017, February): 61-66, 2017.

[3].   K. Kumar & D. D. Shrimankar. "F-DES: Fast and Deep Event Summarization". IEEE Transactions on Multimedia, 20(2), pp. 323-334, 2018.

[4].   S. S. Thomas, S. Gupta, V. K. Subramanian, Perceptual video summarization- "A new framework for video summarization". IEEE Transactions on Circuits and Systems f or Video Technology, 27(8), pp. 1790-1802, 2016.

[5].   F. Cricri, S. Mate, I. D. Curcio & M. Gabbouj. "Salient event detection in basketball mobile videos". In Multimedia (ISM), 2014 IEEE International Symposium. IEEE pp. 63-70, 2014.

[6].   S.J. Andaloussi, A. Mohamed, N. Madrane & A. Sekkaki. "Soccer video summarization using video content analysis and social media streams". In Proceedings of the 2014 IEEE/ACM International Symposium on Big Data Computing, IEEE Computer Society, pp.1-7, 2014.

[7].   M. Cote, F. Jean, A. B. Albu & D. Capson. "Video summarization for remote invigilation of online exams". In Applications of Computer Vision (WACV), 2016 IEEE Winter Conference pp. 1-9, 2016.

[8].   F. Wang, C.W. Ngo, "Summarizing rushes videos by motion, object and event understanding". IEEE Transactions on Multimedia 14 (2012).

[9].   M. Miniakhmetova & M. Zymbler, "An approach to personalized video summarization based on user preferences analysis". In Application of Information and Communication Technologies (AICT), 2015 9th International Conference, pp.153-155, 2015.

[10].   A. A. Rav, Y. Pritch, S. Peleg, Making a long video short: Dynamic video synopsis. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR' 06), Vol. 1, pp. 435-441, 2006.

[11].   K. Davila, R. Zanibbi. "Whiteboard Video Summarization via SpatioTemporal Conflict Minimization". In Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference, vol 1, pp. 355-362, 2017.

[12]. Z. Lu, K. Grauman, "Story Driven Summarization for Egocentric Video. 2013 IEEE Conference on Computer Vision and Pattern" Recognition, pp. 2714-2721, 2013. http://doi.org/10.1109/CVPR.2013.350.

[13]. S. Lu, M. R. Lyu, I. King, H. K. Sar, "Semantic Video Summarization Using Mutual Reinforcement Principle and Shot Arrangement Patterns", 2005.

[14]. J. Varghese, K. R. Nair, "An Algorithmic Approach for General Video Summarization". In 2015 Fifth International Conference on Advances in Computing and Communications (ICACC) pp. 7-11, 2015.

[15]. K. Muhammad, T. Hussain, and S.W. Baik, "Efficient CNN based summarization of surveillance videos for resource constrained devices". Pattern Recognition Letters, 130, pp. 370-375, 2020.

[16]. C. W. Ngo, Y. F. Ma, H. J. Zhang, "Video summarization and scene detection by graph modeling". IEEE Transactions on circuits and systems for video technology, 15(2) pp. 296-305, 2005.

Lung Image Segmentation for Identifying the COVID

Srinithi S, Eliahim Jeevaraj P S
Department of Computer Science,
Bishop Heber College (Autonomous), Tiruchirappalli.
eliahimps.cs@bhc.edu.in

Abstract

Since the onset of 2020, the spread of corona virus disease (COVID-19) has rapidly accelerated worldwide into a state of severe pandemic. COVID-19 has infected more than 29 million people and caused more than 900 thousand deaths at the time of writing. Since it is highly contagious, it causes explosive community transmission. Thus, health care delivery has been disrupted and compromised by the lack of testing kits. COVID-19-infected patients show severe acute respiratory syndrome. Meanwhile, the scientific community has been involved in the implementation of deep learning (DL) techniques to diagnose COVID-19 using computed tomography (CT) lung scans, since CT is a pertinent screening tool due to its higher sensitivity in recognizing early pneumonic changes. However, large datasets of CT-scan images are not publicly available due to privacy concerns and obtaining very accurate models has become difficult. COVID-19 infected people have symptoms that are related to pneumonia, and the virus affects the body's respiratory organs, making breathing difficult. Due to a shortage of kits, suspected patients cannot be treated promptly, resulting in disease spread. To develop an alternative, radiologists looked at the changes in radiological imaging, like CT scans, that produce comprehensive pictures of the body of excellent quality. In this project, lung CT scans are considered to classify the covid-19 injections using Optimal Contour Segmentation and UNet classification algorithm. UNet classification is the conventional Convolutional Neural Network.

Keywords: Lung image classification, COVID-19, U-Net

1. Introduction

Corona virus is a large family of viruses that can cause a human being to developa serious illness. The first reported major epidemic was Severe Acute Respiratory Syndrome (SARS) [1] in 2003, while the second severe outbreak of Middle East Respiratory Syndrome (MERS) [2] in Saudi Arabia began in 2012. The latest outbreak of corona virus disease was announced in late December 2019. This new virus is very infectious and has spread globally rapidly. On January 30, 2020, as it had spread to 18 countries, the World Health Organization (WHO) declared this epidemic a Public Health Emergency of International Concern (PHEIC) [3]. This virus was named 'COVID-19' by the World Health Organization on February 11, 2020 [4]. As of September 2020, the WHO reported that 31.3 million confirmed cases and over 965 thousand deaths have been registered in 213 countries.

Fig. 1.1. shows confirmed cases of global COVID-19 as of September 2020. Thedisease has spread rapidly around the globe since it was first identified and has become an international concern. An analysis performed by Jiang et al. [5] found that COVID- 19's death rate is 4.5% worldwide. In the age group of 70-79 years, the death rate for patients is 8.0%, while 14.8% for patients over 80 years. Patients over 50 years of age with chronic diseases are at the highest risk and it is critically important to find a way todetect illness before getting into serious conditions.
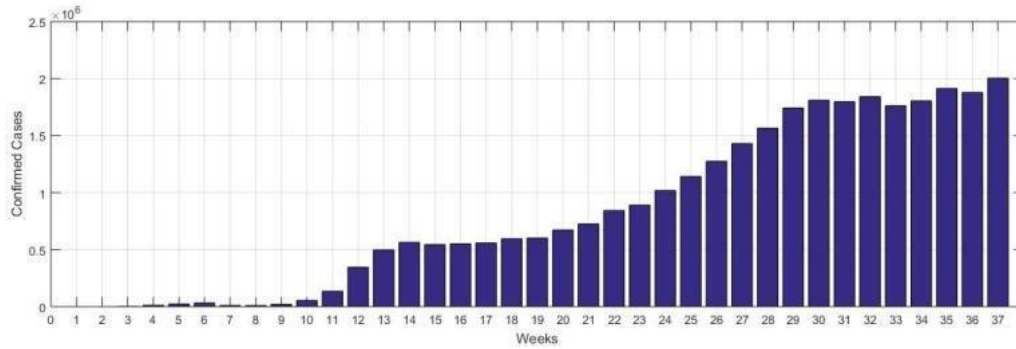
Fig. 1.1. Global COVID 19 confirmed cases as of September 2020

Early identification of patients with COVID-19 pneumonia for timely treatment is crucial to contain the spread, particularly in epidemic regions. According to information shared by the Radiological Society of North America (RSNA), X-ray and CT images of a Chinese person dead by COVID-19 showed the damages done to the human lungs. A research team lead by Lucas [7] at The University of São Paulo demonstrated the chest imaging finding of COVID-19 on different modalities such as Chest Radiography (CXR), Computed Tomography (CT), and Ultra sonography. According to them, chest CT is the main imaging method used in the assessment of COVID-19 pneumonia. A structured chest CT report standardizes imaging results and optimizes contact with the prescribing physician, making it a valuable tool in the pandemic scenario. In addition, the CT imaging properties of infected lungs include Ground-Glass Opacity (GGO) and severity-correlated consolidation. In Hubei Province,China, CT scans have been used widely and on display in order to discover, isolate, and control the epidemic's progress as quickly as possible.

Many investigations have shown that chest CT sensitivity is high in the detectionof COVID-19 pneumonia .Previous studies have shown that the most common CT characteristic of COVID-19 pneumonia is the presence of multifocal Ground-Glass Opacity (GGOs). Fig.1.2. displays the CT scan image of a COVID-19 patient. Arrowheads reveal the recognizable hazy area on the outer edges of the lungs. As per the description, Ground-Glass Opacities refer to the distorted presence of the lungs in imaging experiments, almost as though parts were obscured by Ground-Glass. This maybe due to the fluid filling of pulmonary airspace, the collapsing of airspace, or both. This is a trend that can be seen while the lungs are sick. Regular lung's CT scans look black; rare chest CTs with GGOs reveal lighter color or gray spots. Consolidation refers to the saturation of fluids or other inflammatory products in pulmonary airspace. Pleural effusion refers to abnormal fluids that form in the spaces surrounding the lungs.
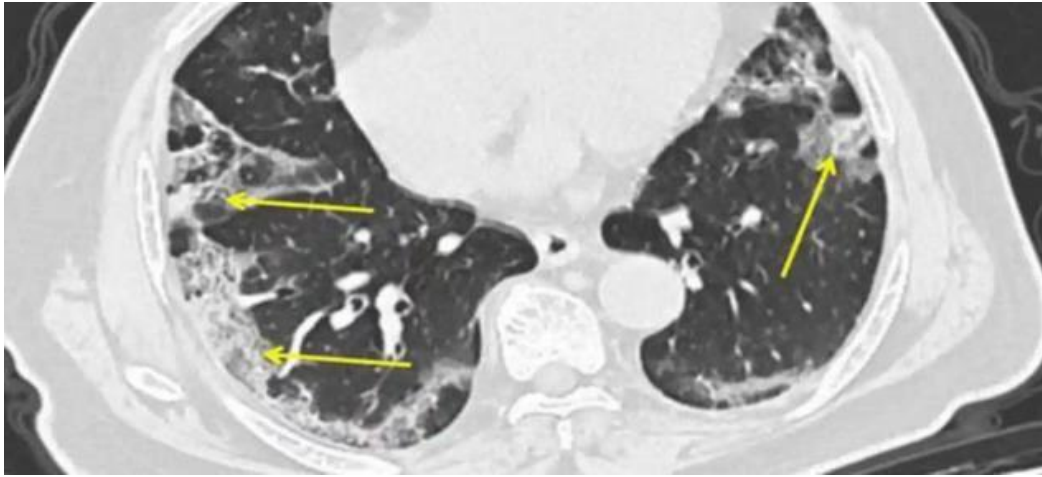
Fig.1.2. CT scan image of a patient with severe COVID 19

PCR tests are taking time to diagnose COVID-19 patients and the test results appear to be of low accuracy compared to CT scan tests. However, CT scans can be used as a simple and quick way of categorizing patients into "probably positive" and "probably negative" cohorts. As the hospital admission rate of COVID-19 patients increases, the PCR test is not appropriate. Nowadays, tools for the identification of COVID-19 patients with high efficacy and accuracy are essential. Due to the poor contrast of infection regions of CT images and the large differences in both the shape and location of the lesions in different patients, the delineation of the infection regionsin CT scans in the chest is very difficult for the physician. Image processing techniques may open new pathways to describe the state of the lungs using CT scans.

The Introduction are given in section 1. Section2 provides detailed review about the proposed method. Section 3 provides methodology used in the system. Result and Discussion in the system are mentioned in section 4. Conclusion was discussed in section 5.

2. Related Works

Bhattacharya, Sweta, et al [1] focused on compiling a list of the most recent research papers on deep learning applications for COVID-19 medical image processing Then we go over deep learning and its applications in healthcare that have been discovered in the last decade. Next, three use cases in China, Korea, and Canada are also presented to show deep learning applications for COVID-19 medical image processing. Finally, we discuss several challenges and issues related to deep learning implementations for COVID-19 medical image processing, which are expected to drive further studies in controlling the epidemic and controlling the crisis, which results in smart healthy cities.

Horry, Michael J., et al [2] proposed an image pre-processing stage to build a dependable image collection for development and testing the deep learning models. The new method aims to remove extraneous noise from photos so that deep learning models may concentrate on detecting diseases with specific characteristics. The results indicate that Ultrasound images provide superior detection accuracy compared to X-Ray and CT scans. The experimental results highlight that with limited data, most of the deeper networks struggle to train well and provides less consistency over the three imaging modes we are using. The selected VGG19 model, which is then extensively tuned with appropriate parameters, performs in considerable levels of COVID-19 detection against pneumonia or normal for all three lung image modes with the precision of up to 86% for

X-Ray, 100% for Ultrasound and 84% for CT scans.

Garain, Avishek, et al. [3] The outbreak of a global pandemic called corona virus has created unprecedented circumstances resulting into a large number of deaths and risk of community spreading throughout the world. Desperate times have called for desperate measures to detect the disease at an early stage via various medically proven methods like chest computed tomography (CT) scan, chest X-Ray, etc., in order to prevent the virus from spreading across the community. Developing deep learning models for analyzing these kinds of radiological images is a well-known methodology in the domain of computer based medical image analysis. However, doing the same by mimicking the biological models and leveraging the newly developed neuromorphic computing chips might be more economical. These chips have been shown to be more powerful and are more efficient than conventional central and graphics processing units. Additionally, these chips facilitate the implementation of spiking neural networks (SNNs) in real-world scenarios. To this end, in this work, we have tried to simulate the SNNs using various deep learning libraries. We have applied them for the classification of chest CT scan images into COVID and non-COVID classes. Our approach has achieved very high F1 score of 0.99 for the potential-based model and outperforms many state-of-the-art models. The working code associated with our present work can be found here.

Wang, Bo, et al. [4] presented the expertise developing and implementing an AI system that scans CT images automatically and calculates the chance of infection in order to detect COVID-19 pneumonia quickly. The suggested approach, which consists of classification and segmentation, will save clinicians 30–40 percent of their detection time and improve COVID-19 detection performance. Working in an interdisciplinary team of around 30 persons with medical and/or AI backgrounds who are geographically spread out between Beijing and Wuhan., the authors ,are In this circumstance, we were able to overcome a number of obstacles (such as data discrepancy, evaluating the model's time-effectiveness, data security, and so on) and deploy the system in four weeks. Furthermore, because the suggested AI approach prioritizes each CT scan based on the likelihood of infection, physicians may confirm and separate sick patients in real time. A total of 1,136 training cases (723 COVID-19 positives) from five hospitals were used. On the test dataset, which comprised a number of lung disorders, the scientists were able to acquire a sensitivity of 0.974 and a specificity of 0.922.

Panwar, Harsh, et al. [5] proposed a deep transfer learning algorithm that accelerates the detection of COVID-19 cases by using X-ray and CT-Scan images of the chest. It is because, in COVID-19, initial screening of chest X-ray (CXR) may provide significant information in the detection of suspected COVID-19 cases. The authors have considered three datasets known as 1) COVID-chest X-ray, 2) SARS-COV-2 CT-scan, and 3) Chest X-Ray Images (Pneumonia). In the obtained results, the proposed deep learning model can detect the COVID-19 positive cases in ≤ 2 seconds which is faster than RT-PCR tests currently being used for detection of COVID-19 cases. The authors have also established a relationship between COVID-19 patients along with the Pneumonia patients which explores the pattern between Pneumonia and COVID-19 radiology images. In all the experiments, the authors have used the Grad-CAM based color visualization approach in order to clearly interpretate the detection of radiology images and taking further course of action.

Ahuja, Sakshi, et al [6] In the proposed research work; the COVID-19 is detected using transfer learning

from CT scan images decomposed to three-level using stationary wavelet. A three-phase detection model is proposed to improve the detection accuracy and the procedures are as follows; Phase1- data augmentation using stationary wavelets, Phase2- COVID-19 detection using pre-trained CNN model and Phase3- abnormality localization in CT scan images. This work has considered the well known pre-trained architectures, such as ResNet18, ResNet50, ResNet101, and SqueezeNet for the experimental evaluation. In this work, 70% of images are considered to train the network and 30% images are considered to validate the network. The performance of the considered architectures is evaluated by computing the common performance measures. The result of the experimental evaluation confirms that the ResNet18 pre- trained transfer learning-based model offered better classification accuracy (training = 99.82%, validation = 97.32%, and testing = 99.4%) on the considered image dataset compared with the alternatives.

Silva, Pedro, et al. [7] proposed COVID-19 screening with a voting-based approach using an efficient deep learning technique. The photos from a certain patient are categorized as a group in this method, which uses a voting system. With a patient- based split, the method is tested in the two largest COVID-19 CT datasets. A cross dataset analysis is also presented to evaluate the models' robustness in a more realistic scenario using data from various distributions. The task's generalization capacity is far from satisfactory, as accuracy reduces from 87.68 percent to 56.16 percent in the best evaluation scenario, according to the cross-dataset study. These findings show that approaches for detecting COVID-19 in CT scans must greatly improve in order to be evaluated to test the approaches in a realistic environment, larger and more diversified datasets are required as a clinical option.

3. Methodology

The first step is to give the input lung CT scan images. Contrast Limited Adaptive Histogram Equalization is used to enhance the input images. Contour Technique is used to segment the images. Segmented images are classified by the use ofU_Net method, to classify that the CT Lung images are Covid-19 or non Covid. Finally predict the values of Precision, Recall, Accuracy, and F1-score.
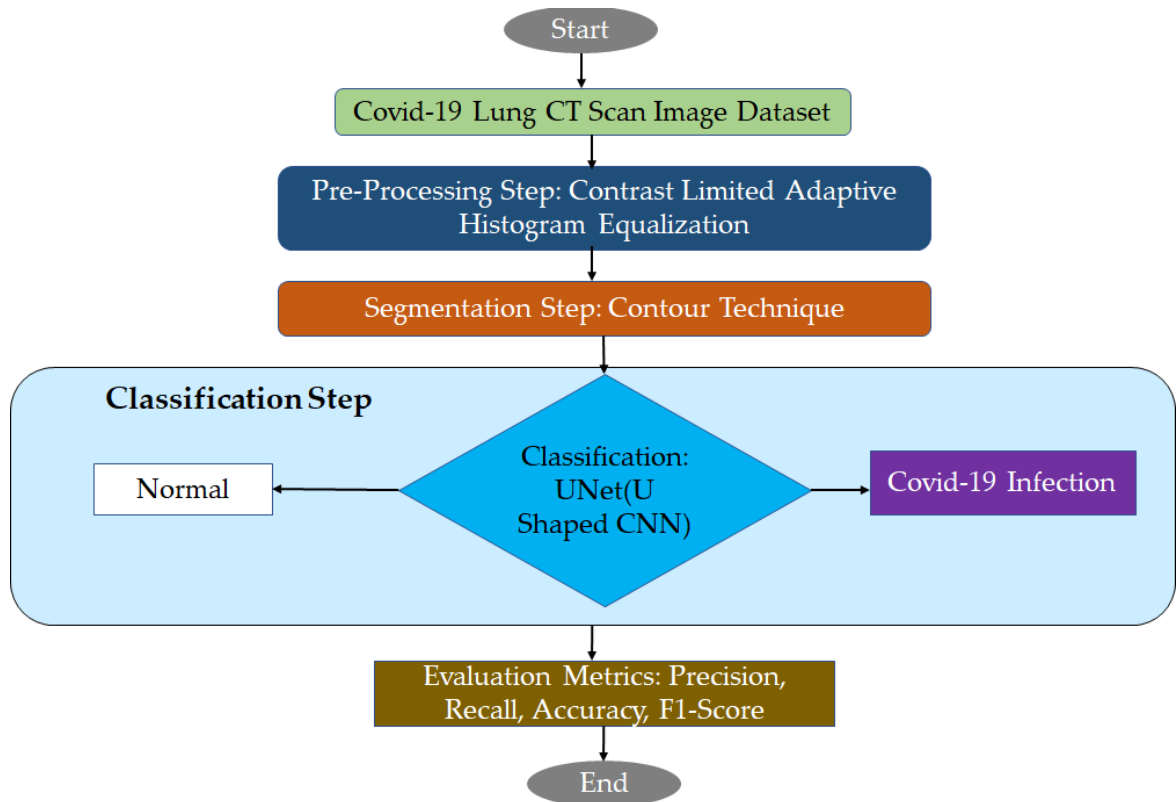
Fig.3.1. Flowchart of the proposed Algorithm

3.1. PRE-PROCESSING USING CONTRAST LIMITED ADAPTIVE HISTOGRAM EQUALIZATION

Image enhancement consists of image quality Improvement processes, allowing a better visual and computational analysis. It is widely used in several applications due to its capability to overcome some of the limitations presented by image acquisition systems. Deblurring, noise removal, and contrast enhancement are some examples of image enhancement operations. The idea behind contrast enhancement is to increase the dynamic range of the gray levels in the image being processed. It plays a major role in digital image processing, computer vision, and pattern recognition. Besides providing a better visual interpretation by improving the imageappearance, the contrast enhancement may also be used to improve the performance of succeeding tasks, such as image analysis, object detection, and image segmentation. In fact, it has contributed in a variety of fields like medical image analysis, high definition television (HDTV), industrial X-ray imaging, microscopic imaging, and remote sensing.

Most of the contrast enhancement techniques are based on histogram adjusts, dueto their straight forward and intuitive implementation qualities. These techniques are often categorized in global or local techniques. The use of global contrast enhancement may be not suitable for images whose local details are necessary or for images containing varying lighting. conditions. On the other hand, if the process is applied locally, it is possible to overcome those limitations. The Contrast Limited Adaptive Histogram Equalization(CLAHE) is a popular method for local contrast enhancement that has been showing powerful and useful for several

applications.CLAHE has been extensively used to enhance image contrast in several computer vision and pattern recognition applications. In the medical field, it was successfully applied in breast ultrasound and mammography image enhancement, in cell image segmentation, in retinal vessel image processing, and in enhancement of bone fracture images. Beyond medical field, CLAHE was applied to enhance underwater images, to perform fruit segmentation in agricultural systems, and to assist driving systems to improve vehicle detection, traffic sign detection, and pedestrian detection.

The basic idea of CLAHE consists in performing the histogram equalization of non-overlapping sub-areas of the image, using interpolation to correct inconsistencies between borders. CLAHE has also two important hyper parameters: the clip limit (CL) and the number of tiles (NT). The first one (CL) is a numeric value that controls the noise amplification. Once the histogram of each sub-area is calculated, they are redistributed in such a way that its height does not exceed a desired "clip limit." Then, the cumulative histogram is calculated to perform the equalization. The second (NT) is an integer value which controls the amount of non-overlapping sub-areas: based on its value, the image is divided into several (usually squared) non-overlapping regions of equal sizes. For 512×512 images, the number of regions is generally selected to be equal to 64 (NT =[8, 8]).

3.2. CONTOUR TECHNIQUE BASED SEGMENTATION TECHNIQUE

Image processing can be defined as computerized processing of images of different types to obtain the desired output. Image processing makes use of a wide rangeof techniques to process the input information which is available in the form of an image. Processing of images is carried out by avoiding certain features like noise and signal distortion that affects the information present in the images. The images can be defined in different dimensions which can be used for processing. Segmentation is a part of image processing used for segregation of regions.

Segmentation is the process of separation of required information from a datafor further processing. Image segmentation can be defined as the segregation of pixels of interest for effective processing. The main aim of image segmentation is to segment the meaningful regions of interest for processing. Region of interest possesses a groupof pixels defined with a boundary and these may contribute to different forms such as circle, ellipse, polygon or irregular shapes. The process of segmentation does not provide information about the entire image rather associates pixel data of only the region of interest. Segmentation is a crucial process in Image analysis because it paves path for future processing of images.

In medical image analysis, segmentation is very much necessary where region ofstudy or research is defined to a particular section of the image. If image segmentationis performed effective, the after stages of image analysis are made easier. Image segmentation provides definite and useful information or data for the high standards of automatic image analysis. Image analysis defines certain objectives for segmentation process:

- Decompose the image into parts for future analysis.

- Change in representation

- Region of interest should be simple, uniform and homogenous with smoothboundary

Active contour is a type of segmentation technique which can be defined as use of energy forces and constraints for segregation of the pixels of interest from the image forfurther processing and analysis. Active contour described as active model for the process of segmentation. Contours are boundaries designed for

the area of interest required in an image. Contour is a collection of points that undergoes interpolation process. The interpolation process can be linear, splines and polynomial which describes the curve in the image. Different models of active contours are applied for the segmentation technique in image processing. The main application of active contours in image processing is to define smooth shape in the image and forms closed contour for the region. Active contour models involve snake model, gradient vector flow snake model, balloon model and geometric or geodesic contours.

Active contours can be defined as the process to obtain deformable models or structures with constraints and forces in an image for segmentation. Contour models describe the object boundaries or any other features of the image to form a parametric curve or contour. Curvature of the models is determined with various contour algorithms using external and internal forces applied. Energy functional is always associated with the curve defined in the image. External energy is defined as the combination of forces due to the image which is specifically used to control the positioning of the contour onto the image and internal energy, to control the deformable changes [3]. Constraints for a particular image in the contour segmentation depend on the requirements. The desired contour is obtained by defining the minimum of the energy functional. Deforming of the contour is described by a collection of points that finds a contour. This contour fits the required image contour defined by minimizing the energy functional.

Active contours can also be used for segmentation of 3-D images derived from different medical imaging modalities. 2-D slices of image data are used for the separation of target object from the 3-D images. These 2-D slices of images in all directions along with the segmented target region are subjected to 3-D reconstruction to segregate the regions. Mesh model of the 3-D image is designed before applying active contour model. The mesh helps in the formation of deformable contours of the target object in the directional 2-D slices of the 3-D images.

3.3. U-NET CLASSIFICATION METHOD

U-net network can be divided into two parts: The first part is the contracting path which uses a typical CNN architecture. Each block in the contracting path consists of two successive 3x3 convolutions followed by a ReLU activation unit and a max- pooling layer. This arrangement is repeated several times. The novelty of u-net comes inthe expansive path where at each stage the feature map is upsampled using 2x2 up- convolution. Then, the feature map from the corresponding layer in the contracting path is cropped and concatenated onto the upsampled feature map. This is followed by two successive 3x3 convolutions and ReLU activation. At the final stage, an additional 1x1 convolution is applied to reduce the feature map to the required number of channels andproduce the segmented image. The cropping is necessary since pixel features in the edges have the least amount of contextual information and therefore need to be discarded. This results in a network resembling a u-shape and more importantly, propagates contextual information along the network which allows it to segment objectsin an area using context from a larger overlapping area.
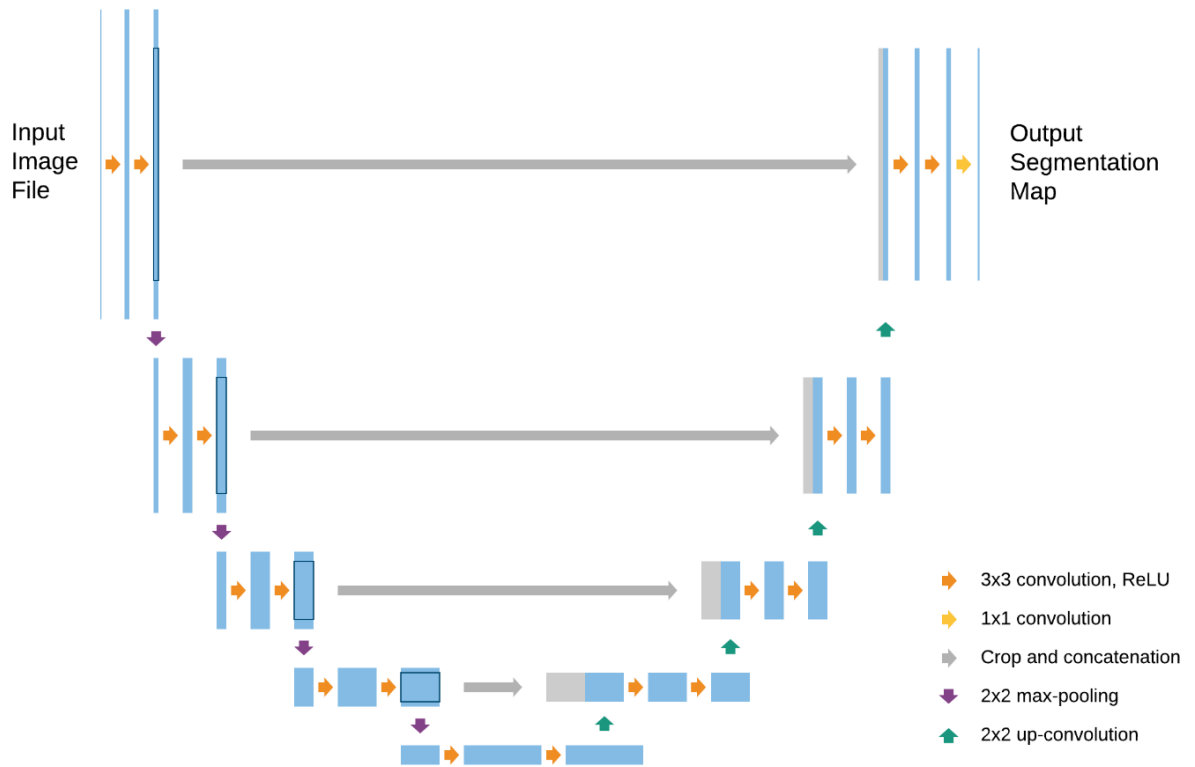
Fig.3.2. Architecture of U-Net Classification Method

3.5. PROPOSED ALGORITHM

-- --

Input: Original Lung CT images

Output: Segmented Lung CT images

-- ----

Step1: Load Covid-19 Lung CT image dataset into the python notebook.

Step2: Apply Contrast Limited Adaptive Histogram Equalization (CLAHE) for pre-processing.

Step 3: Apply Contour segmentation on the pre-processed histogram images.

Step 4: Apply UNet Classification on the segmented images for the detection of Covid-19.

Step 5: Evaluates the classification accuracy, precision, recall and F1-Score.

-- ----

4. Results & Discussion

This dataset is taken from kaggle. CT scans dataset consists four files such as CT scan, Lung mask, Infection mask, and Lung and Infection mask . totally twelve images for each four files in the dataset. Lung image segmentation for identifying the COVID has been enforced in PYTHON in Google Co laboratory with a system configuration of Lenovo Intel Core i5processor.

The proposed algorithm is assessed by Qualitative metrics such as Accuracy, Classification Report, and Confusion Matrix and Qualitative metrics such as Histogramanalysis, Human visual perception.

4.1. QUALITATIVE METRICS

4.1.1. Accuracy

Accuracy is a statistic that sums up how well a model performs across all classes. It's helpful when all of the classes are equally important. The ratio between thenumber of right guesses and the total number of forecasts is used to compute it.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

Table 4.1. Accuracy of proposed algorithm

EPOCHS	TIME	TEST_LABELS		TEST_IMAGES	
		Accuracy	Loss	Accuracy	Loss
Epoch 1	93s	0.5205	1.2618	0.5384	1.1062
Epoch 2	80s	0.5178	1.2446	0.5589	1.1598
Epoch 3	105s	0.2815	1.8463	0.2559	1.8771
Epoch 4	81s	0.2223	1.9930	0.1769	2.1515
Epoch 5	69s	0.1816	2.0714	0.1980	1.9163
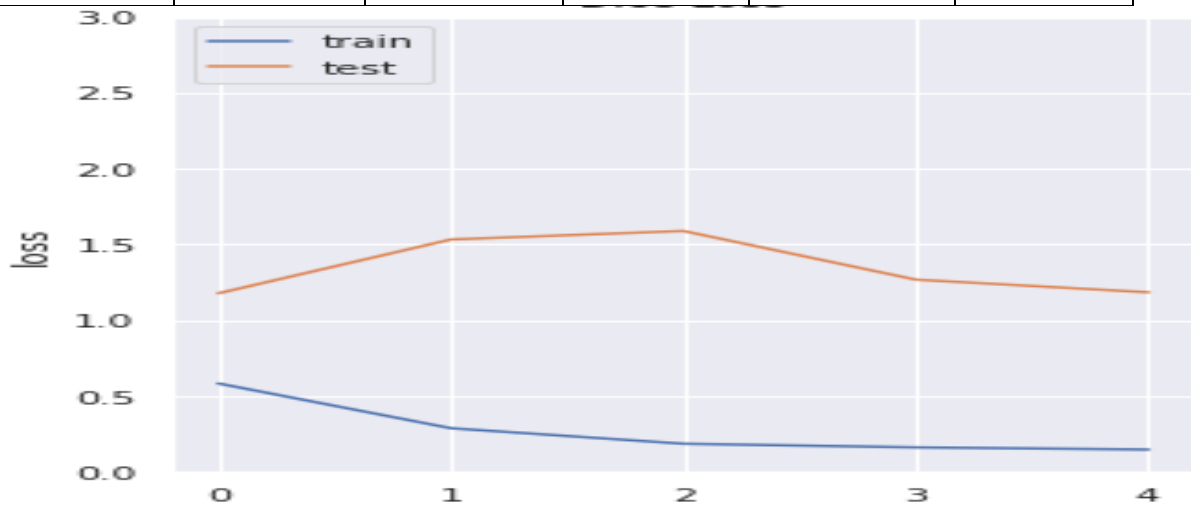
Fig. 4.1. Accuracy of proposed algorithm

4.1.2. Classification Report

A classification report is a statistic for evaluating performance. It is used to display the trained classification model's Precision, Recall, F1 Score, and Support.

Table 4.2. Classification Report of Proposed Algorithm

	PRECISION	RECALL	F1 SCORE	SUPPORT
0	0.60	0.50	0.55	6
1	0.00	0.00	0.00	0
2	0.33	0.33	0.33	3
Weighted Average	0.51	0.44	0.17	9

4.1.3. Confusion Matrix

A confusion matrix is a N x N matrix that is used to evaluate the performance of a classification model, with N denoting the number of target classes. The matrix compares the actual target values to the deep learning model's predictions.

Table 4.3. Confusion Matrix

LABELS	PREDICTION	WEIGHTS
[0 0 0 0 0] [0 1 0 0 0] [0 0 0 2 0] [0 0 0 0 0] [0 0 0 0 5]	shape=(5, 5)	dtype=int32

4.2. QUALITATIVE METRICS

4.2.1. Contrast Limited Adaptive Histogram Equalization
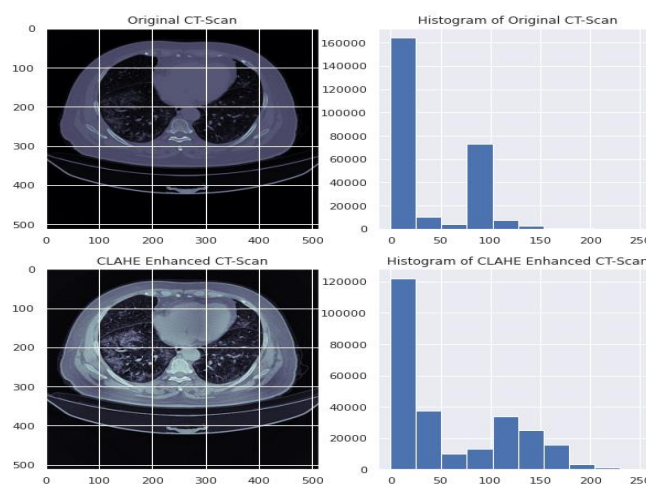
Fig.4.2. Histogram analysis for Original and Enhanced CT image
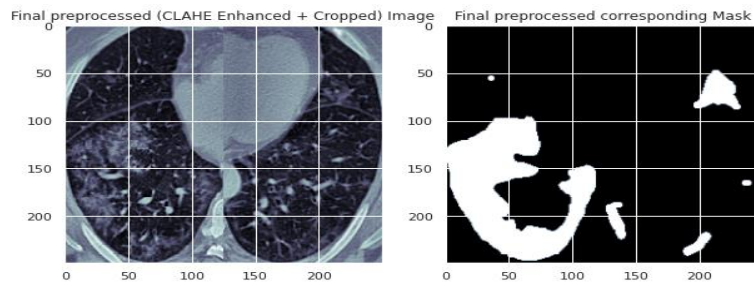
Fig. 4.3. Enhanced image of CT Lung Image

4.2.2. U-Net

U-net is convolutional network architecture of fast and precision segmentation of images.

Fig . 4.4. U-Net Classification of proposed algorithm
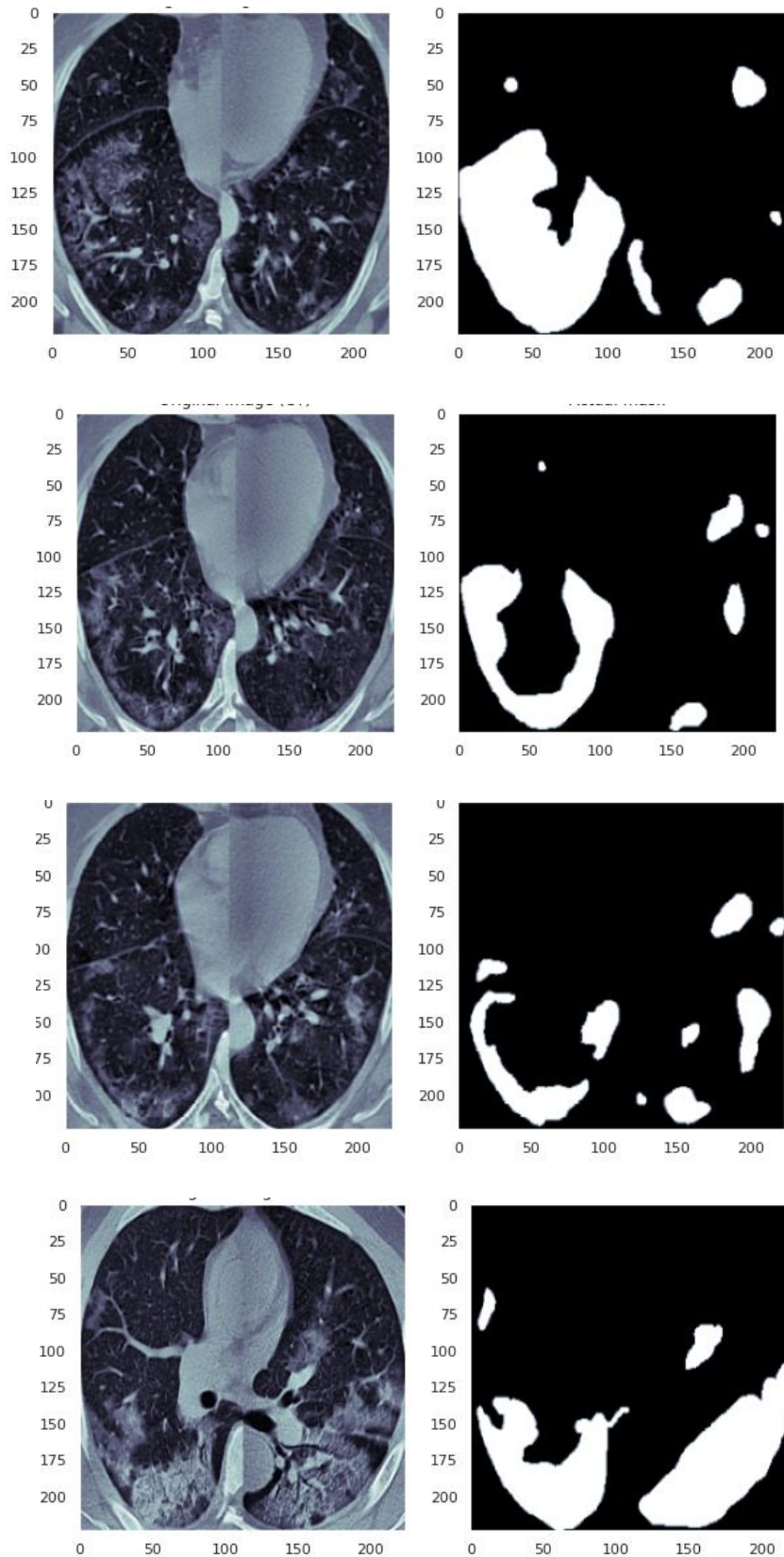
4.2.3. Segmentation
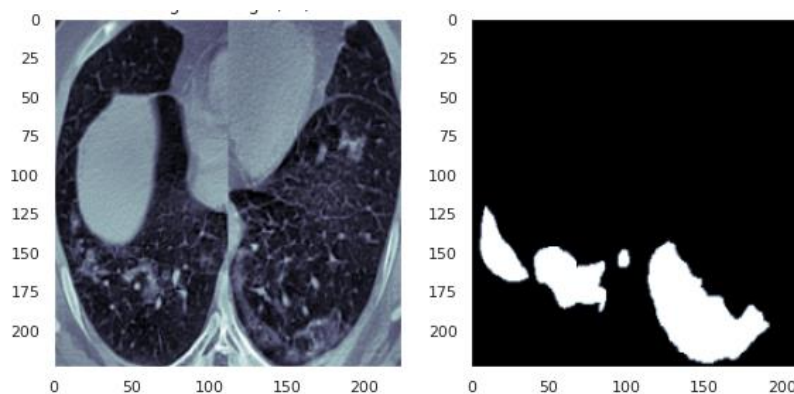
Fig.4.4. Segmented Images

5. Conclusion

In this paper, the classification of Covid-19 is identified by applying Histogram equalisation for removing noise, Contour segmentation and UNet classification. With the threshold of 0.52 and 0.81 in Contour segmentation, the accuracy, precision, recall and F1-score has increased using UNet classifier.

References

[1] Bhattacharya, Sweta, et al. "Deep learning and medical image processing for corona virus (COVID-19) pandemic: A survey." *Sustainable cities and society* 65 (2021): 102589.

[2] Horry, Michael J., et al. "COVID-19 detection through transfer learning using multimodal imaging data." *Ieee Access* 8 (2020): 149808-149824.

[3] Garain, Avishek, et al. "Detection of COVID-19 from CT scan images: A spiking neural network-based approach." *Neural Computing and Applications* 33.19 (2021): 12591-12604.

[4] Wang, Bo, et al. "AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system." *Applied Soft Computing* 98 (2021): 106897.

[5] Panwar, Harsh, et al. "A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images." *Chaos, Solitons & Fractals* 140 (2020): 110190.

[6] Ahuja, Sakshi, et al. "Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices." *Applied Intelligence* 51.1 (2021): 571-585.

[7] Silva, Pedro, et al. "COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis." *Informatics in medicine unlocked* 20 (2020): 100427.

ANALYSIS OF MORTALITY DUE TO COVID-19 USING MACHINE LEARNING TECHNIQUES

Chellammal Surianarayanan, Sharmila Rengasamy and Aravindhan Mani
Department of Computer Science
Government Arts and Science College
Navalurkuttapattu, Srirangam Tk., Tiruchirappalli
Affiliated to Bharathidasan University, Tamilnadu, India
drschellammal@gmail.com, sharmiparam@gmail.com, kumararavindh471@gmail.com

Abstract

Coronavirus disease-2019 (COVID-19) has become a global pandemic that is threatening far more than a health crisis. As the number of confirmed patients has explosively increased, there is a need for risk stratification both for preventing (i.e., home quarantine, social distancing) and for treating confirmed patients (i.e., hospitalization vs. community isolation). Although the COVID-19 pandemic has caused more than 5,000,000 death worldwide, there are also a significant number of asymptomatic patients who become infected and recover asymptomatically. Identification of high-risk confirmed patients is required to allow better allocation of existing available medical resources. According to the Centers for Disease Control and Prevention (CDC), confirmed patients who are over 65 years old, who live in nursing homes, and who have at least one of the following conditions, chronic lung disease, serious heart conditions, severe obesity, diabetes, liver disease, and immunocompromised status, are at a high risk of death due to COVID-19. Although the CDC guideline has been used as a reference for overall patients, more prediction studies are required to find factors which are strongly related to mortality risk due to covid-19. This needs to be experimented with individual patient data. In this project, different factors like age, gender, symptoms, contact, vaccination, comorbidity, etc are being collected from individuals and analysis has been done using two algorithms, Naïve Bayes and Sequential Minimal Optimization(SMO) algorithms. The results presented.

Keyword: Mortality risk, covid-19, Navie Bayes, Sequential Minimal Optimization,

1 INTRODUCTION

The whole world is in distress due to the uninhibited spread of the COVID-19 virus. The news reports and video clippings emanating from different parts of the world are really appalling. Governments across the globe are working overtime to somehow stop this killer pandemic in the budding stage. Four variants of corona virus are of concern, alpha, beta, gamma and delta. As of October 6, 2021, the current dominant coronavirus variant is delta. The identified variant spread faster than other variants, and delta potentially cause more severe disease. Having understood the seriousness, medical practitioners and pioneers are striving hard and stretching further to unearth appropriate vaccines to arrest the epidemic forthwith. On the other side, IT specialists are skilfully leveraging a bevy of proven and potential digital technologies and tools to support the universal goal of surmounting this dangerous endemic. In addition, people are vaccinated. Despite the efforts from various domains, the virus is still spreading in its different variants. Now, it is spreading in its Omicron form everywhere in the World. In this situation, among the various issues around covid-19, one of the major issue to the analysis of mortality due to covid-19. This aspect forms the background of the research work. One of the difficulties involved in mortality prediction, is the unavailability of individual patient record. It is not easy to get patient related record from hospitals. In this project, it is

proposed to collect the data by survey. The primary objective of the proposed work is to analyse the mortality due to covid-19 using artificial intelligence algorithms with data collected from individuals. .

2 LITERATURE SURVEY

As stated in our survey paper [1], artificial intelligence-based techniques are extensively used for resolving various issues associated with corona virus, including surveillance [2], early detection and assessment[3], diagnosis [4-5]. Also, the applications of other digital techniques are discussed in another review paper [6]. Research works related to the prediction of mortality risk due to covid-19 infection using machine learning algorithms are discussed in [7-10].

3 PROPOSED METHOD

The higher-level block diagram of the proposed research work is shown in Fig. 1. It is proposed to analyze the mortality risk due to covid -19. Towards implementing the main objective of the work, questionnaire including various factors namely, age, gender, admission of patient into hospital, covid declaration date, symptoms, Influence like illness or severe acute syndrome infection, travel history, contact details, vaccination details, comorbidity and mortality. The data is collected from various individuals, relatives and known persons, etc. After data collection, the data has been archived into Excel as CSV file.

Also, the data has been converted into .arff file format for analysis using WEKA tool. Factors which are more correlated to mortality due to covid-19 are found out using filter-based technique. Pearson correlation measure is used. The Pearson correlation coefficient is used to measure the strength of a linear association between two variables, where the value r = 1 means a perfect positive correlation and the value r = -1 means a perfect negative correlation. It is calculated using the following formula.

$$r = \frac{\sum_{i}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i}(x_i - \overline{x})^2}\sqrt{\sum_{i}(y_i - \overline{y})^2}}$$

Two classification algorithms namely Naïve Bayes and Sequential Minimal Optimization have been used. Naïve Bayes algorithm is a simple probability based multi-class supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It helps in building the fast machine learning models that can make quick predictions.   The Sequential Minimal Optimization (SMO) algorithm is derived by taking the idea of the decomposition method to its extreme and optimizing a minimal subset of just two points at each iteration. The power of this technique resides in the fact that the optimization problem for two data points admits an analytical solution, eliminating the need to use an iterative quadratic programming optimizer as part of the algorithm.
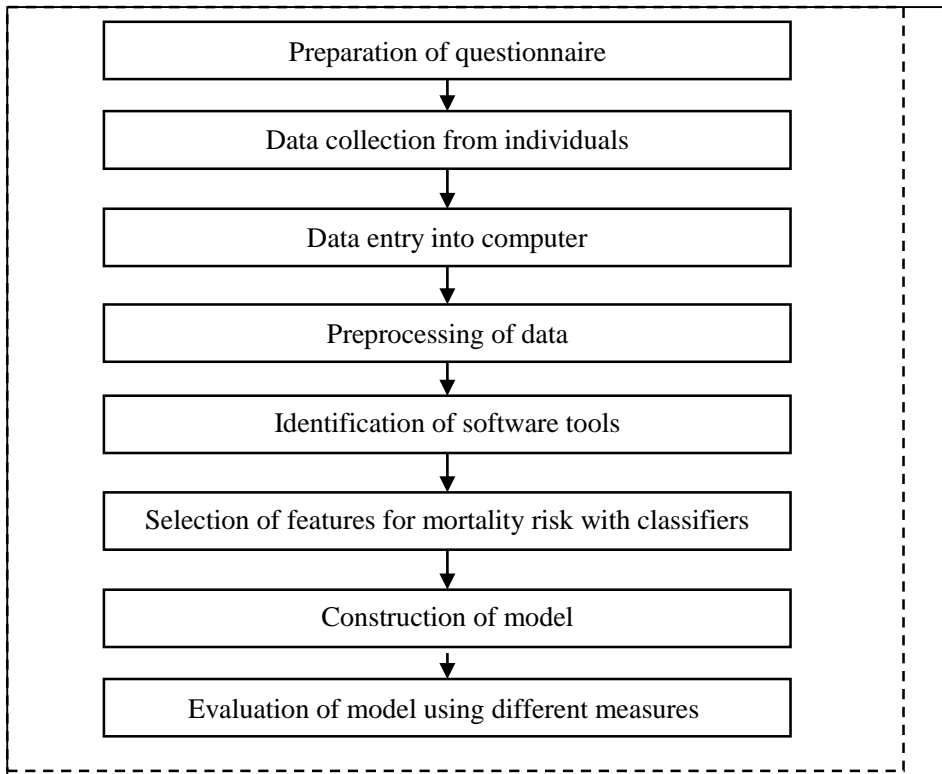
Fig. 1 Block diagram of the proposed methodology

4 EXPERIMENTATION

According to Pearson correlation measure, the attributes are ranked as follows.

- 0.9446 6 attribute_5 (Vaccination)

0.8507 7 attribute_6 (comorbidity)

0.7082 3 attribute_2 (symptoms)

0.5408 4 attribute_3 (Influenza like illness)

0.396 1 attribute_0 (age)

0.2821 5 attribute_4 (Contact)

0.0756 2 attribute_1 (gender)

Two experiments have been conducted using Naïve Bayes and SMO classifiers. In one experiment, all the attributes were used. In another experiment, two attributes namely contact and gender were removed as they are less correlated. The performance of the classifiers have been analyzed using accuracy, precision, recall and F-score. The values of accuracy, precision, recall and f_score obtained using the two algorithms Naïve Bayes and SMO for the two different feature sets are given Table 1.

Table 1 Performance measures of the two classifiers

Algorithm	Dataset	Accuracy	Precision	Recall	F-Score
Naïve Bayes	Complete	96.36%	95.5%	100%	97.7%
Naïve Bayes	Reduced	96.36	95.5%	100%	97.7%
SMO	Complete	98.18%	97.7%	100%	98.8%
SMO	Reduced	98.18%	97.7%	100%	98.8%

Thus, from the above table, both the algorithms could produce the same accuracy for the reduced feature (attribute-0, attribute-2, attribute-3, attribute-5 and attribute-6). The correlation coefficient of the above attributes are given below

-0.9446   6 attribute_5 (Vaccination)

0.8507   7 attribute_6 (comorbidity)

0.7082   3 attribute_2 (symptoms)

0.5408   4 attribute_3 (Influenza like illness)

0.396   1 attribute_0   (age)

CONCLUSION

Machine learning algorithms are extensively used to solve the issues related to COVID-19. In this project, mortality risk due to covid-19 has been analyzed using field data collected from individuals with the help of two algorithms Naïve Bayes and Sequential Minimal Optimization(SMO). Special emphasis is given in determining how attributes are correlated to mortality risk due to covid-19.   Experiments are carried out with two different feature sets and it is found that the attributes -0.9446   6 attribute_5 (Vaccination), 0.8507 7 attribute_6 (comorbidity), 0.7082   3 attribute_2 (symptoms), 0.5408   4 attribute_3 (ILI/SARI) and 0.396 1 attribute_0   (age) are more correlated with mortality risk due to covid-19.

REFERENCES

[1] Chellammal Surianarayanan, Pethuru Raj Chelliah, "Leveraging Artificial Intelligence (AI) Capabilities for COVID-19 Containment", New Generation Computing, 39, 717–741 (2021). https://doi.org/10.1007/s00354-021-00128-0

[2] Long, J.B., Ehrenfeld, J.M, "The Role of Augmented Intelligence (AI) in Detecting and Preventing the Spread of Novel Coronavirus", J Med Syst 44, 59 (2020). https://doi.org/10.1007/s10916-020-1536-6

[3] Srinivasa Rao ASR and Vazquez JA, "Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone–based survey when cities and towns are under quarantine", Infection Control & Hospital Epidemiology, 2020. https://doi.org/10.1017/ice.2020.61

[4] Tao Aiet al, , "Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases", Radiology, Feb 2020, [Online] https://doi.org/10.1148/radiol.2020200642

[5] Udugama Bet al., "Diagnosing COVID-19: The Disease and Tools for Detection", ACS Nano, vol.14, no.4, 30 Mar 2020, [Online] Available: https://doi.org/10.1021/acsnano.0c02624

[6] Chellammal Surianarayanan, Pethuru Raj Chelliah, "Covid-19 Containment: Demystifying the Research Challenges and Contributions Leveraging Digital Intelligence Technologies", In book: Machine Intelligence and Smart Systems (pp.193-214), DOI:10.1007/978-981-33-4893-6_18

[7] Ashis Das, Shiba Mishra, Saji Saraswathy Gopalan, "Predicting community mortality risk due to CoVID-19 using machine learning and development of a prediction tool", 2020, DOI: 10.1101/2020.04.27.20081794

[8] Mohammad Pourhomayoun, Mahdi Shakibi, "Predicting Mortality Risk in Patients with COVID-19

Using Artificial Intelligence to Help Medical Decision-Making", medRxiv, DOI: https://doi.org/10.1101/2020.03.30.20047308

[9]   Yan, L., Zhang, H., Goncalves, J. et al., "An interpretable mortality prediction model for COVID-19 patients", Nat Mach Intell 2, 283–288 (2020). https://doi.org/10.1038/s42256-020-0180-7

[10] Jiang, X., Coffee, M., Bari, A., Wang, J., Jiang, X. et al, "Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coron Clinical Severity". CMC-Computers, Materials & Continua, 63(1), 2020, 537–551.

A SECURE ENERGY EFFICIENT APPROACH IN UNDERWATER SENSOR NETWORK USING EDA CLUSTER

. K. Prabavathy and K.S.Senthilkumar

Department of computer Science,
Rathinam college of Arts and Science, Coimbatore.
innovationinfo6@gmail.com

Abstract

Underwater wireless Sensor Networks (UWSNs) are used for detecting as well as monitoring the underwater environment. This network includes several sensors as well vehicles located in a selected area for performing particular tasks. These sensor networks are connected further with base stations as well as satellites for processing the detected data for processing further. Network of self-powered sensor nodes as well as autonomous vehicles are deployed in under water as well as performing collaborative activities using acoustic links. The two important issues of underwater sensor networks are energy constraints and intrusion detection. Energy efficient intrusion detection techniques have proposed in this research work.

Keywords:UWSN, intrusion detection, EDA, Energy Efficient, EAVAR – LPA

1. INTRODUCTION

UWSN have received great attention due to its importance in oceanic observation as well as exploration. This sensor networks provide a vast range of applications that requires routing of sensed data to the centralized location. Routing is a vital task in all applications. It has demonstrated its strength in various applications like ocean monitoring, surveillance, resource exploration, as well as military in underwater environment.   The researches have mainly focused on self-organization, communication, processing capabilities, connectivity, consumption of low energy, and adaptability.   Underwater wireless sensor networks are similar to wireless sensor networks, however, it has differentiated in some aspects.

1.1 Limited hardware resources

UWSNs are limited in hardware resources that include computational capability, energy, as well as storage space. The power gained for under water acoustic communication is more than the terrestrial communication. In this network, sensor nodes can be deployed either in deep water or shallow where it is difficult for charging or replacing the battery of the nodes. In order to increase the network life time, storage space as well as computational capabilities are constricted. The existing researches have mainly focused on saving energy consumption at the expense of security as well as capability.

1.2 Communication channel unreliability

UWSNs is spatially variable, temporally, limited bandwidth as well as highly depends on both frequency as well as transmission range. For farther distance, acoustic channels' bandwidth is low. Most of the acoustic system may operate below the frequency level 30 kHz. The acoustic channel of under water is affected by path loss, multipath effect, temperature of water, as well as Doppler spread [1]. These factors cause delay variance as well as high bit error that result in packet losses as well as node failures. Nodes in the network shares the acoustic channel results in attacker cam intercept passively as well as analyse packets and also actively disrupt the network services. It is a great challenge for protecting UWSNs from security

threats and malicious attacks.

1.3 Dynamic characteristic of network topology

Terrestrial senor nodes are deployed densely whereas in underwater the deployment of sensor nodes are sparser due to deployment challenges as well as cost involvement [2]. The flow of water makes the underwater sensors to be mobile. The topology of the underwater sensor is highly dynamic due to water conditions and objects' movement speed such as 36 km/h or 2-3 knots. In order to expand the communication and monitoring region, AUVs (Autonomous Underwater Vehicles) are utilised in various applications.
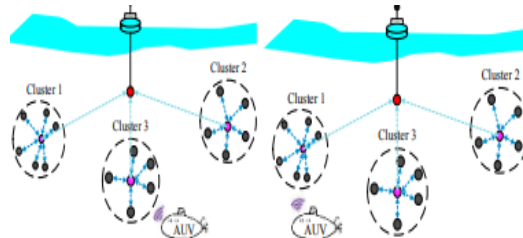

Fig. 1. Dynamic topology

High dynamic network topology is due to the AUVs entry or exit into the network or cluster. In the above figure 1a, AUV has joined in cluster 3 and in Figure 1b, AUV has joined in cluster 1. The changes in the network topology will result in dynamic data routing as well as accuracy changes.

1.4 Vulnerability challenges

Some of the underwater sensor communication network is not much secure namely target tracking as well as security monitoring. The nodes may be deployed in under water for monitoring hostile objects in high seas as well as sea regions. Such nodes are vulnerable for malicious threats and attacks [3].

In addition to the above challenges, UWSNs nodes may also be damaged physically. These are also impacted by the marine organisms as sensor nodes can be deployed at unattended or harsh deep sea that means these are unable for guarding from physical damages.

2. THREATS

Under water sensor networks are vulnerable for various malicious attacks as well as threats. Based on the actions and interest of malicious attacker, malicious attacks can be classified into active and passive.

2.1 Passive attacks

In passive attacks, attackers listen the communication as well as obtain the required information for doing malicious activities later. The attackers do not actively participate in the security attacks and do not disrupt the actual communication or system. Passive attacks do not give any harm for the system or communication rather it compromises confidentiality.

The passive attack is illustrated in the below diagram:
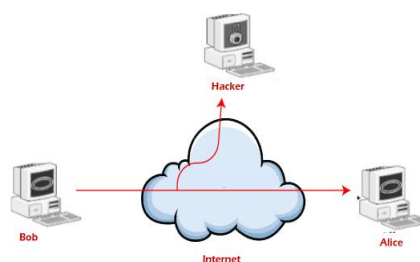
2.2 Active attacks

In active attacks, attacker actively participates in the security attacks. Attackers intercept the communication as well as connection and also modify the message that is being transmitted. Active attacks compromise the integrity because it modifies or alters the actual communication. It also compromises the availability. Active attacks include message modification, masquerade, replay, repudiations and DoS. The impacts of active attacks are harmful for communication or system as well as resources. Different security mechanisms for preventing active attacks are authentication, encryption as well as trust management [4]. This attack can be classified into the following:

2.2.1 Node compromise

Underwater sensor nodes may be put in unmanaged and even hostile marine environments in some specialised fields of application. Furthermore, because the network may have tens or hundreds of nodes deployed at huge proportions, it is impossible to assure the safety of all nodes. To access or modify data from memory, an attacker can capture, crack, and compromise nodes. Worse, the infected nodes could be injecting into the network as a legitimate node to monitor or disrupt, resulting in much more serious consequences [5].

2.2.2 Repudiation

Malicious nodes disclaim any involvement in a specific operation or communication with other nodes in repudiation attacks. A node in a communication denies having participated in all or part of the communication, regardless of whether the communication is malicious or not.

2.2.3 Routing

Routing attacks can prevent packets from reaching their intended destination node and potentially disrupt the operations of the network. Routing table overflow, poisoning, packet replication, and rushing assaults are examples of these types of attacks on routing protocols. Attackers can use these malicious behaviours to attract packets, analyse them, and even drop them at will. Routing attacks are frequently defended using cryptographic techniques [6]. However, because of the high computational complexity of encryption, it not only increases the size of communication messages but also causes increased energy consumption.

2.2.4 DoS attacks

DoS is a type of active attack that tries to prevent valid nodes from accessing resources. The attacker attempted to prohibit genuine nodes from using the network's services. DoS attacks can be carried out in a variety of ways, but they all result in the same issues.

DoS attacks are the most harmful and difficult to detect of these current attacks. To protect UWSNs from DoS attacks, the attack methods must be understood. DoS attacks can occur at any tier of the protocol stack and in a variety of methods. Even if UWSNs are well-protected by encryption, they are still vulnerable to DoS attacks, which can disturb communication and collaboration among nodes, as well as reduce the network's availability [7].

3. SECURITY REQUIREMENTS

The security requirements of UWSNs, being a subset of WSNs, are similar to those of terrestrial WSNs. However, due to the unique characteristics and limits of UWSNs, there are some unique security considerations.

3.1 Confidentiality

This refers to preventing unauthorised nodes from deducing the sensitive data's contents (e.g. security credentials and secret keys). Confidentiality extends not just to the user's information (e.g., strategic or tactical military information), but also to the MAC, routing information, and other systems. Malicious attackers should not be able to read or interfere with this sensitive information. Confidentiality can be ensured using a low-power encryption approach that is appropriate for UWSNs. The Cipher Text Stealing (CTS) approach [10] is a simple encryption method commonly employed in UWSNs.

3.2 Authentication

The acoustic channel is open; additionally, without the use of encryption, a malicious attacker can readily grab packets and manipulate their content. As a result, in order to filter malicious attacks, the receiving node must identify the source of the data. To access and share channel, services, applications, and data on that network, nodes must be authorised. To recognise abnormal behaviours and remove hostile nodes from the network, an intrusion detection and trust management system can be used[8]. These procedures ensure that only authorised nodes have access to the network's resources.

3.3 Integrity

Data integrity assures that data received is not updated, erased, or corrupted by unauthorised nodes throughout the transmission process, whether due to radio failure or malicious activity. This is especially important in situations like military operations and equipment controls, when little modifications could result in significant consequences. The Message Authentication Code (MAC) has been widely used in WSNs and UWSNs for data authentication because of its scalability, low latency, reliability, adaptability, and ease of implementation.

3.4 Availability

Availability means that the network is sufficiently strong. Even if certain nodes malfunction or the system is attacked, services will still be available. UWSNs can have availability if they use the right redundancy and self-adaptive strategies.

3.5 Self-stabilization

It assures that nodes can recover from attacks independently and without intervention in real time. Even if the attacker stayed in the network, if a node is self-stabilizing in the face of malicious attacks, it can recover to its regular condition on its own.

4. SECURITY ISSUES OF UWSN

Confidentiality, authenticity, integrity, and nonrepudiation are the main aims of cryptography and key management. Unauthorized users cannot read or modify sensitive information stored or sent in insecure networks like the underwater acoustic channel thanks to cryptography. Because of the peculiarities and limits of UWSNs, WSN encryption and key management procedures consume bandwidth and energy, making these primitives unsuitable for UWSNs with limited resources.

Internal and external intruders are detected, identified, and isolated from the network using intrusion detection technologies. Intrusion detection mechanisms, on the other hand, normally work after malicious assaults have taken place and been found. Malicious intruders are difficult to identify in the early stages of an attack. As a result, real-time detection systems must be investigated and enhanced. Intrusion tolerance techniques, on the other hand, can be used to safeguard networks while allowing malevolent intruders to exist, and are thought to be an effective security strategy.

5. PROBLEM OBJECTIVE

Energy efficiency and intrusion detection are the important thrust areas for under water wireless sensor networks. The research work proposes EDA cluster as well as exponential procedure methods to detection intrusions as well as improve the energy efficiency of underwater wireless sensor networks. The methodology will be evaluated using the simulation tool NS-2. The methodology is evaluated with evaluation measures namely death rate, packet delivery ratio and end-end delay.

6. UNSUPERVSED MACHINE LEARNING ALGORITH –EDA

An unlabelled dataset D, $|D|= N$, attributes set of dataset D is $X= \{a_1, a_2, \ldots a_k\}$. Select n data samples randomly from D, which is known as initial population POP_0. The degrees of attributes cannot be calculated in the dataset due to no labelling attached to the data samples. The sequence of attributes selection is random. The attributes a_i is chosen from X of POP_0 .a_i is treated based on its features. If a_i is the discrete attribute, a_i's different values count is t, $V(a_i) = \{v_{i1}, v_{i2,\ldots,}v_{it}\}$ where v_{ij} represents the value set of a_i. Probability of $V(a_i)$ is $P(V(a_i))$ which obeys multinomial distribution:

$$P\mid v_{i1}\mid=n1,\mid v_{i2}\mid=n_2,\ldots,\mid v_{it}\mid=n_t) = \frac{(n_1 + n_2 +\ldots+ n_t)!}{n_1!n_2!..n_t!}.P_{i1}^{n1}.P_{i2}^{n2}...P_{it}^{nt}$$

Let's say a researcher was given a big number of potential solution to an issue and wanted to come up with new and (hopefully) superior ones. He or she could choose one of several approaches to this. The goal is to find the probability distribution that will result in higher probabilities to solutions in areas where the greatest options are available. Once this was done, one might move on to the next step. To uncover fresh candidate solutions to the problem, sample this distribution [13].In an ideal world, repeated refinement of the probabilistic model based on representative samples of high quality solutions would increase the probability of generating the global optimum, and the procedure would find the global optimum or an accurate approximation of it after a reasonable number of iterations [9].

This fitness function assigns a numerical ranking to each string, with higher numbers indicating greater string quality. The selection operator selects a subset of the most promising solutions from this ranking population. Truncation selection with threshold = 50% is an example of a selection operator that selects the 50% best solutions. The method then builds a probabilistic model to try and estimate the probability distribution of the chosen solutions. New solutions are generated by sampling the distribution encoded by the model after it has been built. The previous population is then reintroduced to these new solutions (or replacing it entirely).

The technique is repeated until some termination criteria are met (typically when a satisfactory solution is found or when the number of iterations reaches a certain threshold), with each repetition of the procedure referred to as one generation of the EDA. The building of the model that aims to reflect the probability distribution of promising solutions is a crucial step that distinguishes EDAs from many other meta-heuristics [12]. This is not an easy process because the goal is to reflect a more generic distribution that captures the properties of the selected solutions that make them better than other candidate solutions, rather than to perfectly represent the population of potential alternatives. In addition, we must ensure that the model can be constructed and sampled quickly.

EDA algorithm

```
g <----0

generate initial population P (0)

while (not done)

do

select population of the promising solutions S(g)

from P(g)

build the probabilistic model M(g) from S(g)

Sample M(g) for generating new candidate solutions

O(g)

Incorporate O(g) into P(g)

G←g+1

End while
```

7. PROBLEM STATEMENT

Energy efficiency and intrusion detection are the important thrust areas for under water wireless sensor networks [10]. Energy efficient intrusion detection system is the thrust area in under water wireless sensor networks.

8. METHODOLOGY

8.1 Flow Diagram
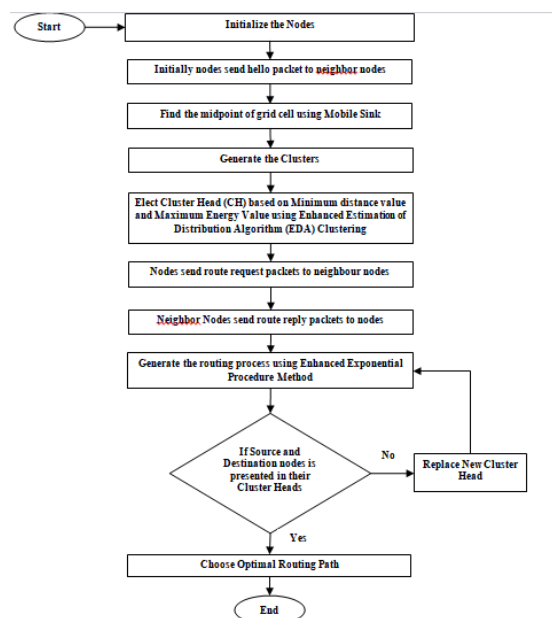
8.1.1 Initialization Phase

Initialize the input values and EDA Cluster and Exponential Procedure method function for the nodes of a designed area.

8.1.2 Cluster Creation Phase

In EDA Cluster Method, Initially generate the x and y co-ordinate value of the nodes. Initiate the

166

Sink node. Sink node is divided into no of grid cells. The grid cell takes the midpoint to form a group of the nodes for grid formation. Create the grid formation to assign a grid id and node id to all nodes. Grids are square in form and static in nature in the network area. It may be a mobile or static during the collection of information at the sink. Among all the other nodes in the grid cell, the midpoint of the grid cell is initially chosen as a cluster head (CH). CH is chosen as the node that has a minimum distance to the midpoint. Node-id and grid ids are allocated to each node in the network.

8.1.3 Path Updating Phase

In Exponential procedure method, Initialize the source node (S) and distance node (D) value. The source node and destination nodes check the present Cluster head of the group nodes. Source node is not up to destination node. After sending the results, the algorithm is implemented to verify the cluster group head. The source input, the source node's value, includes the CH. Finally, to the Destination Node, CH is sent.

Estimation of Distribution Algorithm (EDA) clustering

Step 1: Procedure EDA Cluster

Step 2: Define a cell representation for each data item

Step 3: Initialize the divide a cells of midpoint value

Step 4: Initialize the probability model with a uniform distribution of attribute intervals

Step 5: Then divide a cells allocate cell id and midpoint of the cell

Step 6: Create a population of N individuals each indicating K cluster centers

Step 7: Repeat for G generations

       For each individual X

       For each centerX_k

             Build cluster (Xk);

             Calculate fitness value for X

             Select N/2 individuals by tournament selection

Step 8: Update the probability model considering the selected individuals

Step 9: Generate a new population of individuals according to the updated probability model

Step 10: Add the best individual of the previous population to the current one

- **Procedure for build Cluster**

Step 1: Procedure build cluster (centerXk)

Step 2: Initialize the maximum distance value

Step 3: R ←1 (distance in number of cells between a data item and the cluster center)

Step 4: find a node to node distance value

Step 5: If (distance value > = maximum distance value)

Store the distance value and check minimum energy value

Step 6: Else

Increase the distance value to maximum value node elects the cluster process

Step 7: Create a cluster Cxk containing all data items within a maximum distance R from the centerXk

Step 8: Calculate the density (Cxk) for cluster Cxk

Step 9: Current Density = density (Cxk)

Step 10: Repeat until the density decrease is above a threshold

R ← R + 1

For each data item I

Dxki ← distance to the centerXk

If Dxki <= R

Add I to Cxk

Calculate a new density (Cxk) for the cluster Cxk

If (1 – (density (Cxk) / current density)) >= threshold then

Remove all recently added items for Cxk

Else

Current density = density (Cxk)

Exponential Procedure Algorithm

Step 1: Initially include the source and destination

Step 2: Input source node s, destination node d

Step 3: Initialize for loop and then include the all nodes after all nodes are check source

node

For (j=0; j<n; j++)

Step 4: After matching the source node compare to the grid id and node id based on cluster group

Step 5: Find source node present the group of cluster then,

IF (s! =d && s = = CH) THEN

Find the CH present the head stored in array of the source node

If (n [j, 2] = = src)

Ac [src] =ch [n [j, 1]]

Cluster message send from source s to CH

Step 6: IF (d = = CH1 && CH! =CH1) THEN

If (n [j, 2] = = dst)

Ac [dst] =ch [n [j, 1]]

Cluster message send from CH to CH1

CH1 send cluster message to destination d

End if

End if

END IF

END IF

9. EXPERIMENTAL SETUP

Simulation tools, underwater protocol stack, channel simulators and OS and are significant and crucial software-based components. Commercial, tailored simulation software produced by researchers, as well as other free open source networking software, are among the simulation tools. Many well-known networking software, such as Omnet++, OPNET, and others, have been developed by researchers up to this point.

NS-2, can be used to model/simulate underwater sensor networks. Among these, ns-2 is the most well-known open source discrete-event simulator, providing networks,,a large volume of protocols, and application models to more than 600 universities spread across 50 nations. For this research work, simulations have been conducted with NS-2.

9.1 Evaluation measures

9.1.1 Packet delivery ratio

The PDR is defined as the ratio of total packets delivered to total packets sent from a source node to a destination node in a network. The maximum number of data packets should be delivered to the destination.

9.1.2 End – end delay

The time required for a packet to get from point A to point B is referred to as network delay. Te number of links multiplied by the overall latency in the network gives the delay.

9.1.3 Energy consumption

In wireless sensor networks, energy is a finite resource. During transmit, receive, and idle modes, energy is spent in a wireless sensor network.

10. Experimental Results

Table 1.Death Rate

Time (S)	Death rate (Death rate of sensor node/s)
600	0.95
700	1.01
1000	1.18781
1500	1.23845
1900	1.35428
2000	1.85254

Table 2. Packet Delivery Ratio

No of nodes	Packet Delivery Ratio (Pdr/s)
10	1.80
50	2.1
90	2.18752
130	2.25684
170	2.35875
210	2.40325
250	2.44872

Table 3. Average End-to-End delay

No of nodes	Average End-to-End delay (Delay/s)
50	2.32451
90	2.28546
130	2.24532
170	1.96584
210	1.88234
250	1.77124

Table 4. Energy Consumption

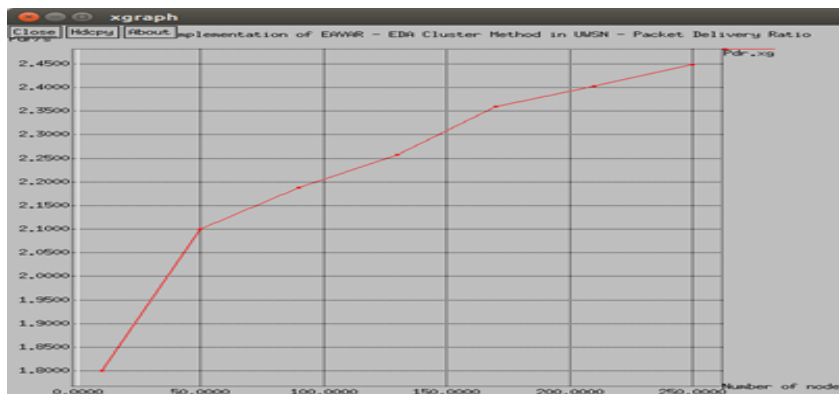| No of nodes	Energy Consumption (J/s)
50	500.54
90	625.240
130	700.524
170	780.925
210	845.124
250	960.452
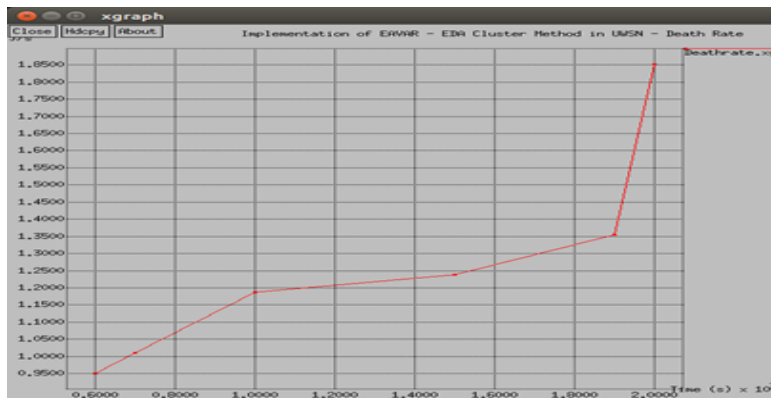
Fig. 2. Death Rate
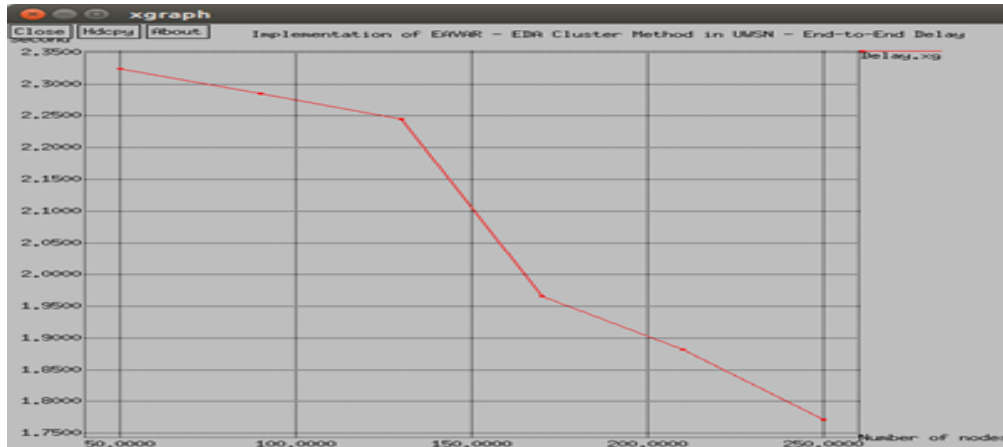

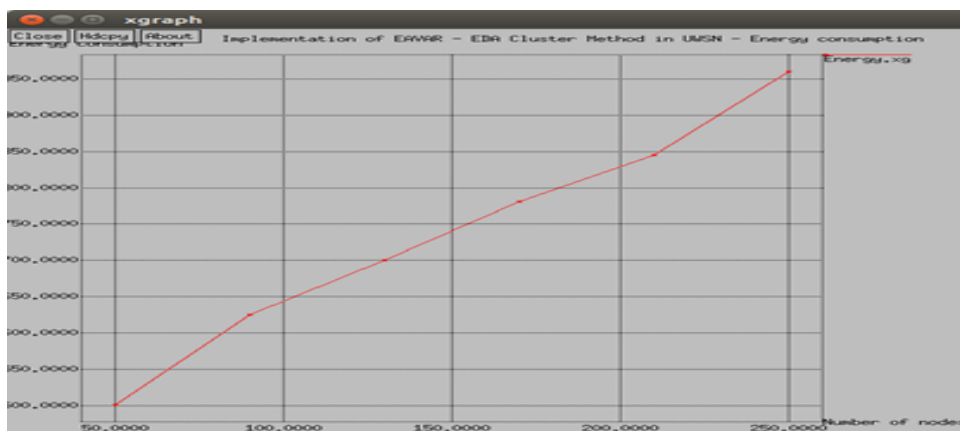*Fig. 3. Packet Delivery Ratio*

Fig. 5. Average End-to-End delay

Fig. 6. Energy Consumption

Table 5. Comparison of Phase 3 and Phase 4

Performance Metric	Death rate (Death rate of sensor node/s)	
Time (S)	EAVAR - LPA	EAVAR - EDA
600	1.00	0.95
700	1.05	1.01
1000	1.30903	1.18781
1500	1.31009	1.23845
1900	1.51007	1.35428
2000	1.99017	1.85254
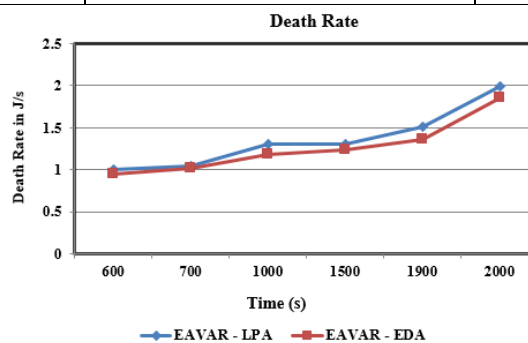
Fig.6. Death Rate

Table 6. Packet Delivery Ratio (PDR)

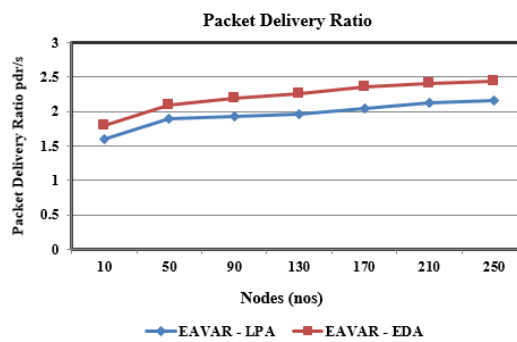| Performance Metric	Packet Delivery Ratio (Pdr/s)	
No of nodes	EAVAR - LPA	EAVAR - EDA
10	1.6	1.80
50	1.9	2.1
90	1.93452	2.18752
130	1.96281	2.25684
170	2.05182	2.35875
210	2.12431	2.40325
250	2.15291	2.44872

Fig. 7. Packet Delivery Ratio (PDR)

Table 7. Average End to End Delay

Performance Metric	Average End-to-End delay (Delay/s)	
No of nodes	EAVAR - LPA	EAVAR - EDA
50	2.42587	2.32451
90	2.40853	2.28546
130	2.38457	2.24532
170	2.01658	1.96584
210	1.95368	1.88234
250	1.85428	1.77124
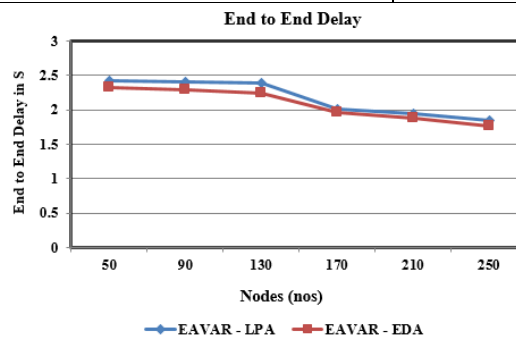
Fig. 8. Average End to End Delay

Table 8. Energy Consumption

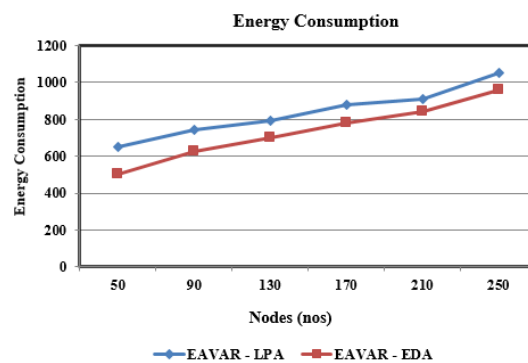| Performance Metrics	Energy Consumption (J/s)	
No of nodes	EAVAR - LPA	EAVAR - EDA
50	650.57	500.54
90	740.352	625.240
130	790.542	700.524
170	876.258	780.925
210	910.578	845.124
250	1050.258	960.452

Fig. 9. Energy Consumption

The experimental results depicts that, the EAVAR – EDA algorithms outperforms than EAVAR-LPA on energy consumption, delay, death rate and delivery ratio.

11. Conclusion

UWSNs are prone to a variety of security threats including malicious assaults that disrupt the network's communication and cooperation. The security standards of UWSNs are introduced to prevent these attacks. Finally, we'll go through several specific security technologies and approaches. Due to the unique characteristics and limits of UWANs, securing them is difficult. Furthermore, different applications may have varied security requirements, and excessive security systems would be a significant energy drain. As a result, the research work focuses on methodology to detect intrusions in the network with effective energy management. The results illustrates that, the proposed methodology EAVAR-EDA outperforms than EAVAR-LPA on various evaluation measures such as energy consumption, packet delivery ration, death rate.

REFERENCES

[1]. Davis A, Chang H. Underwater wireless sensor networks. In2012 Oceans 2012 Oct 14 (pp. 1-5). IEEE.

[2]. Han G, Jiang J, Bao N, Wan L, Guizani M. Routing protocols for underwater wireless sensor networks. IEEE Communications Magazine. 2015 Nov 9;53(11):72-8.

[3]. Gkikopouli A, Nikolakopoulos G, Manesis S. A survey on underwater wireless sensor networks and applications. In2012 20th Mediterranean conference on control & automation (MED) 2012 Jul 3 (pp. 1147-1154). IEEE.

[4]. Domingo MC, Prior R. Energy analysis of routing protocols for underwater wireless sensor networks. Computer communications. 2008 Apr 18;31(6):1227-38.

[5]. Coutinho RW, Boukerche A, Vieira LF, Loureiro AA. Underwater wireless sensor networks: A new challenge for topology control–based systems. ACM Computing Surveys (CSUR). 2018 Jan 4;51(1):1-36.

[6]. Ahmad B, Jian W, Enam RN, Abbas A. Classification of DoS attacks in smart underwater wireless sensor network. Wireless Personal Communications. 2021 Jan;116(2):1055-69.

[7] Ahmad B, Jian W, Enam RN, Abbas A. Classification of DoS attacks in smart underwater wireless sensor network. Wireless Personal Communications. 2021 Jan;116(2):1055-69.

[8] Yang G, Dai L, Wei Z. Challenges, threats, security issues and new trends of underwater wireless sensor networks. Sensors. 2018 Nov;18(11):3907.

[9]. Ahmed MR, Tahsien SM, Aseeri M, Kaiser MS. Malicious attack detection in underwater wireless sensor network. In2015 IEEE International Conference on Telecommunications and Photonics (ICTP) 2015 Dec 26 (pp. 1-5). IEEE.

[10]. Domingo MC. Securing underwater wireless communication networks. IEEE Wireless Communications. 2011 Feb 17;18(1):22-8.

[11]. Jiang J, Han G, Zhu C, Chan S, Rodrigues JJ. A trust cloud model for underwater wireless sensor networks. IEEE Communications Magazine. 2017 Mar 13;55(3):110-6.

[12]. Fan J, Liang Y, Xu Q, Jia R, Cui Z. EDA-USL: unsupervised clustering algorithm based on estimation of distribution algorithm. International Journal of Wireless and Mobile Computing. 2011 Jan 1;5(1):88-97.

[13]. Yang Q, Chen WN, Li Y, Chen CP, Xu XM, Zhang J. Multimodal estimation of distribution algorithms. IEEE transactions on cybernetics. 2016 Feb 15;47(3):636-50.

PUBLISHED BY

DEPARTMENT OF COMPUTER SCIENCE

BISHOP HEBER COLLEGE (AUTONOMOUS)

Reaccredited at the 'A' by NAAC with a CGPA of 3.58 out of 4

Recognized by UGS as the "COLLEGE OF EXCELLENCE"

TIRUCHIRAPPALLI -620 017,

TAMIL NADU.

9 789356 271890