ISSN 2521-0122 (Online)
ISSN 2519-7991 (Print)
Vol. 5 Issue: 1
January - March, 2021

LGURJCSIT

Lahore Garrison University

Research Journal of Computer Science and Information Technology



www.lgurjcsit.lgu.edu.pk

Lahore Garrison University,
Lahore Pakistan

SCOPE OF THE JOURNAL

The LGURJCSIT is an innovative forum for researchers, scientists and engineers in all domains of computer science and technology to publish high quality, refereed papers. The journal offers articles, survey and review from experts in the field, enhancing insight and understanding of the current trends and state of the art modern technology.

Coverage of the journal includes algorithm and computational complexity, distributed and grid computing, computer architecture and high performance, data communication and networks, pattern recognition and image processing, artificial intelligence, cloud computing, VHDL along with emerging domains like Quantum Computing, IoT, Data Sciences, Cognitive Sciences, Vehicular Automation. Subjective regime is not limited to aforementioned areas, Journal policy is to welcome emerging research trends in the general domain of computer science and technology.

SUBMISSION OF ARTICLES

We invite articles with high quality research for publication in all areas of engineering, science and technology. All the manuscripts submitted for publication are first peer reviewed to make sure they are original, relevant and readable. Manuscripts should be submitted via email only.

To submit manuscripts by email with attach file is strongly encouraged, provided that the text, tables, and figures are included in a single Microsoft Word/Pdf file. Submission guidelines along with official format is available on the following link; www.lgurjcsit.lgu.edu.pk

Contact: For all inquiries, regarding call for papers, submission of research articles and correspondence, kindly contact at this address:
LGURJCSIT, Sector C, DHA Phase-VI Lahore, Pakistan
Phone: +92- 042-37181823
Email: lgurjcsit@lgu.edu.pk

Copyright @ 2021, Lahore Garrison University, Lahore, Pakistan. All rights reserved.
Published by: Faculty of Computer Science, Lahore Garrison University

LGURJCSIT

Volume 5 **Issue 1** **January - March 2021**

Qurat ul Ain, Adnan Khalid, Aziz Guergachi, Muhammad Shahbaz, Awais Qasim, Adeel Munawar

Cloud Computing: Needs Enabling Data Mining and Business Intelligent Applications 1-8

Muhammad Amjad Khan, Taimoor Hassan, Zaka Ullah

Information Security for Cloud using Image Steganography 9-18

Haseeb Aslam Rana, Aftab Ahmad Malik, Naureen Riaz, Mujtaba Asad

Analysis of Advantages and Problems in Teaching and Assessment with Online System during Covid-19 19-28

Tanveer Aslam, Salman Qadri, Najia Saher, Syed Ali Nawaz, Muhammad Shehzad, Syed Furqan Qadri, Abdul Razzaq, Faisal Shahzad, Dost Muhammad Khan.

A Spatial Model of K-Nearest Neighbors for Classification of Cotton (Gossypium) Varieties based on Image Segmentation 29-40

Ehsan Elahi, Bilal Shoaib, Muhammad Akmal

Time Dependent Popularity Caching Scheme for NDN Based MANETs 41-46

Zulfiqar Ahmad Noor, Muhammad Khawar Bashir, Yasir Saleem, Zeeshan Ahmad, Muzammil Hussian, Nadia Tabassum

Trading Algorithm Model Based on Technical Indicators 47-58

Syed Shah Muhammad, Salman Qadri, Hasnain Ahmed, Syed Furqan Qadri, Najia Saher, Sarfraz Ahmad Tanveer Aslam , Syed Ali Nawaz and Muhammad Tariq Pervaiz

Traffic Intensity Based Energy Efficiency Architecture for Data-Centers 59-70

Muhammad Waseem, Qaiser Abbas, Muhammad Ilyas, Saad Razzaq, Sadia Niazi, Ijaz Asghar

Classifying Urdu Verbs Using Rule Based Approach 71-78

Binish Naz , Muhammad Rizwan, Fahad Ahmad, Aysha Shabbir

Configuration-Free Systems for WiFi Sensing based Smart Home using the Smart Remote Controller 79-86

Patron in Chief

Major General Shahzad Sikander, HI(M) (Retd) Vice Chancellor Lahore Garrison University

MEMBERS OF ADVISORY BOARD

Dr. VasakiPonnusamy	UniversityTunku Abdul Rahman, Malaysia
Dr. Rehan Akbar	UniversityTunku Abdul Rahman, Malaysia.
Dr. Amir Haider	Sejong University, Seoul, South Korea.
Dr. Atta-ur-Rahman	Imam Abdulrahman Bin Faisal University (IAU), Dammam, Saudi Arabia.
Dr. Rizwan Ali Naqvi	Sejong University, Seoul, 05006 Korea.
Dr. Loc Nguyen	Loc Nguyen's Academic Network,China
Dr. Yin Zhang, Zhongnan	University of Economics and Law, China.
Dr. Baha Ihnaini	College of Science and Technology, Wenzhou Kean University, China
Dr. M. Abdul Basit Ur Rahim	Montana Tech, United States.
Dr. Muhammad Aamir	University of Derby, United Kingdom.
Dr. Abeo Timothy	Ghana, West Africa
Dr. Fahad Ahmad	Jouf University, Sakaka, Saudi Arabia

MEMBERS OF EDITORIAL BOARD

Dr. Khalid Masood	*Lahore Garrison University Lahore, Pakistan*
Dr. Muhammad Asif	*Lahore Garrison University Lahore, Pakistan*
Dr. Sageer Abbas	*NCBA&E, Lahore, Pakistan*
Dr. Muhammad Faheem Mushtaq	*Khawaja Fareed UEIT Rahim Yar Khan, Pakistan*
Dr. Muhammad Sohaib	*Lahore Garrison University Lahore, Pakistan*
Dr. Areej Fatima	*Lahore Garrison University Lahore, Pakistan*
Dr. Ghulam Ali	*University of Okara, Pakistan*

Editor-in-Chief

Dr. Shazia Saqib — *Lahore Garrison University Lahore*

Managing Editor

Dr. Arfan Ali Nagra — *Lahore Garrison University Lahore*

Publication Editor

Dr. Tahir Alyas — *Lahore Garrison University Lahore*

Associate Editors

Mr. Umer Farooq — *Lahore Garrison University Lahore*

Mr. Waqar Azeem — *Lahore Garrison University Lahore*

Assistant Editors

Mr. Muhammad Nadeem Ali — *Lahore Garrison University Lahore*

Ms. Aqsa Iftikhar — *Lahore Garrison University Lahore*

Sub-Editors

Mr. M. Asif Saleem — *Lahore Garrison University Lahore*

Mr. Abdul Rehman — *Lahore Garrison University Lahore*



Munawar et al LGURJCSIT 2021

ISSN: 2521-0122 (Online)
ISSN: 2519-7991(Print)

LGU Research Journal of
Computer Science and IT

Vol.5 Issue1, Jan-Mar 2021

Cloud Computing: Needs Enabling Data Mining and Business Intelligent Applications

Qurat ul ain[1], Adnan Khalid[2], Aziz Guergachi[3], Muhammad Shahbaz[4], Awais Qasim[2], Adeel Munawar[1]
[1,2]Department of Computer Science, GC University, Lahore, Pakistan
[3]Ted Roger's School of Management, Ryerson University Toronto, Canada
[4]Department of Computer Science, UET, Lahore, Pakistan
Adeelmunawar43@gmail.com

ABSTRACT

As a new computational paradigm, cloud computing is attracting a lot of interest from researchers in the field of the business community and information technology sector and it can integrate with several heterogeneous resources makes it distinguished and unique to fulfill the demands of different types of users. The rapid increase in data volume and fixed access to online resources, which is related to all departments need to mine data for the discovery of knowledge. Its principal peculiarities incorporate a versatile asset design and along this line a suitable system for tending to be comprehended in an ideal mode. From the specific situations where cloud computing can be integrated, its use in business information and intelligence also conveys the highest aspirations from data mining to updates. This study gives an outline of the recent condition of the arrangement of Cloud Computing and elaborates, its implications in Business Intelligence and Data Mining. This study defines multiple layers that are expected to create such a framework in distinctive levels of deliberation, from the fundamental equipment stages to the product assets accessible to actualize the applications. At the end of this study, a few cases related to Data mining methodologies have been relocated to the Cloud Computing paradigm.

KEYWORDS: Business Intelligence (BI), MapReduce, Data Mining (DM), Hadoop, Hash function, Service-oriented Architecture (SOA), Service Line Agreement (SLA), Cloud Computing.

1. INTRODUCTION

Cloud computing is an example of modern computing but it is not a desirable idea based on grid computing. There are two similar scenarios for reducing computational costs and increasing framework adaptation. There are also differences in the way work is done. In a computational grid, a wide range of jobs are performed on different machines, separated into several smaller parts, and a comparative office for the computing. Again, through cloud computing, it is planned to allow the client to deprive the various administrations of resources in the basic building design and thus it is not prohibited and from the web, facility to direct word can offer different administration. [1].

Cloud computing is a service-oriented architecture that allows developers to reduce the complexity of computing multiple messages, including application coordination, exchange management, and security arrangements while allowing multiple steps and conventions [2].

The Cloud Computing paradigm provides infrastructure facilities such as hardware, data storage servers, and networks. These administrations provide facilities to the client without obliging them to recognize the area and different points of interest in the computing structure [3]. Now with the advancement in the latest technologies Cloud computing is fruitful for Business Intelligence (BI) applications [4-5].

In every organization data is an important asset business analysts use this data for the decision-making process that informs the organization executives and senior managers with current trends [6]. Data Mining (DM), information visualization, and data warehouse technologies are used to store historical data that helps business analysts in the decision-making process [7].

Managing and processing a huge quantity of data from an organization for a little period require great funding. As expressed above, with these requirements another test for the exploration group turns out with the need to adjust the frameworks to a Cloud Computing construction modeling [8]. The primary point is to reduce data management cost and needs an efficient management system that answers multiple quires and make a wide search environment. Organizations require parallelizing the effort and

enabling fault tolerance in huge data storage as well as wants reduced computational time. Organizations face difficulties in the management of data and applying Data Mining (DM) and Business Intelligence (BI) applications for knowledge discovery in data. In particular, the cloud computing paradigm provides a standard infrastructure with different levels that must be considered to execute the administration of the information, an analogous execution system that answers a query with reduced time [4].

In the case of Business Intelligence (BI) cloud computing offers great benefits with Data Mining (DM) processes and cloud virtualization [5]. It is highly expensive to parallelized high computational processes, for instance, those approaches are based on transformative models [9]–[11]. MapReduce and Hadoop are the latest two programing models used in a cloud environment that provides a platform for developers to run their algorithms.

Both frameworks have been outlined with two critical limitations: in the first place, mists have expected nods within the cloud are co-spotted, i.e. inside one server farm, on the other hand, that there is generally little data transmission accessible between the topographically conveyed groups containing the information [12]. Second, these mists have accepted that individual inputs and yields to the cloud are generally little, even though the total information oversaw and transformed are vast.

Data Mining (DM) applications are needed to apply in cloud computing because of the requirement of data management and knowledge discovery from a huge volume of data in business, science, and engineering [11], [13]. Data Mining (DM) in the cloud paradigm helps in mining knowledge from unstructured or semi-structured internet resource data. In the field of business, it is much needed for organizations to centralize the management of data, and ensure reliable, secure services for their users. Companies don't need to spend funds on the hardware infrastructure to apply data mining, they can apply mining algorithms through the browser and pay for the resources that they use.

In the most recent years, numerous Data Mining (DM) algorithms move to a cloud computing model for productive results. Day by day increase in the data volume needs high computational power for mining algorithms. This study will show a few current recommendations that adjust standard processes to the cloud for building them to be highly proficient and versatile. Developing a variety of uses in genuine building issues, for example, PC vision [4], proposal frameworks [4], or Healthcare frameworks [14] demonstrate the high essentialness of this methodology.

This study organized into sections, Section 2 presents the primary ideas on Cloud Computing, comprising its structure and fundamental layers. Area 3 and 4 introduce the structural planning ways to create Business Intelligence arrangements on a new computing paradigm. The encoding models for executing Data Mining procedures inside this ideal model, along with a few cases, are demonstrated in Section 5. At long last, the fundamental closing comments are given in the last section.

2. BASIC ASPECTS REGARDING CLOUD COMPUTING

Cloud computing provides a variety of services on demand over the internet with dynamic and elastic resource allocation in a multitier environment. Service-Oriented Architecture (SOA) is a blend of technological architecture that support and incorporate all kind of services [4]. The computing framework provides services on demand and is organized in a way that it can be continuously provided to the client.

Cloud computing consists of pay per user business model where resources are dynamically allocated. The taxonomy of cloud computing can be classified as i) homogeneous clouds: having the same physical and virtual architecture. ii) Heterogeneous clouds: having heterogeneity in basic, virtual or physical architecture, e.g. different types of OS or different ways of execution and access of the same application. iii) Hybrid clouds: is the combination of grid and cloud computing. Redundancy, the elasticity of resources, low time of execution, high availability, efficiency, and low cost are some of the well know features of cloud computing. The characteristics of cloud computing architecture increase the scalability and reliability, where an application running on its response in real-time instead of high resource demand, a huge quantity of customers, and an increasing volume of data. The main characteristics of Cloud computing are its loose coupling, high inter-operatively, and the facility of interfaces that separate the services from the implementation and the platform. Cloud computing comprises services-oriented layers such as Infrastructure-as-a-service (IaaS), Platform-as-a-service, and software-as-a-service (SaaS) that may have a different organizational structure.

2.1. Infrastructure-as-a-Service (IaaS) Bottom Layer of Cloud computing

IaaS is comprised of hardware services such as storage, computation power, server, and network with the facility of intermitted software. Client access computational resources on rent as a replacement for buying and installing their hardware and data centers. IaaS is billed for these services based on their actual usage. It is beneficial for the client, who uses what he needs and pay according to the usage. The dynamic resource allocation allows the client to enhance their resource demand according to low or high workload.

IaaS offers a variety of large infrastructure as a service to scientific researchers and technicians who need high computation power to run their tests and analysis of data. It also facilitates with the data storage store huge volume of data client can access any time when they need.

2.2. Platform as a Service (PaaS) Middle Layer of Cloud Computing

Platform as a Service (PaaS) is an integrated set of software with all necessary stuff that a developer needs during application development and execution. The developer does not directly access the infrastructure, the PaaS service provider offer tools that a designer might need for IaaS service [15]. PaaS is a set of independent tools that allow the developer to develop components of the application as well as the application themselves. It reduces the development cost developer no need to buy their software tool kits for making a development environment.

2.3. Software as a Service (SaaS) Upper Layer of Cloud Computing

This layer offers software as a service. SaaS providers offer software programs to millions of users and it was the first implantation of cloud services. The client gets benefits and reduces the cost, needs not to worry about the management of software. The third-party is responsible for updating, installing, maintaining, patching, and licensing the software.

3. CHALLENGES AND LIMITATIONS IN CLOUD COMPUTING

Clouding computing is a highly adopted technology in the last few years due to its characteristics Such as on-demand service, elasticity, a huge pool of resources, and a pay-per-use business model. However, it provides a variety of services but there are some limitations and challenges that need to be tackle.

3.1. Security and privacy: Data is a vital asset in any organization and plays an essential role. Data Security and privacy is the most wanted need for an organization since in cloud computing data can be distributed on multiple hardware servers managed by the third-party customer who has no control over the data. Security needs require computation of encoded data and a physically secure environment for data servers. The service provider needs to increase user confidence levels.

3.2. Service Availability and recovery: Service providers must ensure 24 hours of services with the guarantee and use efficient systems and resolve bandwidth problems.

3.3. Loss of governance: The cloud providers have overall control in services which is also a security gap that may lead to confidentiality and integrity problems.

3.4. Resource Scalability: Cloud computing technology attracts a huge number of clients, changing demand of users requires an efficient resources scheduler and monitoring system. It is necessary to reduce energy consumption.

4. BUSINESS INTELLIGENCE (BI) PROCESSES IN CLOUD COMPUTING

The business intelligence system needs to analyze the huge amounts of data in a very short period usually it is not more than a few minutes. Research shows enterprises are the owner of large volumes of data and it is essential to get the knowledge to predict future trends. This bulk volume of data may consequence of diverse applications, for example, forecast of advance concessions and credit strategies to customers because of dangers, characterization of a client according to their market demand, item proposals, and mining of examples from business exchanges between others.

In the rest of this area, first, present structural planning to create BI arrangements on a Cloud Computing stage. At that point, we will push the decency on the utilization of Cloud Computing concerning other comparable innovations.

4.1. Cloud Computing Environment for Business Intelligence (BI):

It is an essential business community to predict future trends that are hidden in a large amount of data. In [4] researchers reconsidered the essential model and procedure for investigating structured and semi-structured in a business warehouse. Cloud-based data management architecture for organizations is consists of four layers however three are the most common layer in cloud computing architecture [5]. The first three levels are regular for both methodologies, while the previous one is particularly outlined for information warehousing. The portrayal of these segments is counted underneath:

- The First level takes after the structure presented in the above section, that is, it incorporates a system construction modeling with numerous approximately coupled PC nodes for giving decent scalability and an adaptation to the internal failure plan. As recommended, the framework must consider an element/ versatile plan such as the execution, expense as well as energy utilization of each node machines is overseen at run-time.

- The second level is dedicated to the storage of parallel data, based on the storage and recovery of key sets (instead of the social model in light of outside key/essential key connections). A few samples of frameworks that are streamlined for this design are Google BigTable [16] or Sector [7].

➢ This level is the implementation environment. Because of the expansive number of nodes, Distributed computing is particularly pertinent for distributed computing assignments taking a shot at rudimentary operations. The most known case of a distributed computing execution environment is most likely Google MapReduce [17] and its open-source variant Hadoop [18],[19]. However, different projects can be found as doable options [20], [21]. All these situations go for giving flexibility by permitting to arrange assets as indicated by the application, taking care of blunders straightforwardly, and guaranteeing scalability.

➢ The final level is the high-level querying language, which is situated towards OLAP. This level implicates as an interface to the client and it gives straightforwardness to alternate levels of the building design. A few query languages have been proposed like Map-Reduce-Merge [17], [22].

4.2. Data Mining (DM) and Business Intelligence (BI) in Cloud Computing

Generally, large-scale organizations have a huge volume of data, which essentials to be processed in a small-time, at a fast speed. For this purpose, grid computing is considered the best solution to reduce the cost while enhancing the flexibility and overall power of the system. The similitudes with Cloud Computing are apparent since both have the same characteristics such as approximately coupled, heterogeneous, and geologically scattered nodes. On the other hand, the fundamental distinction is the ways errands are processed in every particular setting. In a computational grid, one expansive occupation is partitioned into numerous little divides and executed on different machines, offering a comparable office for computing force. Then again, the computing cloud is proposed to permit the client to get different administrations without putting resources into the basic structural planning and accordingly is not all that prohibitive and can offer a wide range of administrations, from web facilitating, directly managed by a third party [4].

Also, the preferences of this new computational paradigm concerning other contending innovations are rich. First and foremost, Cloud application suppliers endeavor to give the same or better administration and execution as though the product projects were introduced provincially on end-client PCs, so the clients don't have to burn through cash purchasing a complete equipment hardware infrastructure to utilized, i.e. a basic portable gadget is sufficient to run the projects on the Cloud [4], [23].

Another advantage of, this kind of model for the data storage and the computing plans permits organizations to access their applications up and running quicker, with a lesser need of support from the IT office since it naturally deals with the business request by relegating pretty much IT assets (servers, stockpiling and/or systems administration) contingent upon the computational load progressively [24]. At last, this natural versatility of this framework makes the charging of the infrastructure to be carried out as indicated by the previous certainty.

5. DATA MINING (DM) WITHIN THE CLOUD COMPUTING PARADIGM

This part of the study focusing on indicating the guaranteeing future prediction for the Distributed computing paradigm in regards to the execution of DM algorithms in place to manage vast data storage for which it has been restrictively extravagant until this minute.

The ideas behind all the recommendations in this study consistently disperse the completeness of data across all nodes in the cloud and also exchange a small amount of information thus making applications exceptionally adaptable and efficient. It would be prudent, however, to maintain consistency and data protection [25].

It is important to change the data put away into the multidimensional array to "Pig data" [26] to have the capacity to do online investigation process by utilizing the MapReduce/Hadoop technology or related methodologies, (for example, Segment/Sphere [26]), likewise decreasing the stockpiling expenses [7]. In this programming model, clients determine the reckoning regarding a guide and a lessened capacity, and the basic runtime framework naturally parallelizes the processing crosswise over huge bunches of machines, machine disappointments, and timetables between machine correspondence to make productive utilization of the disks, network.

In MapReduce programing model Map, composed by the client, takes data as input and generates intermediate value pairs [27]. The MapReduce library bunches together all transitional qualities related to the same moderate key I and forward to the diminished capacity.

Then Reduce function takes an intermediate Key I and its related set of values. It blends these qualities to structure a conceivably littler arrangement of qualities [19]. The middle qualities are supplied to the client's decrease capacity through an iterator. This permits us to handle arrangements of qualities that are so vast it couldn't be possible to fit in memory.

In MapReduce, Map function summons is divided crosswise over different machines by mechanically dividing the data into an arrangement of M parts. Data divided into splits are processed parallel by all machines in the cluster [22]. In this programming model, the Reduce function divides the intermediate key into r parts with the help of partitioning hash modulus R function. The number of partitions (R)

and the parceling capacity are indicated by the client (see figure 1)

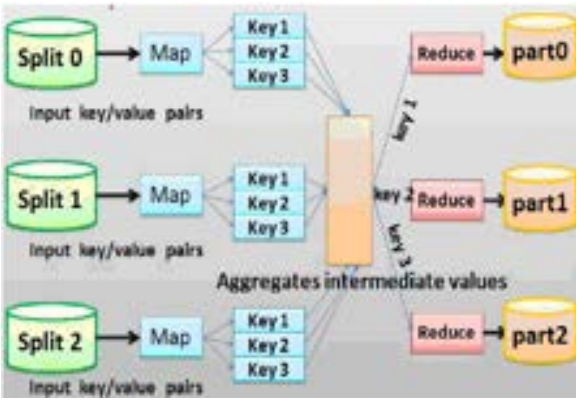

Figure 1: Map-Reduce architecture [27]

To sum things up, these frameworks are arranged to disseminate the data groups besides the cloud and to disperse the processing among the bunches, i.e. as opposed to moving all data among the machines, they characterize mapping functions to make go-between tuples of <key, value> and with the help of the reduced functions lesser the processing time. As a sample, in a system that went for numbering the number of events of an individual word within a vast gathering of reports, the map function will emanate every expression in addition to a related check of events though the diminished capacity totals together all checks radiated for a specific word.

Even though it is a generally new structure, numerous DM algorithms have been present effectively actualized after the rules of this programming model. Case in point, in [16]the creators display an arrangement model which tries to discover middle of the road model in the middle of Bayes and K-closest neighbor procedures by utilizing a positive sort of subtree express to every group which is gotten by bunching on preparing set by applying a base spreading over tree MapReduce execution, and after that perform the characterization utilizing thought like KNN [25].

An alternate approach that is produced utilizing the MapReduce model goes for tending to the dynamic successive example mining issue [22], which inherently experiences the scalability issue. Two Map/Reduce occupations are planned; the hopeful computing occupation registers hopeful successive examples of all arrangements and upgrades the outline of every arrangement for the future reckoning. At that point, utilizing all competitor consecutive examples as information, the bolster collecting employment gathers the event frequencies of competitor successive examples in the present time of investment and reports successive consecutive examples to clients.

Decision Tree Algorithm presents a test investigation utilizing an arbitrary Tree Algorithm under a distributed computing environment by considering two diverse plans keeping in mind the end goal to execute the parallelization of the learning stage [11], [16]. The primary methodology was that every node creates classifiers with its local data simultaneously and all classifiers are accounted for to a focal node. At that point, the focal node will utilize all classifiers together to do expectations. In this model, each node performs a smaller part of the task that may relate one or more classifiers and send its results to the focal node.

Different works are because of diverse distributed computing situations, a Particle Group Optimization was intended for Amazon Cloud, where the hopeful arrangements are displayed by the arrangement of errand administration sets, having every molecule to gain from diverse models, in any case, to take in the other practical sets for distinctive measurements. The valuable position building strategy ensures every position was demonstrated to be plausible and this plan enormously diminishes the hunt space and improves the calculation execution.

Another researcher proposed a novel DM system with the name of FD-Mine that can productively use the cloud nodes to quickly find incessant examples in distributed computing situations with information protection saved. Through observational assessments on different re-enactment conditions, the proposed FD-Mine indicated to convey fabulous execution regarding scalability and reduce processing time. A Data Mining library is available, it is open source and the user can utilize techniques implemented in this library.

6. CONCLUSION AND FUTURE TRENDS

This study identifies business intelligence and data mining applications in a computing environment. In particular, we use the incredible conceptual results, data mining algorithms and business intelligence to calculate high-dimensional issues that the framework deals with standard technology.

this study initially introduces the basic features of the cloud computing structure, discussing it at a diverse level that allows us to understand its trend. We presented its advantages along with the definition of other established innovations, for example, grid computing and correspondingly how this project makes data mining applications profitable. This study examines data mining (DM) and business intelligence (BI) applications in the cloud infrastructure. In particular, a four-tier structure is recommended following the advice given in the specific article. This structure allows management nodes and correspondence systems to be minimized, parallel storage of data to spread data across the cloud, and implementation situations in which the features

offered should be exploited made up of cloud infrastructure. Bid high queries to completely misuse parallel implementation level features. Implemented data mining algorithms in the cloud environment, and reviewed the implementation features that are primarily in the light of the MapReads project, to allow efficient use of information currently available in the cloud. This is a very delicate process.

REFERENCES

[1] Xiao, Z., & Xiao, Y, "Security and privacy in cloud computing", *IEEE communications surveys & tutorials*, *15*(2), 2012.

[2] Namjoshi, J., & Gupte, "A Service oriented architecture for cloud based travel reservation software as a service",In *2009 IEEE International Conference on Cloud Computing* (pp. 147-150) 2009.

[3] Bhardwaj, S., Jain, L., & Jain, S. "Cloud computing: A study of infrastructure as a service (IAAS)", *International Journal of engineering and information Technology*, *2*(1), (60-63),2010.

[4] Gupta, P., & Narang, B., "Role of text mining in business intelligence.", *Gian Jyoti E-Journal*, *1*(2) 2012.

[5] Vossen, G. "Big data as the new enabler in business and other intelligence." ,*Vietnam Journal of Computer Science*, *1*(1), 3-14, 2014.

[6] Mishra, S.,"Financial management and forecasting using business intelligence and big data analytic tools.", *International Journal of Financial Engineering*, *5*(02), 2018.

[7] Gupta, A., Seshasai, S., Mukherji, S., & Ganguly, "A. Offshoring: the transition from economic drivers toward strategic global partnership and 24-hour knowledge factory.", *Journal of Electronic Commerce in Organizations (JECO)*, *5*(2), 1-23, 2007.

[8] Martens, B., & Teuteberg, F., "Risk and compliance management for cloud computing services:", Designing a reference model. *Risk*, *8*5, 2011.

[9] Li, Q. Q., Zhang, T., & Yu, Y. Using cloud computing to process intensive floating car data for urban traffic surveillance. *International Journal of Geographical Information Science*, *25*(8), 1303-1322, 2011.

[10] Wu, C., Buyya, R., & Ramamohanarao, K., "Big data analytics= machine learning+ cloud computing.", *arXiv preprint arXiv:1601.03115,* 2016.

[11] Ouf, S., & Nasr, M.," Business intelligence in the cloud". In *2011 IEEE 3rd International Conference on Communication Software and Networks* (pp. 650-655), 2011.

[12] Verma, N., & Singh, J., "An intelligent approach to Big Data analytics for sustainable retail environment using Apriori-MapReduce framework". *Industrial Management & Data Systems*, 2017.

[13] Padhy, N., Mishra, D., & Panigrahi, R., "The survey of data mining applications and feature scope.", *arXiv preprint arXiv:1211.5723,* 2012.

[14] Koh, H. C., & Tan, G., "Data mining applications in healthcare.", *Journal of healthcare information management*, *19*(2), 65, 2011.

[15] Rani, D., & Ranjan, R. K., "A comparative study of SaaS, PaaS and IaaS in cloud computing.", *International Journal of Advanced Research in Computer Science and Software Engineering*, *4*(6), 2014.

[16] Hall, K. B., Gilpin, S., & Mann, "G. Mapreduce/bigtable for distributed optimization.", 2010.

[17] Zaharia, M., Konwinski, A., Joseph, A. D., Katz, R. H., & Stoica, "IImproving MapReduce performance in heterogeneous environments.", In *Osdi* (Vol. 8, No. 4, p. 7), 2008.

[18] Zhao, J., Wang, L., Tao, J., Chen, J., Sun, W., Ranjan, R., ... & Georgakopoulos, D., "A security framework in G-Hadoop for big data computing across distributed Cloud data centres. "*Journal of Computer and System Sciences*, *80*(5), 994-1007, 2014.

[19] Cassales, G. W., Charao, A. S., Kirsch-Pinheiro, M., Souveyet, C., & Steffenel, L. A., "Improving the performance of Apache Hadoop on pervasive environments through context-aware scheduling.", *Journal of Ambient Intelligence and Humanized Computing*, *7*(3), 333-345, 2016.

[20] Cavallo, M., Di Modica, G., Polito, C., & Tomarchio, O., "Application profiling in hierarchical Hadoop for geo-distributed computing environments.", In *2016 IEEE symposium on computers and communication (ISCC)* (pp. 555-560), 2016.

[21] Gunarathne, T., Wu, T. L., Qiu, J., & Fox, G.,"MapReduce in the Clouds for Science.", In *2010 IEEE second international conference on*

cloud computing technology and science (pp. 565-572), 2010.

[22] Ranger, C., Raghuraman, R., Penmetsa, A., Bradski, G., & Kozyrakis, C., "Evaluating mapreduce for multi-core and multiprocessor systems.", In *2007 IEEE 13th International Symposium on High Performance Computer Architecture* (pp. 13-24),2007.

[23] Al-Absi, A. A., Kang, D. K., & Kim, M. J. , "Enhancing Dataset Processing in Hadoop YARN Performance for Big Data Applications.", In *Advanced Multimedia and Ubiquitous Engineering* (pp. 9-15) ,2016.

[24] Almeer, M. H., "Cloud hadoop map reduce for remote sensing image analysis.", *Journal of Emerging Trends in Computing and Information Sciences*, *3*(4), 637-644, 2012.

[25] Maillo, J., Triguero, I., & Herrera, F., "A mapreduce-based k-nearest neighbor approach for big data classification.", In *2015 IEEE Trustcom/BigDataSE/ISPA* (Vol. 2, pp. 167-172),2015.

[26] Al-Absi, A. A., Kang, D. K., & Kim, M. J. , "Enhancing Dataset Processing in Hadoop YARN Performance for Big Data Applications.", In *Advanced Multimedia and Ubiquitous Engineering* (pp. 9-15) ,2016.

[27] Khezr, S. N., & Navimipour, N. J., "MapReduce and its applications, challenges, and architecture: a comprehensive review and directions for future research.", *Journal of Grid Computing*, *15*(3),295-321,2017. .



Khan et al LGURJCSIT 2021

ISSN: 2521-0122 (Online)
ISSN: 2519-7991(Print)

LGU Research Journal of
Computer Science and IT

Vol.5 Issue1, Jan-Mar 2021

Information Security for Cloud using Image Steganography

Muhammad Amjad Khan[1], Taimoor Hassan[2], Zaka Ullah[3]
[1]Department of Computer Science, Quaid-I-Azam University, Islamabad.
[2]National College of Business Administration and Economics (NCBA&E) Lahore.
[3]COMSATS University Wah, Pakistan
amjad97@gmail.com

ABSTRACT

Cloud computing is getting involved in almost every technological field to serve customers in a more efficient way. The shared resources (pool) with different configurations according to the user's needs are provided by cloud vendors. Users stores their data on the cloud, data can be personal, or organizational, and data of every type must be secure on clouds. Making the data secure and reducing the integrity of data during transfer through public channels. In this paper, we will try to make data secure using image steganography. Using the steganography technique, we use encryption-decryption of data into images and make data invisible.

KEYWORDS: Cloud Computing, Steganography, K-Strange, Clustering Algorithm, Cryptography.

1. INTRODUCTION

Cloud Computing is helpful in day to day life. It is mainly being used for many apps/software by customers. The main properties of CC are scalability, service on demand, pooling, metered service & many more. National Institute of Standards and Technology (NIST) [1] defines cloud services used by customer customers can control and used in different manners:

- Software as a Service (SAAS)
- Platform as a service (PAAS)
- Infrastructure as a Service (IAAS)

In daily life cloud usage, storing data on the cloud, data can be personal or organizational. As usage increases massively, the ultimatum also increases. Most data can contain the virus for interruption due to the transfer of data using open channels. [2].

Many different methods are available for securing the data while transferring data hop to hop. Cryptography is most famous,

in this technique sensitive data is encrypted and then send, using some of the encryption algorithms or keys (key can be a phrase).

Steganography is the practice of concealing a message within another message or a physical object. It is a technique of hiding data using images as a cover. Any image can be used for steganography. Image format can be in the form of a bitmap, in png, etc. Hiding data in images can save the data for a very long time without being noticed. This method can use secure private messages being transferred through the public network.[3].

Generally, it's possible to bind information(data) in images, but this technique requires digital media for securing data. Steganography can be done using different multimedia images like binary images, color images, grayscale images. The homogeneity in data and cover being used is not mandatory. This meant that audio can be embedded with the data or in the same manner data can be embedded with videos also.

Together steganographic and cryptographic methods are controlling and vigorous. In this paper, the main purpose is to evaluate numerous ways of joining steganography [12] and cryptographic methods to accomplish a cross-system. Moreover, some of the changes between cryptographic and steganographic methods were obtainable as well.

In this paper, we conduct a comparative study of steganography and cryptography. We review several approaches joining cryptography and steganography methods in one system. Moreover, we present an arrangement of these approaches, and comparison them in terms of the algorithm [13] used for encoding, the steganography technique and the file type used to cover the information.

2. RELATED WORK

Securing the digital medium (images) can be run by encrypting media-image. Encryption can be performed in a method that the actual image is reconstructed into the same appearance. Now, the main encryption algorithms or we can use such

algorithms that are suitable for working with images and pixels [4].

The encryption system which is described in the above section, plays' important role in containing key-generating and objects starting in the rule of cryptography. The running apps for such systems have been discouraged due to the pivot's fault [6].

The methods described in this paper are on a novel approach for hiding data in images. The 1st way is, to identifying of 2-bit code for information to be attached. The 2nd step is to use the image as a cover, hide the data using the selected cover using a 2-bit code. Then the data and image get combined and converted into bits [6].

The simple messages or text files can be embedded directly into the cover in the basic steganography technique. When the bits are to be inserted into the cover bits. Then at the end of the recovery, it tries to match the data bits in the bits of the pixel. If there are found any match between data bits and the pixel bits & the extreme nibble bits in the image, random pixels and process isn't hidden in any method. But if no matching found then this would be would result in's problem to store data [7].

3.PURPOSED METHODOLOGY

In this section we will describe the model of our work, in short, we can say that this will describe the working. We do implement the model we need a cover as an image, encrypted, decrypted, data to be hidden with the cover and after that actual secured image will be stored on the cloud.



Figure 1: Encrypt/Decrypt Data



Figure 2: Data Stored

This area gives the means associated with concealing the information. Typically, the useless code is produced as clarified in the accompanying section. As an underlying advance pick a spread picture, CI. Haphazardly elect the pixel (a, b) from CL here a & b signifies line and segment of picked pixels individually. Give Aka a chance to speak to higher snack bit places of the pixel (a, b) i.e. Ak = 8, 7, 6, & 5. Give 'd' a chance for the information to be inserted and its period is meant by 'l' bits. Suppose that the nth piece is Dn.

The useless code is produced utilizes the accompanying advances.

1. Inspect the bit location Ak in Pixel (a, b), pixel is arbitrarily picked.
2. Check if an nth piece of information Dn is a matched bit location of Ak.
3. Distinguish code good for nothing i.e. 01, 00, 01 and 11 to be embedded along the coordinating location speaking to the extreme snack's bit locations.
4. Chances that the information bit doesn't coordinate, at that point bits in the higher snack are not quite the same as the information bit. suppose that there is a matching bit location 5 [8].
5. During translating the message, the beneficiary will be searched for this case. The beneficiary will find a supplement of the bit's in the 5th location which would be the information bit.
6. Rehash stage 1-4 will be information bit of Dn where n < l 2) Passcode base coordinating: Select a secret key 'Go' with 'CH' string length.

It can be of any length & could have any grouping of a string with capitalized or lowercased letter sets, & uncommon words. The secret phrase will be changed in the form of parallel code which would be utilized for putting away 1 bit for useless code.

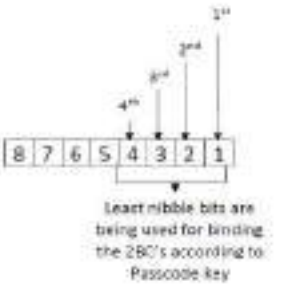

Figure 3:Nibble Bits

The flowchart portrayed gives a pattern of exercises to be performed to insert information on the sender's side.

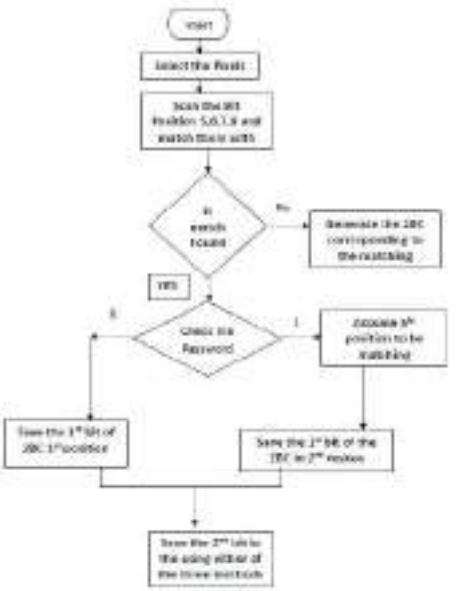

Figure 4: Insert Information

The flowchart portrayed gives a pattern of exercises to be performed to insert information on the sender's side.

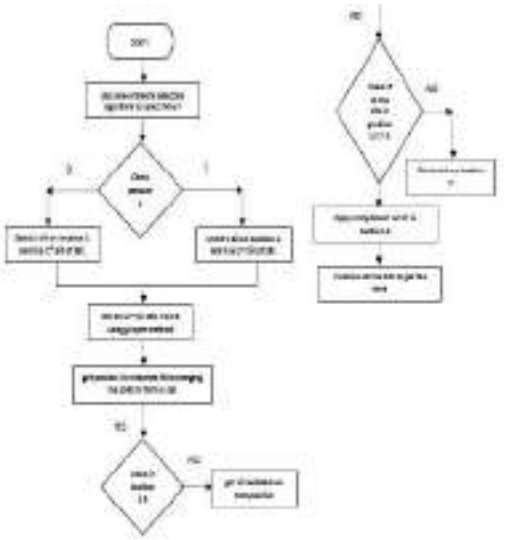

Figure 5: Password Checking

4.PROPOSEDWORK IMPLEMENTATION

The proposed strategies work perfectly according to giving the privileged of the data or information covered up in a picture. So, information without much of a stretch is implanted into the spread picture and send using the public network to the collector. Speculation tests and examination demonstrates that this is a remarkable effect of techniques to be used for concealing information according to the number of strings to be installed on spread picture. Consistently range for upcoming improvements for information privacy is a ceaseless problem in the present situation.

4.1. Improved K-Strange Clustering Algorithm:

Improved K-Strange Clustered Algo, improved K-Strange spike [10] shown in examination going to scan for the focuses which are max separated each other. Calculation likewise denotes impact according to the most distant focuses according to the informational collection. It finds the base K-minimum from the dataset & it's worth finding using calculation [11] in-between everyone according to the dataset. Besides, it's discovered maximum point from K-minimum. That figure is to be spoken K-maximum. Calculation upgrades chances by finding the 1st base point. By using that technique stays away from the choice of different focuses by calculation of the separation among focuses from the dataset. This acquires 0(nth) multifaceted nature periodically searching K-minimum & K-maximum. At that point, it will assess 3rd bizarre location inside the informational collection would be maximally isolated from focuses. On the other hand, there is some chance that the 3rd possibility is nearer to K-minimum, at that point location of the 3rd point is changed utilizing by choosing the in-between point next to 3rd distinct point & K-maximum & this last bizarre extraction. Be that as it may, if the third new point is nearer to K-maximum, at that point the situation of 3rd distinct point is moved by finding an in-between point [6].

4.2. Least Significant Bit Technique

It determined picture steganography epitomizes shrouded data at all critical bits of picture estimations to bearer picture. It divides the picture pixel to concealed information.

Table 1: Divide picture pixel

P1	10011111	01010100	10110101
P2	10000100	00010101	01010001
P3	11000100	11100111	01001011

The binary number 11111110 whose number is 254 is encapsulated into the LSB of the part of the image; and we obtain the data sets as shown below after encapsulating:

Table 2: Data sets after encapsulating

P1	10011111	01010100	10110101
P2	10000100	00010101	01010001
P3	11000100	11100111	01001011

Just 3 bits changed over as per the exemplified message. Yet, just not many bits in the picture will be adjusted to shroud mystery information. For

potential powers of every essential shading are 256, changing over the LSB of a pixel brings about a little believer in the forces of the hues.

Initially, actualizing K-Mean calculation & indicating a yield of both, sending end and recipient end individually. Presently working on similar tasks with Enhanced K-Strange.



Figure 6: A picture chooses for implementation

The optimized K-Strange Clustering Algorithm:

While expanding the size and higher dimensionality of information, Enhanced K-Strange Points Clustering Algorithm works superior to anything k-Means grouping. The table gives the immaculacy of the second strategy, which is a proportion of the level to which a group has things appropriate for the long bunch. Our proposed calculation runs quicker as well as gives great nature of bunches.

Table 3: Cluster algorithm values

Clustered Algo	Values
K-M	.81
Optimized	.84



Figure 7: Contribution for Enhanced K Strange technique

The bunches framed for the picture is as demonstrated as follows:



Figure 8: Information to be embedded

Adding information chunks in the image pixels. Appeared in the below fig:

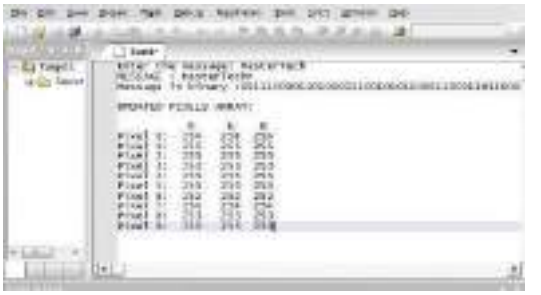

Figure 9: Message addition into the biggest bunch

The information will be spread over 0-9 pieces on the image. With the information, its quantity is additionally added to the beneficiary through chosen media.



Figure 10: Stegoed picture at the receiver end

Information quantity is expected to provide just these strings characters of information. Trash esteem show, no information embedded in a specific location internally in pixels. The picture got on the beneficiary side is appeared in fig 10.

The bunches framed on the staged picture areas demonstrated as follows:



Figure 11: Output after optimized K-Strange Clustered method

At recipient nod staged picture is gotten & bunching is permeated & a similar procedure happens.

Correlation with the proposed system is: The groups shaped for the picture is as demonstrated as follows:



Figure 12: Final Cluster

Beginning along got yield in groups picks biggest 1 & addition's information similarly. As appeared below:



Figure 13: Matrix pixels output

5. RESULTANT

After the implementation and work, the following image was obtained by applying steganography using k-String min and max methods as discussed and implemented above. The image is shown in Figure13: contains the data that needs to be stored on the cloud showing as the image, but this image contains the hidden data

The following image is obtained after embedding the information using the k-strange method to image pixels. After applying the steganography, the image's actual pixels have not fluctuated. The picture is similar to before applying the steganography. The picture got in the beneficiary end is as demonstrated as follows:



Figure 14: Receiver side procedure - The staged picture.

The image received at the receiver's end will provide information as shown below after passing the image through decrypt or the information will be shown like below. The bunches framed on the staged picture are as demonstrated as follows:



Figure 15: Receiver side procedure: Output of K Means Clustering at beneficiary side alongside message evoked

6. CONCLUSION

In paper to acquire steganography, we have used K-Strange spikes Clustered Algo. In this paper, we have discussed steganography using different methods. In short, we are hiding the data by using an image as a cover. Cover images can be in any form bitmap, png, jpeg, etc. At that point we thought about the result of steganography strategies by using K-Mean grouping calculation, reason's that purposed technique of optimized K-Strange spikes Clustered Algo works perfectly as described.

REFERENCES

[1] P. Mell and T. Grance, "NIST definition of cloud computing," National Institute of Standards and Technology, 2009.

[2] P. Marwaha and P. Marwaha, "Visual Cryptographic Steganography in Images," in Proceedings of International Conference on Computing Communication and Networking Technologies, Karur, India, 2010.

[3] W. Mazurczyk and K. Szczypiorski, "Is cloud computing steganographyproof?" 2011 Third International Conference on Multimedia Information Networking and Security, pp. 441–442, 2011.

[4] X. Wang, J. Zhao, and H. Liu, “A New Image Encryption Algorithm Based on Chaos,” Optics Communications, vol. 285, no. 5, pp. 562 – 566, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0030401811012430

[5] C. Zhu, “A Novel Image Encryption Scheme Based on Improved Hyperchaotic Sequences,” Optics Communications, vol. 285, no. 1, pp. 29 – 37, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0030401811009540

[6] M. Amin, O. Faragallah, and A. Ahmed, “A Chaotic Block Cipher Algorithm for Image Cryptosystems,” Communications in Nonlinear Science and Numerical Simulations, vol. 15, pp. 3484–3497, 2010.

[7] A. Ahmed, X. Niu, and N. Wang, “Chaotic Image Encryption Using Bezier Curve in DCT Domain Scrambling,” Communications in Computer and Information Science, vol. 194, pp. 30–41, 2011.

[8] V. Patidar, N. Pareek, G. Purohit, and K. Sud, “A Robust and Secure Chaotic Standard Map based Pseudorandom Permutation-Substitution Scheme for Image Encryption,” Optics communications, vol. 284, pp. 4331–4339, 2011.

[9] W. Zhang, K.-w. Wong, H. Yu, and Z.-l. Zhu, “An Image Encryption Scheme Using Lightweight Bit-Level Confusion and Cascade Cross Circular Diffusion,” in Optics Communications, vol. 285, 05 ,pp. 2343 – 2354,2012.

[10] Johnson, Terence,S.K. Singh, “Quantitative Performance Analysis for the Family of Enhanced Strange Points Clustering Algorithms”, International Journal of Applied Engineering Research, Volume 11, Issue 9,, pp: 6872-6880,2016.

[11] Alfakih, Abdo Y, Henry Wolkowicz, and Amir Khandani, “Solving Euclidean distance matrix completion problems via semidefinite programming”, Computational optimization and applications, Volume 12, Issue 1, pp: 13-30, 2015.

[12] Taha, M. S., Rahim, M. S. M., Lafta, S. A., Hashim, M. M., &Alzuabidi, H. M.,“Combination of steganography and cryptography: A short survey.”,In IOP conference series: materials science and engineering (Vol. 518, No. 5, p. 052003),2019.

[13] Almuhammadi, S., & Al-Shaaby, A. “A survey on recent approaches combining cryptography and steganography. Computer Science Information Technology (CS IT).”,2017.



Malik et al LGURJCSIT 2021

ISSN: 2521-0122 (Online)
ISSN: 2519-7991(Print)

LGU Research Journal of Computer Science and IT

Vol.5 Issue1, Jan-Mar 2021

Analysis of Advantages and Problems in Teaching and Assessment with Online System during Covid-19

[1]Haseeb Aslam Rana, [2*]Aftab Ahmad Malik, [3]Naureen Riaz,[4]Mujtaba Asad
[1 2,3,4] Lahore Garrison University Lahore
dr_aftab_malik@yahoo.com

ABSTRACT

During Covid-19 spread over, throughout the world, online classes were arranged at school and University level. In certain institutions, the experiments of online systems were successful while in certain cases difficulties were observed and reported. In this paper, we are interested to highlight and analyze the problems occurring in assessment and evaluation. The delivery of lectures and grades assessment, comparison of time spent on in-class lectures and online, the provided infrastructure and technology to faculty and its synchronization with that arranged by students. The objective type questions are automatically marked by the system for example Moodle, whereas the essay type (subjective type) questions are normally preferred to be marked by faculty on the system. This is indeed a facility that all students are provided with different question papers. The online system is more economical, time-saving and easily usable. The most serious issues have been unfair means and cheating used by students during examinations. A detailed analysis of the state-of-the-art is presented in this paper. We also present a comparison of online and in-class teaching and assessment, but irresistibly the benefits of online system are more advantageous. For example, a lot of stationary is saved. The online system causes breaks and pauses during delivery of lessons due to instability of internet and concavity issues. The most important is the choice concerning system subject to available facility with leaner and the faculty and its compatibility with emphasis on technology. The use of anti-cheating software makes the examination secure.

KEYWORDS: Teaching online, Learning Management System (LMS), Online Examination System (OES)

1. INTRODUCTION

The learning systems are of three types, on-line, in-class on campus and hybrid. In view of the massive spread of Covid-19, online teaching and assessment gained rapid popularity around the world to save time for students during the closure of academic institutions affecting 150 crore students. The second author of this paper has shown a comparison of OES and LMS in [1] which is very beneficial in using an appropriate online system. However, a series of solutions have also emerged to cause improvements. In USA and Europe, according to [1], the platforms such as Google Class Room, Cloud and Moodle are functioning satisfactorily. The academic institutions and faculty are happy with the performance of Virtual systems of teaching its reliability and efficiency. The e-textbooks are also provided to students online for study. According to [2], a methodology to develop online curriculum and its teaching has been effective. It has been advocated that the camera of faculty must be on during teaching for the learner's concentration. According to [3], digital, mobile technologies , Wi-Fi have been combined with other electronic devices to maintain requirements of education and describes the experiences in Nepal to implement online policies. Further [3] also highlights the role of the Indonesia Education program "learning House" which described useful information to strengthen online education, while Malaysian Television Channel has been delivering online lectures/lessons for all students not having internet facility.

According to [4], UNESCO has compiled a document that 1.9 billion students round the world in about 191 countries used internet and attended online classes during covid-19. Moreover, faculty, members of organizations have also working from home and instructing their subordinates to perform official responsibilities.

While teaching online and conducting the examination, a usual trend of students was observed that they do unfair means by sending screenshots (questions) for help from outside mostly by screen sharing. This is how

they dodge the technology by using a friend for help. Some do copy answers by writing on the desk or by keeping answers in their lap. The Zoom provides every user a unique identity and students face problems using unauthorized friends for cheating and getting help. However, Zoom itself is helpless to detect cheating. But setting the students video on, we may overcome this problem.

In Pakistan, according to [5] statistics for 2020-21 are as follows:

- The number of observed Coronavirus cases: 549032
- Deaths: 11802
- Recovered : 504,046

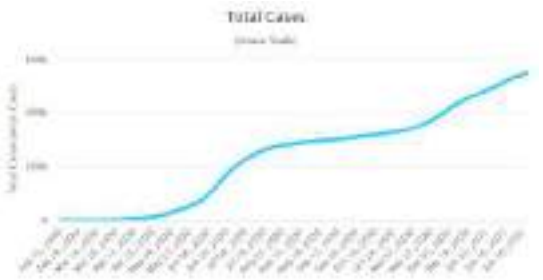

Figure 1: Total corona cases in Pakistan (Source [5])

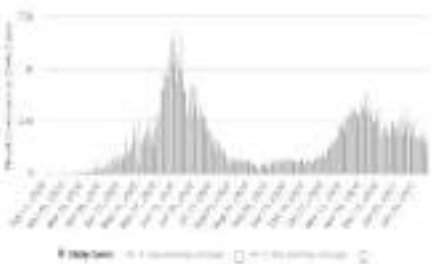

Figure 2: Daily new Corona cases in Pakistan (Source [5])

2. LITERATURE REVIEW

According to [6], a few useful hypotheses have been presented regarding the relationship of communication technology and learning as well as the quality of the presentation on E-learning.

According to the survey conducted by the authors of [7], it is accomplished that in Pakistan online education for Medical and Dental students is not suitable as compared to on-campus classes and recommend that the Management and faculty must take adequate steps to improve the quality of online classes.

According to [8], the authors are of the view that online learning is more sustainable if a hybrid system is introduced with more opportunities. Further, it is suggested that technologies to develop the instructional material for online lessons must be strengthened for effectiveness and efficiency by well-trained faculty having comprehensive knowledge and skill in this area. In certain cases, it is supposed by authors of [8] that a complete curriculum is not delivered remotely with available internet facilities.

The students may not have the requisite compatible technology to synchronize with online lessons.

According to the survey conducted reported in [8], the majority of students faced internet signal strength problems and wished that learners must have face-to-face contact with the faculty.

According to findings reported in [10], the authors are of the opinion that student's convenience and inspiration plays a pivotal role while the faculty should use technology to enhance the learning online. Other points highlighted by the authors relate to transmission, motivation of learners, students achievements after online learning, synchronization of lectures with curriculum and support of the staff delivering online lectures.

According to [11], presents a quantitative analysis method and suggests for improvement on online system for practicing student assessments and evaluations. The authors of [11] advocate that the course contents must be well organized, supported by faculty, need to develop a liking for online education amongst the students, effective strategy for university faculty and to advance the quality of higher education.

The significance of HR development in areas of Science and Arts has been discussed in [12], motivating to encourage students and faculty to enhance their potentials in the key areas of Human Development for professional practices.

According to [13], it is a matter of basic human rights to have access to education on the basis of impartiality and equity especially during emergencies like Covid-19, disaster and violence. The experience of online teaching has been successful around the world and Pakistan. The experiments of online classes in war and emergencies in Afghanistan, South Africa, Palestine and Syria have been successful.

A model of financing E-education has been reported in [14] in Russia to improve online learning for public and private sector that receive financing. The model provides standards of e-education. Moreover, it highlights a vision in the development of an educational structure for online education.

According to Word Bank Report [15] provides a guideline on remote on online mode of learning during

Covid-19. The world was practicing online teaching to fight out poverty. The is of the view to further improve the quality of online teaching; for this purpose supporting various nations financially to make the best use of usefulness and maximize the use of online working.

2.1. Experience of China in Online Teaching during Covid-19

China managed to successfully overcome the gap during the closure of the educational institutions by introducing online teaching up to the contentment of learners. According to [16], a survey was conducted in this regard based on a questionnaire and web crawler to explore the quality-related issues.

The data was collected online and offline for the survey to perform quantitative analysis.

3. METHODOLOGY

3.1. Features of Information Technology Frameworks for Online Classes

During the last year's teaching and training regarding online classes, it is emphasized that all the components should be functional, automated, combined and intercommunicating for the virtual classroom. According to [17, the most important points of structures are individualities and characteristics such as integration of components of the platform, the methodology of tracking, generation of reports, auto-evaluation scheme for assessment and mechanism for setting up a virtual classroom rather than a virtual classroom. There are several online learning platforms described by [1]; moreover, a list of 33 platforms have by provided by [19]. Table 1 provides some useful platforms and frameworks for online education.

Table 1: Well-known Platforms for online Teaching

1.	Adobe Captivate Prime LMS
2.	Administrate LMS
3.	Accord LMS
4.	Blackboard Education Technology Solutions
5.	Best LMS for Banking & Finance Companies
6.	Financial Services – Best LMS for Banking & Finance Companies
7.	Looop: Next Generation Learning Platform
8.	Moodle LMS – the world's most popular learning platform. An open source LMS.
9.	Pathway Cloud LMS Review
10.	Xyleme LCMS

All the features of the online technology must possess the requirement of learning such as self-registration, reporting regarding examination, attendance, course-library and databank of questions of every course to be taught, consistent with accreditation principles. The engrossment and participation of students must be made mandatory. The course library may consist of all lectures date-wise along with video Zoom Lecture files. According to [18], an understandable user interface and dashboard is of pivotal importance in every learning platform. In the next section, we briefly highlight the MOODLE, GOOGLE CLASSROOM and activities performed with the ZOOM system.

3.2. Moodle System

The institutions offering online have LMS (Learning Management System) on the private website filled with learning courses using Moodle, which is a free online system. It is a well-known and useful way for all sorts of tasks such as online learning for academic and industrial organizations. It possesses facilities such as provision for Chat Room, assignment, Quizzes and Questions data bank, etc. Using Moodle, every learner can easily obtain lecture material such as e-books and lecture slides and appear in Quizzes, submit assignments and appear in the examination. Each student is allocated a unique login for online work. Apart from being a user-friendly system, Moodle can be easily integrated with the LMS components and is a secure system. The features like course admiration, review option and customization make Moodle more attractive for online teaching. In order to overcome cheating by students, Zoom activity may be created to Moodle course space and a meeting room during the examination to watch them with a camera by screen sharing.

3.3. Google Classroom

It was developed by Google in 2014 which uses operating systems: Android, IOS and web available free. It handles the sharing files and grading assignments between learners and instructor and handles unlimited classroom size. A unique folder is created when the student submits their work, which is not used, scanned, or sent to other companies. It is easier to use and supports online learning efficiently.

3.4. Zoom

For audio/ video conferencing software, Zoom plays a very important role as a leader and reliable platform and is extensively used for chat, conferences and online seminars (webinars) for 100 participants. It uses Windows, macOS, IOS and Android.

3.5. Feasibility of Online Learning

The well-known benefits of online learning are a conversion of classroom lectures into Power-point Slides and electronic form with voice using Zoom Video has been very useful in connection with a framework, use of technology, design issues, assessment and evaluation of examinations. The roll

call of students can also be displayed and managed on an e-portal specially designed for this purpose.

The online learning systems provide prospects, ease, enhance access to the benefits of learning especially remote learning and updating skills of faculty.

Online learning systems are financially feasible being cost-effective. The quality, capacity and ability of the academic system are improved. There is a great deal of reduction manual work in online assessment and evaluation. Apart from manual labor, the cost is also reduced and security is enhanced. It saves student's expenses in traveling and residing in Hostels.

The provision of a quick training facility in a specified geographical area is the salient feature of e-learning, assuming the arability of requisite technology. We may improve further it by introducing international characteristics. The digital evaluation and assessment increases the confidence of the learners in the system and decreases their doubts about manual marking. The new methodologies can be formulated in digital orientation as for as online learning is concerned

3.5.1. Problems Confronted in Online Learning

The online lecture preparation requires more time as compared to the classroom environment where students are physically present. It is easier to postpone the class and reschedule. The online system develops a feeling of loneliness and segregation. The browser containing the databank can be hacked, if not properly protected. Some students or outsiders may interfere and disrupt the online class. However, there is a facility in Zoom to mute all or for a particular person causing the disturbance. Such activities may be controlled by ensuring the security of the browser.

The students living in remote areas where internet facility is either not available or have problems with bandwidth or not seamless (continuous). It then causes interruptions. There also exists inequality of internet in certain geographical areas, especially with context to Pakistan. The background of the student's family should also be kept in mind regarding their locations in remote areas as well as the financial position. The important issue is that of the faculty training regarding guidance of students to use online technologies.

3.5.2. Working with Online Examination System

In a manual system of examination, a written or typed question paper is given to students, they use a pen and paper approach to write a solution to the given problem. Whereas in an online system no stationary is involved. Every student may have a unique and different question paper. The system, for example, Moodle, develops different question papers for all students appearing in the examination, selecting a question from a Data Bank. The students have to complete the question paper within the stipulated time limit. The objective type questions and multiple-choice questions are automatically graded by the system. Thus resulting in a lot of saving of time for the faculty.

The online conducting the examination remotely has been a wonderful experience during Covid-19 which guaranteed the continuousness of teaching and learning. One may calculate saving in the cutting of trees for preparing paper stationery around the world during covid-19. As for learning management systems are concerned, the use of Blackboard, Moodle and cloud technologies have been extensively used in recent times. The currently available system is advanced, comical, return quick over,

Moreover, the online examination system also possesses the following beneficial characteristics:

- It is user friendly and easy to use
- Inexpensive,
- Efficient and faster,
- Operationally and technically reliable
- Provides auto-grading mechanism,
- Computes and provides the evaluation,
- Assesses and compute overall grades
- Provides graphs on the performances of students

The online examination system is inclined and subject to possible fraud. An online exam system is a little bit more susceptible to fraud because the students sitting at home use their own devices. The vigilance is comparatively naïve There are some difficulties faced by learners due to the implementation of requisite technologies in online teaching. The essay type question is marked by faculty most of the time personally. Sometimes students behave as if they are appearing in an open-book examination, if not permitted, it amounts to cheating and fraud. The use of smart i-phone and smartwatches can be used to support unfair means in the online examination.

3.6. Data Collection for Proposed Methodology

A survey using a questionnaire consisting of 16 questions given in Table 3 was provided online to 300 students who attended online classes throughout the covid-19 pandemic period. The number of total responses received was 270 and 30 students did not respond. Different questions regarding issues such as the usefulness of online classes, evaluation and assessment, use of Zoom, problems with internet faced by students at their home, availability of instructional material and comparative quality of theory and Lab instructions.All the questions are self-explanatory. Table 2 contains the codes of a variable used

Table 2: Codes of Variables

Code	SD	DA	N	A	SA
Variable name	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

The collected data were used to estimate various trends among the students during online learning, especially the level of satisfaction were estimated to know the confidence level of learners on the online education system.

4. RESULTS AND ANALYSIS

4.1. Data Analysis of Survey

The numerical entries of Table 3 show the number of responses for each question for the five variables given in table 2. That indicates the level of opinion for each category among students.

Table 3: Analysis of Advantages and Problems in Teaching and Assessment with the online system during Covid-19

Q#	QUESTION	SD	DA	N	A	SA
1.	I cannot concentrate on my studies, as I keep worrying and thinking about getting Covid-19	21	50	79	79	41
2.	Do you think teachers are more lenient in online classes while evaluation of your assignments and quizzes?	28	64	33	76	12
3.	Do you think teachers are more lenient in online classes while awarding your sessional marks?	29	72	89	70	10
4.	Are you satisfied with online classes on Zoom?	36	61	55	82	36
5.	Online classes make me comfortable during covid-19 pandemic	17	42	51	83	77
6.	Do you think online classes are more flexible than on campus classes?	53	75	48	51	43
7.	Do you think while online education, students can have more time for other activities?	33	42	40	97	58
8.	Do you think online classes save traveling time and provide an opportunity to students for self-study?	12	26	42	105	85
9.	Do you think interest is lost in study during online classes?	26	39	48	80	77
10.	Do you think interaction between teachers and students has significantly reduced during online classes?	15	31	67	104	60
11.	During online classes can teachers interact with students using various ways?	14	32	74	113	37
12.	Were you provided online e-books, lectures/slides and instructional videos for Lab/theory?	6	21	75	133	35
13.	Do you think up to some extent teaching is effective online?	24	41	104	66	35
14.	Do you think theoretical courses can be better explained in online classes?	28	39	64	84	55
15.	Do you think Lab Courses can be better explained in online classes	79	81	55	28	10
16.	should health be considered an important factor while deciding to reopen universities	6	10	43	95	116

Figure 3 shows a level of opinion for each category among all students who participated in the online survey.

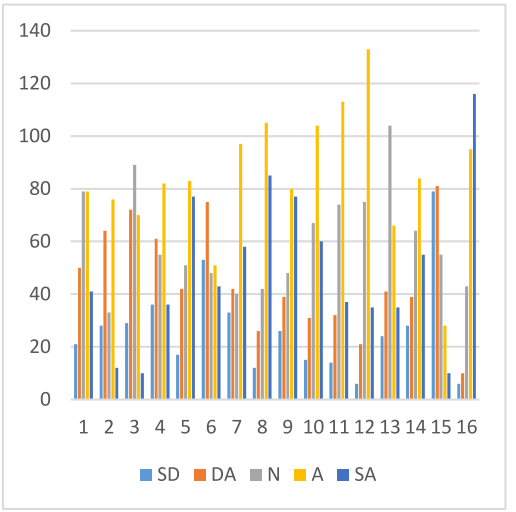

Figure 3: Bar graph of overall response showing a various level of opinion

4.2. Estimation of Satisfaction Coefficient

It is beneficial to know the overall impact of online education and learning satisfaction of users (learners). To estimate the satisfaction and dissatisfaction coefficients from the data, the SPSS Statistics Software was used. SPSS Statistics is used to manage and analyze quantitative data based on statistical analysis. Following mathematical estimations were established to calculate the satisfaction coefficient.

$$X_i = \frac{1}{270} * \sum_{i=0}^{16} [\,(A_i + k * SA_i) + (\frac{N_i}{N_i + A_i})] \quad \ldots.(1)$$

$$Y_i = \frac{1}{270} * \sum_{i=0}^{16} [\,(D_i + k * SDA_i) + (\frac{N_i}{N_i + D_i})] \quad \ldots.(2)$$

Where X_i and Y_i are satisfaction and dissatisfaction coefficients, respectively. A_i is the total agreed responses to each ith question, D_i is the total disagreed responses to each ith question, SA_i is the total strongly agreed responses to each ith question, and SDA_i is the total strongly disagreed responses to each ith question. The k is the constant to give an additional 10% multiplicative weight to satisfaction and dissatisfaction coefficients based on a total of strongly agree and disagree responses.

Table 4: Satisfaction and dissatisfaction coefficients

Q #	Satisfaction Coefficient (Xi)	Dissatisfaction Coefficient (Yi)
1.	0.46	0.27
2.	0.33	0.35
3.	0.30	0.39
4.	0.45	0.37
5.	0.62	0.23
6.	0.37	0.50
7	0.60	0.29
8.	0.74	0.15
9.	0.61	0.25
10.	0.63	0.18
11.	0.57	0.18
12.	0.64	0.11
13	0.39	0.25
14.	0.54	0.26
15	0.15	0.62
16.	0.83	0.06

Figure 4 shows the overall trend of both the satisfaction and dissatisfaction coefficient

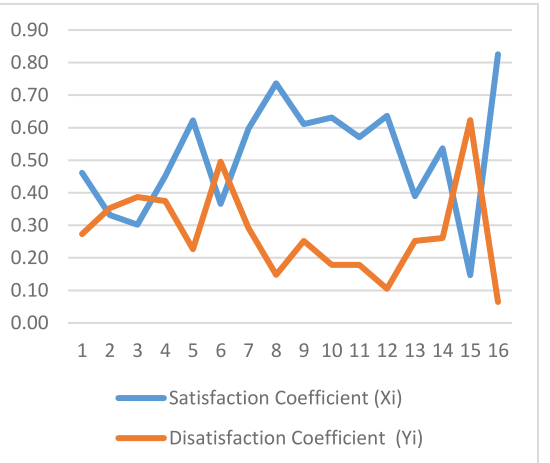

Figure 4: Satisfaction and dissatisfaction coefficient

The satisfaction coefficient calculated in figure 4 was compared with Cronbach's Alpha Coefficient for the results of 11 questions and the comparison shows the consistency. Table 5 presents the Cronbach's Alpha Coefficient for satisfaction coefficient.

Table 5: Data reliability test using Cronbach's Alpha Coefficient

Index	Cronbach's Alpha Coefficient	N of Questions
Degree Of Satisfaction	0.699	11

4.2.1. Students Satisfaction based on Comfort and Grading for Online Education on Zoom

In figure 4 the slope from points 3 to 5 shows an increasing trend that indicates the strengthening of satisfaction coefficient among the students getting their online education on Zoom. While learning on Zoom users can aspects higher grades in a comfortable and convenient environment. Figure 5(a-c) supports the evidence by presenting the percentage of responses against questions 3 to 5. The green and purple portions of the pie-charts shows the percentage of agree and strongly agree for overall responses in the survey from 3 to 5.

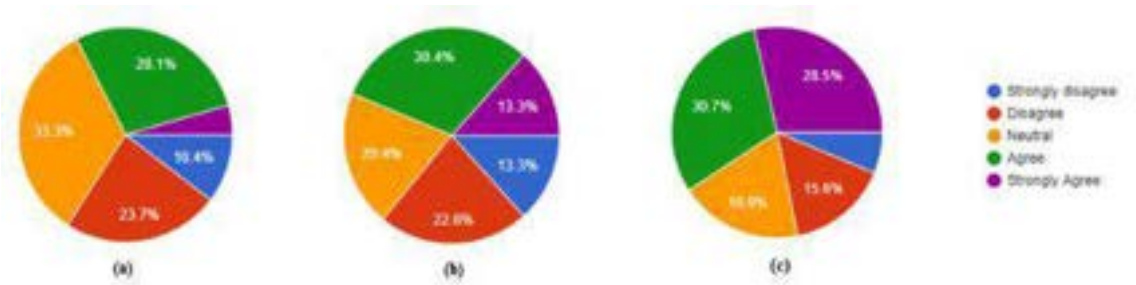

Figure 5: Percentage of responses against questions 3 to 5

To verify the correlation between multiple factors of the online education system for user satisfaction, Chi-Square Tests were performed using SPSS Statistics Software. The Chi-Square Test is a nonparametric test that is used to verify the association between categorical variables and to check the dependence of correlated variables with each other. For this purpose, the strategy of [20] was followed up to compute Chi-Square Test results of questions 3 to 5 and questions 6 to 8. Table 6 shows the results of the Chi-Square Test performed on the results of questions 3 to 5 to verify the students' satisfaction based on comfort and grading for online education on Zoom:

Table 6: Chi-Square Test for questions 3 to 5

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	277.316	25	.000
Likelihood Ratio	34.607	25	.096
Linear-by-Linear Association	0.835	1	.361
N of Valid Cases	256		

Table 6 checks the student satisfaction based on comfort and grading for online education on zoom. According to the table results, it is evident that there is a statistically significant association.

4.2.2. Students Satisfaction based on Flexibility of Online Education System

In figure 4, the slope from point 6 to 8 shows an increasing trend that indicates the strengthening of the satisfaction coefficient for online education based on the flexibility of the online education system. Users' data convince that a higher percentage of learners agree that the online education system provides flexibility and various approaches to adopt for the learning that could also save much time. Figure 6(a-c) supports the evidence by presenting the percentage of responses against questions 6 to 8, respectively. The green and purple portions of the pie-charts shows the percentage of agree and 'Strongly Agree' for the overall responses in the survey from Figure 6 to 8.

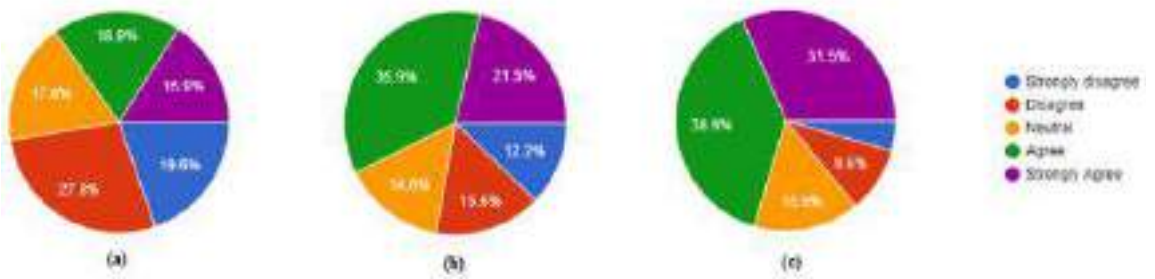

Figure 6: Percentage of responses against questions 6 to 8

Similarly, the Chi-Square Test was performed on the results of questions 6 to 8 to verify the student's satisfaction based on the flexibility of the online education system. The table shows the results of the Chi-Square Test performed on the results of questions 6 to 8.

Table 7: Chi-Square Tests for questions 3 to 5

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	420.086	25	.000
Likelihood Ratio	173.178	25	.000
Linear-by-Linear Association	25.497	1	.000
N of Valid Cases	256		

Table 7 checks the students' satisfaction based on the flexibility of online education. From the results, of table 7 it is evident that there is a statistically significant association.

4.3. Recommendations for Online Education System

a. The online system is more economical, time-saving and easily usable
b. The working experience of successful implementation of online communications arranged by organizations in case of websites, video conferencing, Google Classroom and Zoom must share and told to learners.
c. More emphasis must be on the skills development for online education in case of university students undergoing towards degree or diplomas.
d. The body language of faculty must be exhibited through the online camera during teaching
e. It is strongly recommended that the faculty and student must get full benefits of advanced online courses projected by various foreign universities and non-profitable private organizations discussed in detail by [11]. The important organizations are Coursera, edX and Ted-Ex.
f. The concerns of the World Bank about online and distance learning must be kept into consideration.
g. While switching on online teaching teachers lost tools such as body language and facial impression be managed by making the video live.
h. The course library may consist of all lectures date-wise along with video Zoom Lecture files.
i. The use of anti-cheating software makes the examination secure.

j. In Pakistan online education for Medical and Dental students is not suitable as compared to on-campus classes and it is recommended that the Management and faculty must take adequate steps to improve the quality of online classes.

k. The requirement of learning such as self-registration, reporting regarding examination, attendance, course-library and 'databank of questions' of every course to be taught.

5. CONCLUSION

The entire success of online classes depends mainly on the experience of faculty, organizational standing, status, faculty receptiveness and special care of learners, efficiency and reliability of involved technologies. The experiments of authors using Moodle, Goodloe Classroom and Zoom for online education have been encouraging and successful.

The advantages and disadvantages of online assessment and evaluation and teaching have been discussed in this paper. With the advancement of Computer and Communication Technology, the instructional techniques and procedures of imparting instructions have improved with the right online systems. More secure, authentic, vigorous and latest systems are available for use for this purpose.

The elegant, stylish and well-designed lectures have a more effective impact on the learner. There exist no experimental, pragmatic, realistic, or empirical evidence that the delivery of online education has been 100% fruitful and beneficial around the world. In the future, there may be experiments to enhance, achieve and adoption of valuable systems for online teaching.

REFERENCES

[1] G. Noureen1, Dr. Aftab Ahmad Malik, Dr. M. A. Fahiem, "Selecting a Right Online Educational Paradigm in Covid-19 Pandemic, Pakistan.", Journal of Science. Vol. 72 No. 3, PP:164-168, 2020.

[2] Sun, A., & Chen, X.,"Online education and its effective practice: A research review.", Journal of Information Technology Education, 15, PP157-190, 2016.

[3] Dawadi, S., Giri, R., & Simkhada, P., " Impact of COVID-19 on the Education Sector in Nepal-Challenges and Coping Strategies.", Sage Submissions. Preprint. doi: https://doi. org/10.31124/advance, 12344336, v1, 2020.

[4] UNESCO, "*COVID-19 Impact on Education,"* UNESCO, Jan. 25, 2021.

[5] *"Total Coronavirus Cases in Pakistan",* Jan. 25, 2021. Accessed on: Feb, 2021.

[6] Shehzadi, S., Nisar, Q. A., Hussain, M. S., Basheer, M. F., Hameed, W. U., & Chaudhry, N. I. , "The role of digital learning toward students' satisfaction and university brand image at educational institutes of Pakistan:" a post-effect of COVID-19. Asian Education and Development Studies, 2020.

[7] Abbasi, S., Ayoob, T., Malik, A., & Memon, S. I., "Perceptions of students regarding E-learning during Covid-19 at a private medical college." Pakistan Journal of Medical Sciences, 36(COVID19-S4),2020.

[8] Adedoyin, O. B., & Soykan, E. , "Covid-19 pandemic and online learning: the challenges and opportunities. Interactive Learning Environments," pg 1-13, 2020.

[9] Adnan, M., & Anwar, K, "Online Learning amid the COVID-19 Pandemic: Students' Perspectives.", Online Submission, 2(1), 45-51,2020.

[10] Ali, W., "Online and remote learning in higher education institutes: A necessity in light of COVID-19 pandemic.", Higher Education Studies, 10(3), pg 16-25, 2020.

[11] Sun, A., & Chen, X., "Online education and its effective practice:", A research review. Journal of Information Technology Education, 15.PP-157-190,2016.

[12] Schroeder, H, "The importance of human resource management in strategic sustainability: An art and science perspective." Journal of Environmental Sustainability, 2(2),2012.

[13] Affouneh, S., Salha, S., & Khlaif, Z. N., "Designing quality e-learning environments for emergency remote teaching in coronavirus crisis.", Interdisciplinary Journal of Virtual Learning in Medical Sciences, 11(2), 135-137, 2020.

[14] Nie, D., Panfilova, E., Samusenkov, V., & Mikhaylov, A., "E-Learning financing models in Russia for sustainable development.", Sustainability, 12(11), 4412, 2020.

[15] Guidance Note on Remote Learning and COVID-19 (English),2020.

[16] Chen, T., Peng, L., Yin, X., Rong, J., Yang, J., & Cong, G, "Analysis of user satisfaction with online education platforms in China during the

COVID-19 pandemic.", in Healthcare (Vol. 8, No. 3, p. 200). Multidisciplinary Digital Publishing Institute.,2020.

[17] Zhang, W., "Entering the 3rd Generation of e-Learning: Characteristics and Strategies.", Journal of Educational Technology Development and Exchange (JETDE), 6(1), 2,2013.

[18] Kumar, P., Keserwani, P. K., & Samaddar, S. G., ,"e-Learning Web Services and Their Composition Strategy in SOA.", In Smart Computing Paradigms: New Progresses and Challenges (pp. 291-302),2020.

[19] Freire, L. L., Arezes, P. M., & Campos, J. C., "A literature review about usability evaluation methods for e-learning platforms.",Work, 41(Supplement 1), 1038-1044,2012.

[20] Munyoka, W., & Maharaj, M., "The effect of UTAUT2 moderator factors on citizens' intention to adopt e-government: the case of two SADC countries." Problems and perspectives in management, (15, Iss. 1), 115-123,2017.



Qadri et al LGURJCSIT 2021

ISSN: 2521-0122 (Online)
ISSN: 2519-7991(Print)

LGU Research Journal of Computer Science and IT

Vol.5 Issue1, Jan-Mar 2021

A Spatial Model of K-Nearest Neighbors for Classification of Cotton (Gossypium) Varieties based on Image Segmentation

Tanveer Aslam[1], Salman Qadri * [1], Najia Saher[1], Syed Ali Nawaz[1], Muhammad Shehzad[1], Syed Furqan Qadri [2], Abdul Razzaq[3], Faisal Shahzad, [1] Dost Muhammad Khan.

[1]Department of IT, Islamia University of Bahawalpur, 63100 Bahawalpur Punjab, Pakistan.

[2]School of Computer Science and S.E, Shenzhen University, Shenzhen, China.

[3]Department of Computer Science MNS-UAM Multan, Punjab Pakistan.

salman.qadri@iub.edu.pk

ABSTRACT

In this study, we describe a technique that used a machine learning (ML) approach to classify four (4) different cotton leaf varieties namely; BS-15, S-32, Z-31, and Z-32. Each variety of cotton leaves were collected from 500 Farmers. These image datasets are captured by using the cell phone camera in the open agricultural field area, and every image was captured from both sides (Front and Back) of the cotton leaf. Each variety of cotton has used over 300 (150 Front Side and 150 Back Side of the leaves) leaf images and the total calculated cotton leaves are 1200 (300 x 4) as leaf image samples. These sample datasets have analyzed through image preprocessing and image segmentation process. Each image was employing four different non-over-lapping regions of interest (ROI's) and calculated a total of 4800 (1200 x 4) ROI's. The acquired datasets are employed different machine learning features such as Scalability, Texture, Spectral, Binary, Histogram, Rotational, and translational (R-S-T). A total of fifty-seven (57) machine learning features were evaluated on each ROI and a total calculated 273,600 (4800 x 57) features. Furthermore, the Correlation-Based Feature Selection (CFS) genetic algorithm technique was employed for feature optimization. It has been evaluated 22 optimized features and applying different machine learning (M-L) classifiers namely; K-Nearest Neighbor (K-NN), K*, Random Forest (RF) Tree, and Naive Bayes (NB) Tree. The resulting accuracy produced by K-NN presented is 98.9167% on (512 x 512) ROI's. The individually overall result accuracy dataset values by using K-NN classifier on the four varieties of cotton leaf namely; BS-15, S-32, Z-31, and Z-32 were evaluated 97.83%, 99.50%, 99%, and 99.33%, respectively.

KEYWORDS: Cotton Leaves, image & classification process, feature optimizing, machine learning process

1. INTRODUCTION

In the field of the agricultural system, machine learning plays an important role in the classification of plant categories [1] [2]. The structure of leaves are the primary or main section of plants and every leaf area is different according to size, shape, texture, color, and structure. This primary section was very helpful for the classification of different plant leaf varieties [3], and cotton leaf is one of them. Nowadays, the identification of cotton (Gossypium) leaves was performed by different researchers in Southern Asia (sub-continental) countries, and Pakistan is one of them. Gossypium is the most important source or crop of vegetable oil, pollen fiber, and produce high-quality protein meals for livestock [4]. Cotton (Gossypium Hirsutum L.) is also called the 'White Gold' and 'Fiber King'. Gossypium is also a most beneficial source for the textile industry, and 60 million (approx.) peoples are engaged with the cotton industry and agricultural system [5]. In Pakistan, cotton (Gossypium) is the second-largest crop cultivated locally and ranked fourth (4th) cultivation country of Gossypium world-wide. Pakistani Gossypium is highly produced fiber and vegetable oil, which is most beneficial for the oil industry [6].

Gossypium crop is sowing on the area of 2527 (hectares) during 2019-20, and it produces 9.178 billion cotton bales [7]. Gossypium is cultivated among the Sindh and Punjab (Pakistan) provinces [8]. Punjab produces 82% of cotton, and the Sindh production rate is 16% [9]. Cotton (Gossypium Arboreum L.) variety was cultivated in Monjadharo Sindh (Pakistan) in early 6000 B.C [10]. The different genotypes of cotton varieties were cultivating locally such as BS-15, S-32, Z-31, Z-32, CIM-610, CIM-608, Bt.CIM-63 and Bt.CIM-600 etc. These different genotypes of Gossypium are very famous categories and nearly 90% of growers are produced cotton by using the CIM-608 variety of Gossypium in Pakistan [11]. In [12], Pakistan is the biggest exporter of cotton, and it is also earned a million dollars in the fiber and oil industry.

Therefore, we need an automated or artificial

intelligence system that recognized automatically different varieties of Gossypium plants, rather than the complete area of the cotton leaf. The open eye or visual-based identification of the cotton leaf is very challenging due to the same features of leaf such as texture, color, geometrical, size, and shape feature between its variant cotton varieties. The confusing arguments about cotton varieties describe major problems such as the high use of pesticides (chemicals) and urea (Fertilizer) without proper study or knowledge and research. That is why the cotton production rate is a little low based on purity and quantity. Here, we discuss three different phases for resolving the following issues

a) the pre-processing technique was employed on raw image dataset for evaluation of features;
b) employed different methodologies for feature optimization;
c) different machine learning classifiers were implemented on available optimized features datasets.

2. LITERATURE REVIEW

Seed quality is the most important element for the cotton cultivation process in the agricultural system. The procedure of seed storage was an essential part to secure seeds from deterioration. Non-Consumer-Seed-Growers (NCSG) and Consumer-Seed-Growers (CSG) were used for seed storage in district Vehaari of Pakistan. CSG techniques are more beneficial for seed storage as compared to NCSG growers. [13]. Cotton-Leaf-Curl-Virus (c-l-c-v) diseases were described in Punjab, Pakistan early 1990s. Four different cotton varieties were used for the experimental process namely; FH-900, CIM-448, FH-901, and CIM-1100. The random Amplified Polymorphic DNA (RAPD) approach was describing the genetic distance as only 3% [14]. The choice of cotton seeds verity was based on geographical environment, crop requirement, and agricultural resources. Two different varieties of cotton are used in the cultivation process such as Zhaongmian-16 and J-4B. The growth of cotton leaf functionally accuracy is 15.9% - 10.7% and 12.3% - 10.5% based on two years, respectively [15].

The cotton (Gossypium) leaf verity is affected by Mosaic disease, which is a cause of the reduction in plant growth and genetic characteristics. CRIS-168 variety of cotton is identifying the Mosaic disease and c-l-c-v virus. Mosaic diseases described the different issues on the plant are morphology with following rate such as 25% effected in length, 24.1% in height of the plant and 37.5% of sympodia. The different characteristics are improved slightly inaccuracy of result namely; Uniformity ratio, fiber fineness, and maturity index increase about 20.5%, 14.4%, and 0.9%, respectively [16].

C-L-C-V disease has been the most serious issue of the cotton cultivation process in Pakistan. It is also identified in some areas of plants in Pakistan (1967). Climate and temperature changed were also infected the cotton crops. Fiber index, length, height, fineness, and maturity ratio also affect cotton crop yield and production. Different methods were employed to reduce C-L-C-V diseases along with fertilizer (urea), control insets, agronomic and biotechnology methods [17]. Leaf varieties classification leads to multiple applications. Many researchers describe enormous techniques for pattern recognition in the leaf of the plant. CNN based techniques were employed on different Indian leaves varieties [18]. Ethylene Insensitive-3 (EIN-3) and Ethylene Insensitive-3 like (EIL) protein values are managed as a major part of plant development and growth during diverse cultivation environment. EIN-3 and EIL proteins are the main part of ethylene signaling regulators and also helped in cotton production stages [19].

In [20], the researchers describe 32 plants by employing among twelve (12) different features of the leaf. A Probabilistic Neural Network (PNN) approach was deploying on the available 1800 sample image dataset, and it's evaluating 90% result accuracy based on PNN. A genetic algorithm was used for the feature optimization process by using apply fruit. Support Vector Machine (SVM) classifier was achieved a classification accuracy is 98.10% [21]. Two different varieties of fused data of 5 variant types of whets were described for the leaf classification process. These datasets were collected from photographic and radiometric information, it is achieved accuracy result is 93.14% and 96%, respectively [22]. The authors [23] introduced a novel based dataset of features, which describe the shape and image color feature. Artificial Neural Network (ANN) approaches were employed on 8 different pepper seed varieties and evaluate 84.94% result accuracy. The different leaves of citrus have been used to describe texture-based features on the color image. Discriminant Analysis (DA) technique deployed for feature optimization, and its observed 95% accuracy based on available features [24].

In this study [25], Deep Learning (DL) technique was used for leaf feature classification. That also describes a new model technique of hybrid, which helps in exploring the hidden information about leaf features. DL approach was improving feature identification in image processing. The shape and texture-based segmentation or object identification techniques were deployed on mango variety detection [26]. The accuracy of the result was compared to an existing one and same technique report in [27], author recognize different method for feature extraction such as binary, histogram and texture. The authors [28] describe, two dimensional (2D) images of the leaf dataset were used for the plant classification process. It's also used two different methods for feature optimization such as one-dimensional (1-D) and two-dimensional (2-D) based on the classifier of Bagging. The five different machine learning (ML) techniques were employed on Flavia's publically colored leaf images dataset for the experimental process namely; Principal Component Analysis (P-C-A), Direct Linear Discriminate Analysis (D-L-D-A), LDA + PCA, 2-DLDA, and 2-DPA. It was evaluated that, the 2-DLDA and 2-DPCA result accuracy much better than all other techniques. In this

study [29], many researchers have adopted different machine vision (M-V) techniques for the varieties identification based on ten different varieties of potatoes. PCA approach was deployed for the feature optimization process. It is observed that the Artificial Neural Network (ANN) was achieved some better accuracy in the result.

It was concluding based on the above literature discussion, huge research or experimental work, it was adopted for the identification of cotton varieties by using different feature techniques. So, there is necessary required a reliable or automatic system for the classification of cotton leaf varieties, that is cost-effective, efficient in work, reliable, and time-saving. The objective of this research or study was to describe the novel technique for the identification of cotton leaves by using optimized features.

3. MATERIALS AND METHOD

3.1. Data Collection

The proposed technique is employed on the Pakistan cotton leaf samples [30]. These leaf datasets were collected from 500 Farmers' cotton land of agricultural and 2 tehsil of District Bahawalpur Punjab (Pakistan) such as Bahawalpur and Ahmed Pur East zone, which is described in Table 1 and Figure 1. The geographical or global position location of cotton was field area in Bahawalpur (29° 23' 44" (North) latitude and 71° 41' 1" (East) longitude) [31] and Ahmad Pur East (29° 5' 13" (North) latitude and 71° 16' 61" (East) longitude).

3.2. *Image Capturing and Acquisition*

In this experimental process, about 500 growers of cotton field plants with 600 fresh leaves samples have been collected for every single type of cotton. The leaves samples were managed or found in average height and width of 17.5 $\pm$ 0.7 cm and 9.05 cm. The sample image acquisition was performing by using a cell phone (OPPO) camera with a resolution of 1080 x 1920 pixels. A cell phone camera has been adjusting vertically at the height of 2 feet on each leaf image, and eliminate shadow leaf. So, all image samples were collected at the open climate and temperature of 40^0 C (July, 2020) from 1:00 (P.M) to 3:00 (P.M) according to the Pakistani time zone.

All images were captured from the front and backside of 4 different Cotton leaves varieties, and evaluate 600 (150 x 4) front and 600 (150 x 4) backside image of leaf, which is shown in figure 2. A total calculated of 1200 (600 x 2) cotton leaves image datasets were acquired.

3.3. *Leaf Area Method*

Information about leaf area and Index of Leaf Area (ILA) has been collected from the grower's cotton land and it's calculated by using the following method [32].

$$\text{Area of Leaf } (ab^2) = M \times N \times K$$

Table 1: Different Area or Zone for Data Collection

Province	District	Area / Tehsil	Union Area or Council (Owner)	Rural / Village	Growers (Farmers)	Cotton Leaves	Total Leaves Sample
Punjab	Bahawalpur	Bahawalpur	Khan Pur (M. Shehzad)	Basti Punraan	45	55	1200 (600 Front Side, 600 Back Side of leaf)
				Basti Chuhan	70	100	
			Noor Pur (Mashood)	Noor Pur Pull	40	50	
				Wahi Jogiyaan	45	60	
			Mubarak Pur (M. Imran)	Basti Kaloo Khan	80	85	
		Ahmed Pur East	Dera Nawab Sahib (Abdullah)	Basti Bhatiyan	60	75	
				Basti Balouch	50	60	
			Sukhail (Rashid Dharala)	Basti Khamisa	40	50	
				Basti Dharala	70	65	

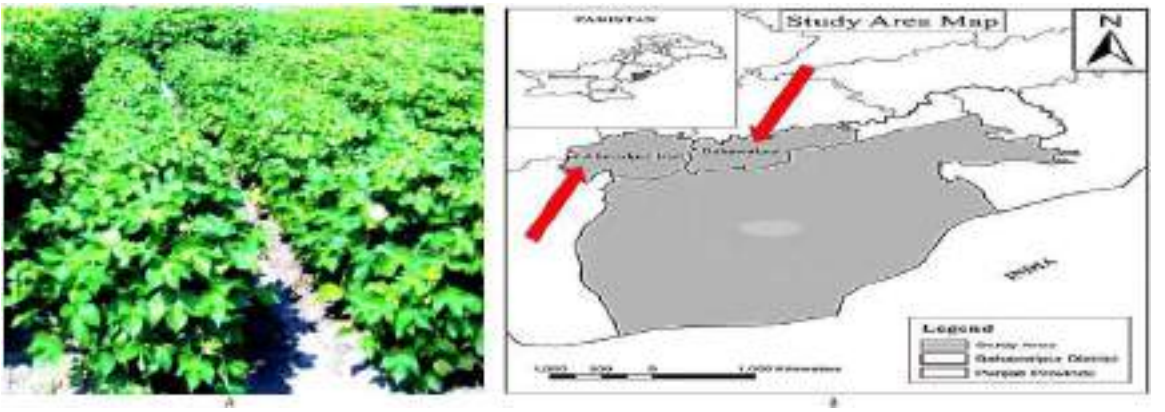

Figure 1: (A) Field Area of Crop Image and (B) Geographical or Study Area Map in Pakistan



Figure 2: Gossypium Leaf Variety Image Dataset; (A) BS-15 (B) S-32 (C) Z-31 (D) Z-32

Here, M is the length of the leaf; N is the highest width of the cotton leaf; and K is the value of correction factor of the cotton leaf is 0.75.

Index of Leaf Area (ILA):

$$ILA = \frac{Total\ Area\ of\ Leaf\ from\ all\ Cotton\ Leaf}{Field\ Area\ Land\ Where\ Data\ Collected}$$

3.4. Image Preprocessing

In this section, initial processes on images were described in the layout to obtain feature optimization, and classification techniques are very important procedures or ways [33]. Image accuracy or the processing is difficult as it directly expresses our results of feature optimization and classification [34]. So, we describe a better technique during the image processing step.

Furthermore, all leave images of cotton were arranged in the same shape (Width and Length) by using image resizer software. All the collected images of cotton leaves were resized to resolution 512 x 512 pixels and employed a Median filter to evaluate the prominent area of the cotton leaf

3.5. *Selection or Segmentation ROI*

In this proposed technique, the most prominent area of the cotton leaf is considered for ROI selection with four (4) non_over_lapping ROIs. The segmentation of cotton leaf is done by using gray_scale_level image, and it's described statistically feature [35]. To classification of the leaf as ROI, the real color is Red, Green, and Blue (RGB) image is exchanged into H_S_V color space structure or model.

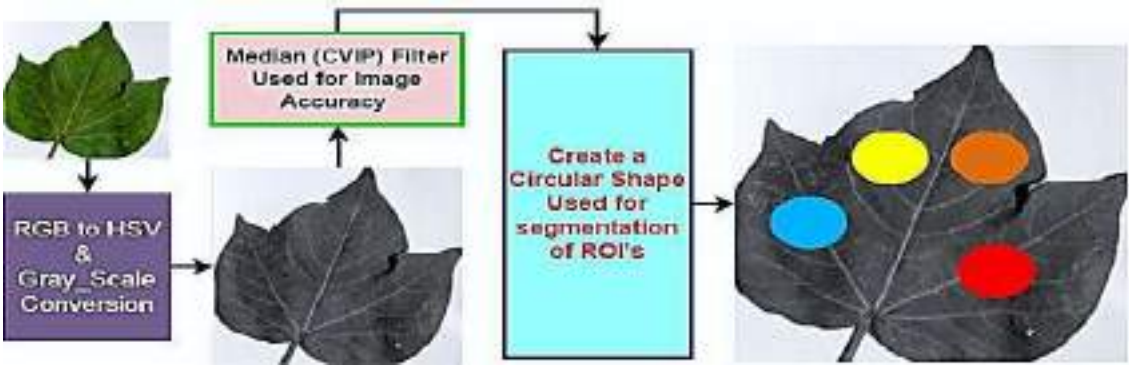

Figure 3: Block-Based Diagram for ROI selection Procedure

Three different coordinate values were evaluated such as shade, brightness, and shade [36]. ROI is developing in gray_scale_level describe in algorithm 1, which describes the all necessary steps of ROI selection and segmentation. RGB image conversion into H_S_V or gray_scale_level is shown in Figure 3.

Algorithm 1: Segmentation of Leaf ROI

Phase 1: Input the RGB based image $I_{R\ G\ B}$ of the cotton leaf

Phase 2: RGB leaf image $I_{R_G_B}$ is converted into gray_scale_level $P_{_\ gray}$.

$P_{_\ gray}$ = (0.299 * Red) + (0.59 * Green) + (0.114 * Blue)

1: Leaf co-ordinate is evaluate by employing gray_scale_level

2: Evaluate the circular shape of image by use equation

$$(a - a_0)^2 + (b - b_0)^2 - M^2 = 0$$

Here, $(a_0 - b_0)$ is the mid-point value of the circle and (a, b) are expected coordinate

Phase 3: Optimize ROI's are evaluate from $P_{_\ gray}$ by employing binary values

ROI = $I_{R_G_B}$ – transpose (Binary_Values)

Phase 4: H_S_V value is calculated form ROI

$$Red' = {Red}/{255}$$

$$Green' = {Green}/{255}$$

$$Blue' = {Blue}/{255}$$

$$D_{maxi} = maxi\ (Red', Green', Blue')$$

$$D_{mini} = mini\ (Red', Green', Blue')$$

The value conversion method is given as:

Δ = Delta = $D_{maxi} = D_{mini}$

$$H = \{60 * \left(\frac{Green' - Blue'}{Delta} * MOD2\right)\}$$

$$H = \{60 * \left(\frac{Blur' - Red'}{Delta} * +2\right)\}$$

Phase 5:

$$H = \{60 * \left(\frac{Red' - Green'}{Delta} + 4\right)\}$$

$$S = \{\ O\ D_{maxi}\ = 0\ \}$$

$$V = D_{maxi}$$

Extracted ROI

4. PROPOSED METHODOLOGY

At the start, the image dataset of cotton leaves are input and employing the preprocessing procedure. Furthermore [37], the Transductive Parameter Transfer (TPT) approach was used for image segmentation and machine learning classification successfully, this approach was evaluated in three phases, which is shown in Algorithm 2.

In Phase 1, the ROI selection procedure is employed for image segmentation. The complete procedure for ROI selection is discussed in Algorithm 1.

In Phase 2, we employed the correlation feature selection (CFS) genetic search algorithm for feature optimization or feature reduction procedure. Phase 3 and finally, these optimized feature datasets were employed different machine learning classifiers for output or result.

5. FEATURES SEGMENTATION OR ACQUISITION

The binary, rotational, spectral, translational (R_S_T), first and second-order (GLCM), and scalability features were deployed on each cotton leaf sample image dataset [38]. These extracted features were based on 28 binary based on 10 pixels (width and height) distance, 6 spectral (3 sectors and 3 rings) features, 7 R_S_T features, 5 first order

(texture), and second-order (GLCM) features [39], [40].

A total of fifty-seven (57) features were extracted on each ROI's, and it's calculated 273600 (57 x 4800) features dataset on available image sample. Computer Vision and Image Processing (CVIP) tools were deployed for feature extraction.

Algorithm 2: Proposed Transductive Parameter Transfer (TPT) Method	
Input Data:	$A_1^c, \dots\dots\dots, A_n^c, T^x$, the parameter of $\tau_B, \tau_C, \in$.
Phase 1:	ROI selection procedure, which is describe in Algorithm 1.
Phase 2:	Feature optimization or computation procedure $\theta_j, x_i = 1, 2, \dots\dots\dots, n$
Phase 3:	Evaluate the machine learning (ML) classifiers and calculate the target ML classifiersM^t. $M_j^t = r\,(x_j, T^x)$ Given M^t, evaluate $\theta^t = \hat{f}\,(T^x)$
Output:	θ^t (ML Results)

All the statistical feature segmentation was discussed below:

5.1 Histogram or 1ST Order Features

Histogram (1ST order) features are calculated by applying the intensity value on each pixel. Histogram features are also called the 1ST order feature. The probability of the Histogram is L (m), which is shown in equation 1.

$$\boldsymbol{Histogram} = \mathbf{L\,(m)} = \frac{\mathbf{K\,(a)}}{\mathbf{N}} \qquad (\mathbf{1})$$

Here, N evaluates the total value of a pixel, and L (m) represents all instances of gray_scale_level value of m. First-order (Histogram) features were calculated five different statistical features namely; Entropy, Mean, Energy, Skewness, and Standard Deviation. **Mean** is evaluate the dark and brightness value of an image, which is shown in equation 2.

$$\boldsymbol{Mean} = \bar{\boldsymbol{s}} = \sum_{s\,=\,0}^{n-1} \boldsymbol{sn(s)} = \sum_{p}\sum_{q} \frac{\boldsymbol{M\,(p,q)}}{\boldsymbol{M}} \qquad (\mathbf{2})$$

Here, n represents the value of gray_scale_level from range 0 to 255, and m and n describe the pixel value of horizontal (rows) and vertical (columns) area. **Standard Deviation (SD)** is evaluating the contrast of the image; this is shown in equation 3.

$$\boldsymbol{SD} = \boldsymbol{\sigma_n} = \sqrt{\sum_{n=0}^{D-1} (\boldsymbol{n} - \bar{\boldsymbol{n}})^2\, \boldsymbol{D\,(n)}} \qquad (\mathbf{3})$$

Skewness is describing the center value of symmetry or image, which is shown in equation 4.

$$\boldsymbol{Skew} = \frac{\mathbf{1}}{\boldsymbol{\sigma_n^3}} \sum_{n=0}^{D-1} (\boldsymbol{n} - \bar{\boldsymbol{n}})^3\, \boldsymbol{D\,(n)} \qquad (\mathbf{4})$$

Energy is describe the value of gray_scale_level, and is describe in equation 5.

$$\boldsymbol{Energy} = \sum_{n=0}^{D-1} [\boldsymbol{D\,(n)}]^2 \qquad (\mathbf{5})$$

Entropy is calculating the contents of image data, which is shown in equation 6.

$$\boldsymbol{Entropy} = -\sum_{n=0}^{D-1} \boldsymbol{p\,(n) log_2 [p(n)]} \qquad (\mathbf{6})$$

5.2 Histogram or 2ND Order (Texture) Features

The 2ND order feature is also called statics texture feature or gray-level co-occurrence matrix (GLCM) based feature. Five different texture features are calculated with 4 dimensional (0^0, 45^0, 90^0, and 135^0) areas by using a five-pixel distance. These features are energy, correlation, entropy, inverse-difference, and inertia.

Energy is evaluating the gray_level value in equation 7.

$$Energy = \sum_{a}\sum_{b} (D_{ab})^2 \qquad (7)$$

Correlation calculated the similarity between pixels

with pixel distance. It is evaluated in equation 8.

$$\mathbf{Corr} = \frac{\mathbf{1}}{\boldsymbol{\sigma_c \sigma_d}} \sum_{\mathbf{i}} \sum_{\mathbf{j}} (\mathbf{i} - \boldsymbol{\mu_c})(\mathbf{j} - \boldsymbol{\mu_d})\mathbf{D_{ij}} \qquad (8)$$

The total content of the image is described by using **Entropy**, which is shown in equation 9.

$$\boldsymbol{Entropy} = -\sum_{c} \sum_{d} \boldsymbol{K_{cd} log_2 K_{cd}} \qquad (9)$$

The inverse difference is also called local homogeneity of image, which is described in equation 10.

$$\boldsymbol{Inver - Diff} = \sum_{i} \sum_{j} \frac{\boldsymbol{K_{ij}}}{\boldsymbol{|i-j|}} \qquad (10)$$

Inertia is evaluating the contrast of image, which is shown in equation 11.

$$\boldsymbol{Inertia} = \sum_{c} \sum_{d} \boldsymbol{(c-d)^2 K_{cd}} \qquad (11)$$

5.3 Binary Feature

The binary feature is also called image shape-based feature such as a least second moment of an axis, area, the area of the center, number of Euler, thinness, projection, and aspect of ratio. Twenty Eight (28) binary shaped features were evaluated by using height and width (10, 10) pixels of projection. The area of the jth object was described in equation 12.

$$\boldsymbol{A_j} = \sum_{\boldsymbol{m=0}}^{\boldsymbol{height-1}} \sum_{\boldsymbol{n=0}}^{\boldsymbol{width-1}} \boldsymbol{J_j\,(m,n)} \qquad (12)$$

The rows and column coordinates have calculated the centroid of the image by using the jth object, which is shown in equations 13 and 14.

$$\boldsymbol{\overline{A}} = \frac{\mathbf{1}}{\boldsymbol{P_j}} \sum_{\boldsymbol{p=0}}^{\boldsymbol{height-1}} \sum_{\boldsymbol{q=0}}^{\boldsymbol{width-1}} \boldsymbol{A\,P_j\,(p,q)} \qquad (13)$$

$$\boldsymbol{\overline{D}} = \frac{\mathbf{1}}{\boldsymbol{P_j}} \sum_{\boldsymbol{p=0}}^{\boldsymbol{height-1}} \sum_{\boldsymbol{q=0}}^{\boldsymbol{width-1}} \boldsymbol{D\,J_j\,(p,q)} \qquad (14)$$

In this binary feature also evaluated the orientation, perimeter Euler number and aspect ratio of image.

5.4 Translation, Rotation and Scaling (R_S_T) Features

R_S_T dataset or values are the segmentation of invariant features which are evaluated by employing architectural knowledge of histogram, that is described seven R_S_T invariant segmentation features. Spectral features are based on the frequency domain and it describes successfully when an image is classifying on texture values. Spectral features were evaluated in the different regions and these are based on sector and region, which is defined in equation 15.

$$\boldsymbol{S.R.P} = \sum_{\boldsymbol{\sigma\, \varepsilon\, region\ area}} \sum \boldsymbol{|M\,(m,n)^2|} \qquad (15)$$

5.5. K- Nearest Neighbors (K-NN) Classifier

Different ML classifiers were available and K-NN is one of them [41]. K-NN algorithm is the simplest and very helpful for the deep learning process [42]. In the K-NN algorithm, an element of the object is identifying by using the maximum number of its nearest or neighbors. The element of the object has continuously evaluated to the class, which was the nearest value between K-NN, here the value of K is positive, which represents a small value. If the value of K is 1 (K=1), then the nearest value of neighbors is assigned to the selected class.

The first employed algorithm of K-NN was evaluating some statistical notation N= $(a_{j,}\, b_j)$, j= 1,2,3,4,…, M has described the set of training values. Here a_j is describe the dimensional vector of features, and $b_j \in (-1, +1)$ was evaluated with the label of class. Simply we consider this as a classification of binary class. The previous sample labeled as a training value of N. So, the algorithm of K-NN builds a sub-region locally ($D\,(a) \in R^c$) as input, that is described at the point of a. The estimated region or area D (a) measure the nearest point of training is a, which is shown in equation 16.

$$\boldsymbol{D\,(a)} = [\,\boldsymbol{\hat{a}} \mid \boldsymbol{P\,(a,\hat{a})} \leq \boldsymbol{S_k} \qquad (16)$$

Here S_k is the kth is static order of $[D\,(a,\hat{a})]_1^M$, and $D\,(a,\hat{a})]$ is the metric distance and k is denoted the value of the region in sample D(a). The algorithm of K-NN described the probability of statistics $T\,(\,n\,|a\,)$, which is shown in equation 17.

$$\boldsymbol{T\,(\,n\,|a\,)} = \frac{\boldsymbol{T\,(\,a\,|\,n\,)T\,(\,n\,)}}{\boldsymbol{T\,(a)}} = \frac{\boldsymbol{S_k}}{\boldsymbol{S}} \qquad (17)$$

For the basis of a, the decision of h (a) is evaluating the data value of S_k and electing the class, which describe the value of the height of S_k, which is shown in equation 18.

$$\boldsymbol{h\,(a)} = \begin{bmatrix} \boldsymbol{-1\,S\,[\,K=1\,] \geq S\,[\,n=1\,]} \\ \boldsymbol{1\,S\,[\,K=1\,] \geq S\,[\,n=-1\,]} \end{bmatrix} \qquad (18)$$

Therefore the heights probability value is deployed in the algorithm of K-NN. Here, $b_j \in (-1, +1)$ is evaluating the binary problem of classification, which is shown in equation 19.

$$\boldsymbol{h\,(a) = nsg\,\left(averg_{a_j} \in D(a) b_j\right)} \qquad (19)$$

6. FEATURE ACQUISITION OR SELECTION

It is perceived that the evaluation of fifty-seven (57) variant features on each ROIs by using cotton leaf sample images. It is not satisfactory of ML classification process, and to manage this 273,600 large amount of features are taken more processing time for model building [43]. So, this problem is resolved by using a feature optimization process as possible, as discussed in [3]. So, that is the easiest way for the classification process with a minimum rate of error. The Principal Component Analysis (P-C-A) technique described good results accuracy on repeated data. But this optimized feature dataset was not providing better accuracy on the whole dataset [44].

Furthermore, we employed the represent correlation-based feature selection (C-F-S) genetic search algorithm to express optimized features from a huge size dataset, which is shown in equation 20

$$\boldsymbol{M_k = \frac{M\overline{\mu}_{aB}}{\sqrt{M + M\,(S-1)\overline{\mu}\,aB}}} \qquad (20)$$

After employed CFS genetic search on the large size dataset, it evaluated 22 optimized features on each cotton leaf image, which is shown in table 2. It was produced that 273600 (4800 x 57) huge features were minimized into 105,600 (4800 x 22) features.

Finally, different machine learning classifiers are plugged into available optimized features. 10-k fold cross-validation technique has been employed to eliminate the difficulties about the testing and training ratio. The experimental layout on the framework of cotton types is described in figure 4. In this experimental process, there are different ML classifiers were applied to our proposed features dataset such as K-NN, K*, Random Forest Tree, and Naive Bayes Tree.

It has been evaluated that according to the above discussion classifier evaluates a very low accuracy (below 90%) result on the backside image dataset of the cotton leaf. However, the same classifier set was employed on the front side image dataset of cotton leaf and it is showing a better accuracy result (above 90%).

Table 2: CFS Genetic Algo Based Optimized Features

Sr. No.	Optimized Features
1.	Objects_row_coordinate
2.	Objects_column_coordinate
3.	Centroid (row_column)
4.	Orientation (Axis_of_least_second_moment)
5.	Histro-Standard-Deviation
6.	Histo-Skew
7.	Histo-Entropy
8.	Inertia-range
9.	Inertia-average
10.	Texture-energy-range
11.	Texture-energy-average
12.	Correlation-range
13.	Inverse-diff-average
14.	Inverse-diff-range
15.	Spectral-DC
16.	Texture-entropy-range
17.	Texture-entropy-average
18.	Ring1
19.	Ring2
20.	Ring3
21.	Sector2
22.	Sector3

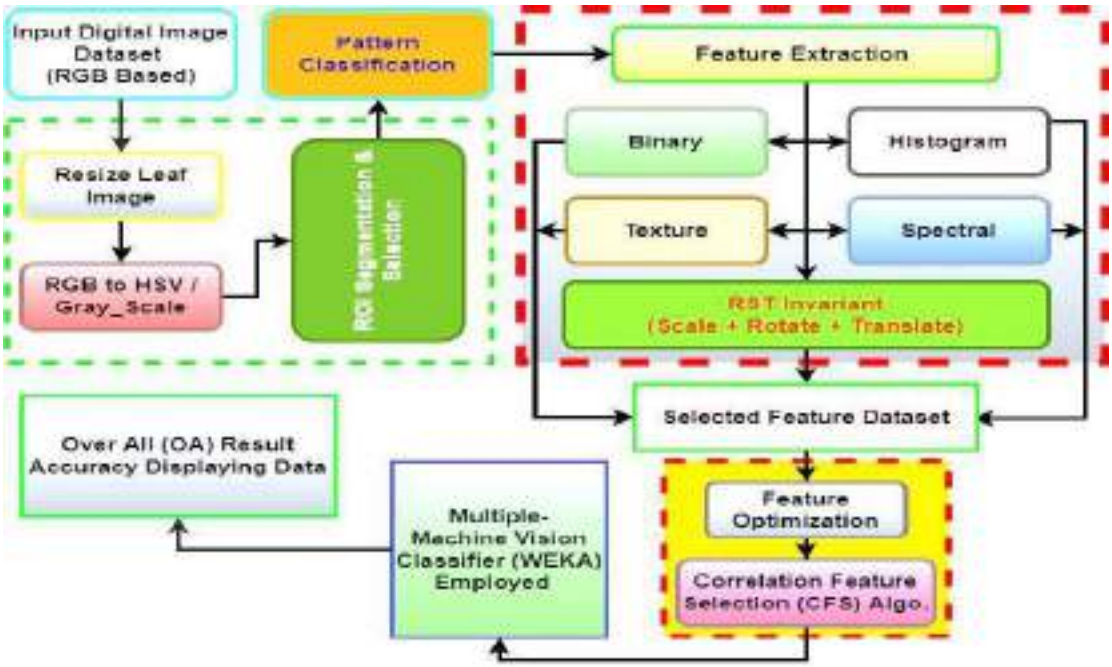

Figure 4: Feature Dataset Layout or Framework of Cotton varieties classification based on Leaves

7. RESULT AND DISCUSSION

In this proposed study, we employed 4 main machine learning classifiers such as K-Nearest Neighbor (K-NN), K*, Random Forest Tree, and Naive Bayes Tree by using 4 Gossypium varieties with different features dataset. Initially, we employed the 10-k fold cross_validation technique for the classification process. After this, for the creating ROI's of used 60 x 60-pixel size and with image resolution 512 x 512 pixels. Here we discuss some important factors or an element, that's helped us for result accuracy.

1. The acquired dataset of the cotton image is healthy and clear
2. Pre-processing or image accuracy phases and Transductive Parameter Transfer (TPT) method of image segmentation helped us for image clarification
3. A feature optimization algorithm was used for extracted prominent features

In section 6.1, the Backside image dataset of cotton leaf-based employed different machine learning classifiers (ML) were discuss and section 6.2, the Front side of image dataset of cotton leaf-based employed ML classifiers were discussed.

7.1. Back Side Image Dataset Result

For this proposed experimental process, WEKA 3.6.12 software is employed as a tool and used a different algorithm along with it. The resulting accuracy was described by using 52800 (2400 x 22) total optimized features dataset as input and four different machine learning classifiers were employed for the experimental process such as K-Nearest Neighbor (K-NN), K*, Random Forest Tree, and Naive Bayes Tree, and shown following accuracy result of 94.375%, 93.8333%, 92.875%, and 84.25% respectively. It has been shown that only K-NN achieved better accuracy results, which is shown in Table 3. The individually overall result accuracy of the 4 cotton varieties namely; BS-15, S-32, Z-31 and Z-32 were 93.33%, 95.33%, 92.67% and 96.17% respectively, which is shown in figure 5 and the miss classification result was shown in figure 6. The confusion-matrix (C-M) of K-NN classifier on ROI's image resolution 512 x 512 of four cotton leaves varieties were described in Table 4.

Table 3: Performance of Accuracy Result of Different Machine-Learning Classifiers (Back Side Image of Cotton)

Classifiers	Kappa-Static Value	TP Values	FP Values	R.O.C Values	R-M-S-E Values	Time (Second)	Miss (%) Classify	(%) Accuracy
K-NN	0.925	0.944	0.019	0.963	0.1676	0.03	5.625	94.375%
K*	0.9178	0.938	0.027	0.992	0.1608	0	6.1667	93.8333%
Random Forest	0.905	0.929	0.024	0.987	0.1986	3.25	7.125	92.875%
Naive Bayes	0.79	0.843	0.052	0.939	0.2578	20.55	15.75	84.25%

Table 4: Confusion-Matrix (C-M) of K-NN classifier (Back Side Image of Cotton)

Cotton Classes	BS-15	S-32	Z-31	Z-32	Total Images	(%) Accuracy
BS-15	560	10	27	3	600	93.33
S-32	17	572	2	9	600	95.33
Z-31	32	3	556	9	600	92.67
Z-32	1	7	15	577	600	96.17

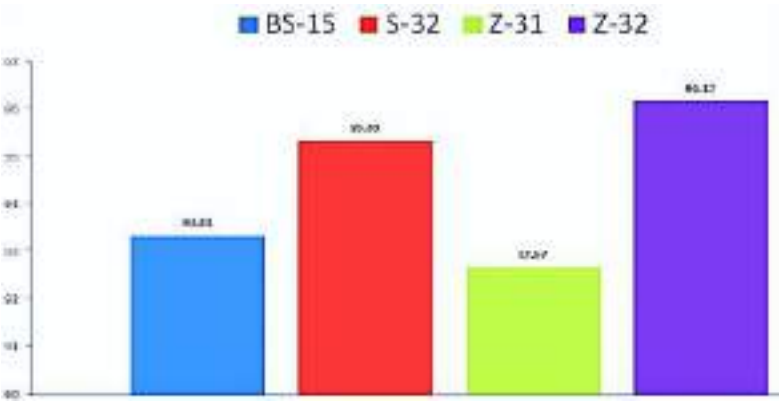

Figure 5: Result Accuracy of Cotton Leaves Image classification on Four Varieties (Back Side Image of Cotton)

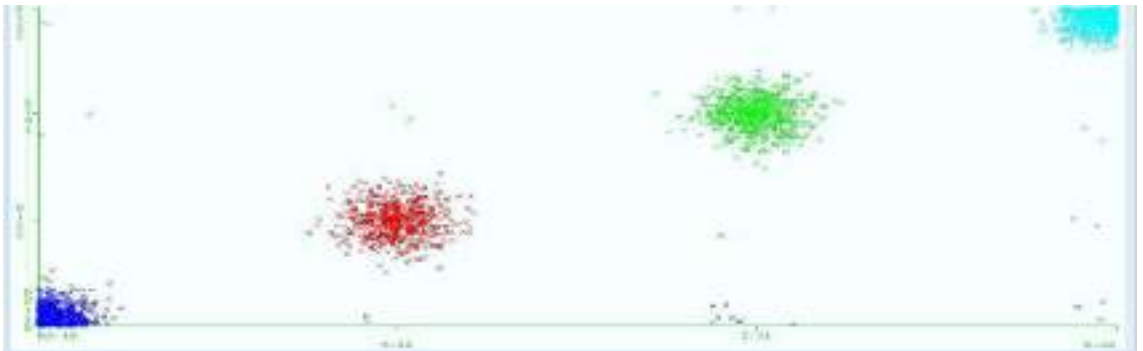

Figure 6: Miss Classification of Result on K-NN based Classifier (Back Side Image of Cotton) and Color Represents are Blue= BS-15, Red= S-32, Green = Z-31 and Light Blue = Z-32

7.2. Front Side Image Dataset Result

To achieve better accuracy results, the same dataset of machine learning classifiers was also employed on the cotton leaves image dataset with 512 x 512 ROI's. The resulting accuracy is being described by using 52800 (2400 x 22) total optimized features dataset input and four different machine learning classifiers were employed for an experimental process such as K-Nearest Neighbor (K-NN), K*, Random Forest Tree and Naive Bayes Tree, and shown following accuracy result of 98.9167%, 98.2083%, 97.0417%, and 90% respectively. It is showed that only K-NN achieved a better accuracy result, which is shown in Table 5. The individually overall result accuracy of the 4 cotton varieties namely; BS-15, S-32, Z-31, and Z-32 were 97.83%, 99.50%, 99%, and 99.33% respectively, which is shown in figure 7 and the miss classification result was shown in figure 8. The confusion-matrix (C-M) of the K-NN classifier on ROI's image resolution 512 x 512 of four cotton leaves varieties were described in Table 6.

Finally, the K-NN feature is given better overall accuracy results based on open climate and noisy image dataset, as compared to the following approach [12]. Complete descriptions of the following techniques with existing techniques were shown in Table 7. Our presented system was the ability to discriminate the cotton varieties leaves based on multiple features, which was very beneficial for growers and industrial sectors to accurately identifying the exact variety of cotton. It is an efficient and robust technique to eliminate human error. All the available digital cameras were their resolution and change in the pixel resolution of cameras may become the variation in results. The image dataset used in this experimental process was very less. We positively encourage to researcher and researchers in developing an open or public distribution network for agricultural areas or datasets, where we describe the problems, that is related to the agricultural study.

Table 5: Performance of Accuracy Result of Different Machine-Learning Classifiers (Front Side Image of Cotton)

Classifiers	Kappa-Static Value	TP Values	FP Values	R.O.C Values	R-M-S-E Values	Time (Second)	Miss (%) Classify	(%) Accuracy
K-NN	0.9856	0.989	0.004	0.889	0.0735	0	1.0831	98.9167%
K*	0.9761	0.982	0.006	1.00	0.0846	0	1.7917	98.2083%
Random Forest	0.9606	0.97	0.01	0.998	0.1459	3.04	2.9583	97.041%
Naive Bayes	0.8667	0.9	0.033	0.965	0.2074	18.67	10	90%

Table 6: Confusion-Matrix (C-M) of K-NN classifier (Front Side Image of Cotton)

Cotton Classes	BS-15	S-32	Z-31	Z-32	Total Images	(%) Accuracy
BS-15	587	2	8	3	600	97.83
S-32	0	597	1	2	600	99.50
Z-31	2	2	594	2	600	99
Z-32	2	0	2	596	600	99.33

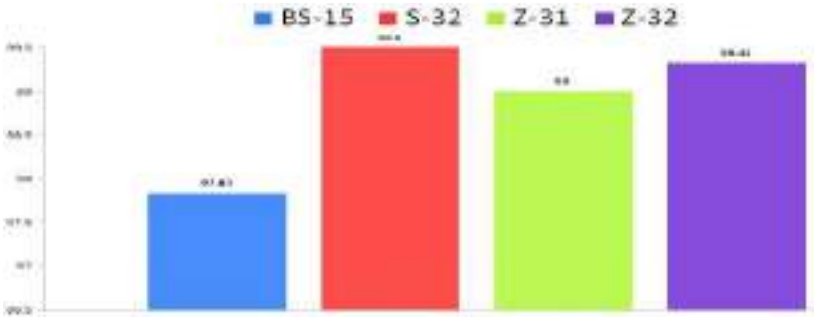

Figure 7: Result Accuracy of Cotton Leaves Image classification on Four Varieties (Front Side Image of Cotton)



Figure 8: Miss Classification of Result on K-NN based Classifier (Front Side Image of Cotton) and Color Represents are Blue= BS-15, Red= S-32, Green = Z-31 and Light Blue = Z-32

Table 7: Comparison Result Accuracy between Proposed and Existing Approach

Source	Features	Algorithm	Results / Overall Accuracy
[2]	ANN, MAPE	CNN, LSTM, DNN	60%
[4]	SVM, LR, PLS-DA	PCA, ResNet, CNN	80%
[5]	MLP, SVM, K-NN	GLCM, HoG, HIST	96%
[3]	MLP, J48, NB	CFS Best Search Algorithm	98%, 97.5%
[18]	SVM, ANN, MLP	KNN ,Robust Algorithm	43.9%, 88%
[12]	NIGAB, ABRI	CRS, MTN, CCRI, APTMA	39.4%
[20]	PNN	PCA, ANN	90.312%
Proposed Approach	K-NN, K*, RF, NB	CFS Genetic Search Algorithm	98.2583%

8. CONCLUSION

In this proposed study, we described the classification of four cotton varieties using multiple features leaves dataset. For this experimental process, four different machine learning techniques were employed successfully namely; K-Nearest Neighbor (K-NN), K*, Random Forest Tree, and Naive Bayes Tree. Different methodologies were applied in the description of the image dataset for the classification of the cotton leaves. The entire machine learning classifier described the efficient resultant accuracy but the K-NN's resultant accuracy was better as compared to the other machine learning classifiers' results by achieving 98.2583% accuracy on the four cotton varieties. The achieved overall accuracy results evaluate that the proposed technique is efficient and can be employed in real-life applications. In this proposed study, it was evaluated

that the feature optimization technique is more helpful for result accuracy and time-saving procedure. In the featured study, we used the variation among the texture features dataset with the addition of extra attributes to the same cotton leaves.

Acknowledgments: The authors would like to especially thank Dr. Salman Qadri, Ali Nawaz Shah and Muhammad Shehzad for their technical support during this experimental process. Muhammad Abdullah Ismail, Muhammad Mashood, Muhammad Imran and Muddasir Ijaz works as a field assistant for the collection of cotton leaf process.

Funding**:** "This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors".

Author Contributions**:** Tanveer Aslam supervised all the fieldwork of data collection, wrote the manuscript, and description of the available dataset. Dr. Salman Qadri supervised this study process, provides all technical support, and analyzed all datasets.

Conflicts of Interest**:** The authors declare no conflict of interest.

ORCID**:**

Salman Qadri https://orcid.org/0000-0002-3503-6535

TANVEER ASLAM https://orcid.org/0000-0001-5625-6206

REFERENCES

[1] R. Barth, J. IJsselmuiden, J. Hemming and E. V. Henten, "Synthetic bootstrapping of convolutional neural networks for semantic plant part segmentation," *Computers and Electronics in Agriculture,* vol. 161, no. 2019, pp. 291-304, November 2017.

[2] T. V. Klompenburg, A. Kassahun and C. Catal, "Crop Yield Prediction Using Machnie Learning: A Systematic Litreature Review," *Coputer and Electronic in Agriculture Elesvier ,* pp. 1-12, Auguest, 2020.

[3] S. Qadri, S. F. Qadri, M. Husnain, M. M. S. Missen, D. M. Khan, M.-u.-R. A. Razzaq and S. U. , "Machine Vision Approach for Classification of Citrus Leaves using Fused Features," *International Journal of Food Properties ,* pp. 2071-2088, December, 2019.

[4] S. Zhu, L. Zhou, P. Gao, Y. Bao, Y. He and L. Feng, "Near-Infrared Hyperspectral Imaging Combined with Deep Learning to Identify Cotton Seed Varieties," *Molecules MDPI,* vol. 24, no. 3268, pp. 1-17, September 2019.

[5] K. Prashar, D. R. Talwar and D. C. Kant, "CNN based on Overlapping Pooling Method and Multi-layered Learning with SVM &KNN for American Cotton Leaf Disease Recognition," *Journal of Engineering Research,* vol. 3, no. 4, pp. 1511-1514, 2014.

[6] F. Shuli, A. H. Jarwar, X. Wang, L. Wang and Q. Ma, "Overview of the Cotton in Pakistan and its FutureProspects," *Pakistan Journal of Agricultural Research ,* vol. 31, no. 4, pp. 396-407, November 2018.

[7] D. I. Ahmad, M. S. Hanif, M. Asif, M. S. Malik, M. Z. Husnain, M. N. Gul, M. O. Farooq, M. A. Shah, M. N. Mazhar, M. S. Khatoon, M. A. Nawaz, M. T. Islam, H. S. M. Azeem, M. R. Akbar, M. U. Zahid, M. S. Saddiq, M. A. M. S. Khan, M. S. Ameer, M. F. Anwar and M. F. Shamim, "Pakistan Economic Survey 2019-20," Islamabad, 2020.

[8] W. Wei, Z. Mushtaq, A. Ikram, M. Faisal, Z. Wan-Li and M. I. Ahmad, "Estimating the Economic Viability of Cotton Growers in Punjab Province, Pakistan," *SAGE Open Original Reseach ,* pp. 1-12, June, 2020.

[9] S. Ashraf, A. H. Sangi, Z. Y. Hassan and M. Luqman, "Future of Cotton Sector in Paksitan: A 2025 Outlook," *Pakistan Journal of Agricultural Research ,* vol. 31, no. 2, pp. 145-150, June 2018.

[10] M. Ur-Rehman, Z. Rahmat, A. Mahmood, K. Abdullah and Y. Zafar, "Cotton Germplasm of Pakistan," *World Cotton Germplasm Resources INTECH,* pp. 137-166, June 2014.

[11] Z. Anjum, K. H. S. C. T. Azhar, U. S. F. A. M. T. H. Mehmood and M. A. , "Development of Cotton Leaf Curls Virus Tolerance Varieties Through Interspecific Hybridization," *African Journal of Agricultural Research ,* vol. 10, no. 13, pp. 1612-1618, March 2015.

[12] Y. M. A. A. M. S. A. W. Ali L., A. B. K. A. and A. Q. , "Evaluation of Release of New Cotton Variety "RH-668"," *International Joural of Biology, Pharmacy and Allied Science (IJBPAS),* vol. 9, no. 2, pp. 270-281, Febraurary 2020.

[13] M. F. Saeed, A. Jamal, I. Ahmad, S. Ali , G. M. Shah, S. K. Husnain, A. Farooq and J. Wang, "Storage Condition Deteriorate Cotton and Wheat Seed Quality: An Assessment of Farmers Awareness in Pakistan," *Agronomy MDPI,* vol. 10, no. 1246, pp. 1-19, August 2020.

[14] I. Ullah, M. Nawaz, N. Akram, M. Younas, M. Z. Iqbal and H. L. Shaheen, "PHYLOGENETIC ANALYSIS OF FIRST GENERATION COTTON LEAF CURL VIRUS (CLCuV) RESISTANT COTTON VARIETIES," *Biotechnol. & Biotechnol. eq. 2011 (A&EB),* vol. 25, no. 4, pp. 2610-2612, April 2011.

[15] A. Khan, J. Zheng, D. K. Yuen Tan, A. Khan, K. Akhtar, X. Kong, F. Munsif, A. Iqbal, M. Z. Afridi, A. U. S. Fahad and R. Zhou, "Changes in Leaf Structural and Functional Characteristics when Changing Planting Density at Di under a New Planting Model," *Agronomy MDPI,* vol. 9, no. 859, pp. 1-15, December 2019.

[16] K. P. Akhtar, M. H. W. Ishaque, M. R. Khan and A. I. Khan, "Effect of Cotton Leaf Mosaic Disease on Morphology, Yield and Fibre Characterstics of Upland Cotton in Pakistan," *The Plant Pathalogy Journal,* vol. 21, no. 2, pp. 137-141, Febraurary 2005.

[17] J. F. A. F. M. R. M. S. F. S. M.S Iqbal , T. A. B. and A. M. , "Cotton Leaf Curl Virus Disease a Principle Cause of Decline in Cotton Productivity in Pakistan (A Mini Review)," *Candian Journal of Plant Protection ,* vol. 2, no. 1, pp. 9-16, 2014.

[18] M. V. and P. R. , "CNN Approaches for Classification of Indian Leaf Species Using Smartphones," *Computer, Materials & Continua Techscience Journal ,* vol. 62, no. 3, pp. 1445-1472, 2020.

[19] H. Salih, S. He, H. Li, Z. Peng and X. Du, "Investigation of the EIL/EIN3 Transcription Factor Gene Family Members and Their Expression Levels in the Early Stage of Cotton Fiber Development," *Plants MDPI,* vol. 9, no. 128, pp. 1-16, Janury 2020.

[20] S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang and Q.-L. Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network," *Agricultural Journal,* July 2007.

[21] M. A. Khan, M. I. LaLI, M. Sharif, K. Javed, K. AURANGZEB, S. I. HAIDER, A. S. ALTAMRAH and T. AKRAM, "An Optimized Method for Segmentation and Classification of Apple Diseases based on Strong Correlation and Genetic Algorithm based Feature Selection," *igital Object Identifier IEEE Access ,* vol. 1, pp. 1-18, 2019.

[22] S. Qadri, M. D. Khan, F. Ahmad, S. F. Qadri, M. E. Babar, M. Shahid, M. Ul-Rehman, A. Razzaq, S. S. Muhammad, M. Fahad, S. Ahmad, M. T. Pervez, N. Naveed, N. Aslam, M. Jamil, E. A. Rehmani, N. Ahmad and N. A. Khan, "A Comparative Study of Land Cover Classification by Using Multispectral and Texture Data," *BioMed Research International Hindwai,* pp. 1-12, April 2016.

[23] F. Kurtulmuş, İ. Alibaş and I. Kavdir, "Classification of pepper seeds using machine vision based on neural network," *International Journal of Agricultural and Biological Engineering,* vol. 9, no. 1, pp. 51-61, January 2016.

[24] M. Clerc and S. Mallat, "The Texture Gradiant Equation for Recovring Shape from Texture," *IEEE TRANSACTION ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE ,* vol. 24, no. 4, pp. 536-549, April 2002.

[25] S. H. Lee, C. S. Chan, S. J. Mayo and P. Remagnino, "Howdeeplearning extractsandlearns leaf features for plant Classification," *Pattern Recognition Elsevier,* vol. 71, pp. 1-13, May 2017.

[26] N. Turner, A. Hearn, J. Begg and G. Constable, "COTTON (GOSSYPIUM HI_RSUTUM L.): PHYSIOLOGICAL AND MORPHOLOGICAL RESPONSES TO WATER DEFICITS AND THEIR RELATIONSHIP TO YIELD," *Field Crops Research Elsevier,* pp. 153-170, 1986.

[27] T. Aslam, S. Qadri, M. Shehzad, S. F. Qadri,

A. Razzaq, S. S. Muhammad, S. Ahmad, S. A. Nawaz and N. Ahmad, "Emotion Based Facial Expression Detection Using Machine Learning," *Life Science Journal ,* vol. 17, no. 8, pp. 35-43, August 2020.

[28] A. Tharwat, T. Gaber and A. E. Hassanien, *University of Salford Manchester ,* pp. 1-31, April 2016.

[29] A. Azizi and Y. Abbaspour-Gilandeh, "Identifying Irregular Potatoes by Developing an Intelligent Algorithm Based on Image Processing," *Journal of Agricultural Sciences,* vol. 22, pp. 32-41, December 2014.

[30] A. Rehman, L. Jingdong, A. A. Chandio, I. Hussain, S. A. Wagan and Q. U. A. Memon, "Economic perspectives of cotton crop in Pakistan: A time series analysis (1970–2015) (Part 1)," *Journal of the Saudi Society of Agricultural Sciences,* pp. 1-7, December 2016.

[31] T. Aslam and S. Qadri , "The Islamia University of Bahawalpur," December 2020. [Online]. Available: https://www.iub.edu.pk/.

[32] V. J. Pascual and Y.-M. Wang, "Impact of Water Management on Rice Varieties, Yield, and Water Productivity under the System of Rice Intensification in Southern Taiwan," *Water MDPI,* pp. 1-15, December, 2016.

[33] F. Khan, M. A. Khan, S. Abbas, A. Athar, S. Y. Siddiqui, A. H. Khan, M. A. Saeed and M. Hussain, "Cloud-Based Breast Cancer Prediction Empowered with Soft Computing Approaches," *Journal of Health Care Engineering Hindwai,* pp. 1-16, May, 2020.

[34] R. H. Ajaz and L. Hussain, "Seed Classification using Machine Learning Techniques," *Journal of Multidisciplinary Engineering Science and Technology (JMEST),* vol. 2, no. 5, pp. 1098-1102, May 2015.

[35] P. Chaudhary, A. K. Chaudhari, D. A. N. Cheeran and S. Godara, "Color Transform Based Approach for Disease Spot Detection on Plant Leaf," *International Journal of Computer Science and Telecommunications ,* vol. 3, no. 6, pp. 65-70, June 2012.

[36] N. Sengar, M. K. Dutta and B. Sarkar, "Computer vision based technique for identification of fish quality after pesticide exposure," *International Journal of Food Properties Taylor & Francis ,* pp. 2192-2206, August, 2017.

[37] G. Zen, L. Porzi, E. Sangineto, E. Ricci and N. Sebe, "Learning Personalized Models for Facial Expression Analysis and Gesture Recognition," *IEEE TRANSACTIONS ON MULTIMEDIA,* pp. 775-788, April, 2016.

[38] A. Humeau-Heurtier, "Texture Feature Extraction Methods: A Survey," *IEEE Digital Object Identifier ,* vol. 7, no. 2019, pp. 8975-8986, January 2019.

[39] G. M. M. E Elahi, "Gradient Based Novel Texture Feature Extraction Methods For Texture Classification," 2016.

[40] D. Kaur and Y. Kaur, "Various Image Segmentation Techniques: A Review," *International Journal of Computer Science and Mobile Computing,* vol. 3, no. 5, pp. 809-814, May 2014.

[41] G. Guo, H. Wang, D. Bell, Y. Bi and K. Greer, "KNN Model-Based Approach in Classification," in *Lecture Notes in Computer Science*, 2003.

[42] C. Li, S. Zhang, H. Zhang, L. Pang, K. Lam, C. Hui and S. Zhang, "Using the K-Nearest Neighbor Algorithm for the Classification of LymphNodeMetastasis inGastric Cancer," *Computational and Mathematical Methods in Medicine,* vol. 2012, pp. 1-12, September 2012.

[43] G. Chandrashekar and F. Sahin, "A Survey on Feature Selection Methods," *Computer and Electrical Engineering Elsevier,* vol. 40, pp. 16-28, December 2013.

[44] M. Shehzad , S. Qadri, T. Aslam, S. F. Qadri, A. Razzaq, S. S. Muhammad, S. Ahmad, S. A. Nawaz and N. Ahmad, "Machine Vision Based Identification of Eye Cataract Stages Using Texture Features," *Life Science Journal,* pp. 44-50, August, 2020.



Elahi et al LGURJCSIT 2021

ISSN: 2521-0122 (Online)
ISSN: 2519-7991(Print)

LGU Research Journal of
Computer Science and IT

Vol.5 Issue1, Jan-Mar 2021

Time Dependent Popularity Caching Scheme for NDN Based MANETs

Ehsan Elahi[1], Bilal Shoaib[2], Muhammad Akmal[3]
[1,2,3]Computer Science Department,Minhaj University Lahore
ehsan.cs@mul.edu.pk

ABSTRACT

Named data networking (NDN) approach has natural benefits within Mobile Ad hoc Network (MANET) but presents different issues as well. Space for cache, energy, and mobility of devices in a MANET is limited; therefore, we need for an enhanced judgement concerning which data to be store and where to be cache. A Time dependent Popularity Caching Scheme (TDPC) has suggested which selects nodes for caching the content on the forwarding path of packet and chooses the contents which have cached constructed on their time dependent popularity. At this interval, the cache distribution of the content and the storage capability of the devices are also measured. Results of the suggested TDPC approach are evaluated by using the simulator ndnSIM which is beached on Network Simulator 3 (NS-3). Simulation outcomes show that TDPC has good performance in expression of cache hit ratio, content retrieval interval, total cache copies and compared to the Dynamic Caching Strategy for CCN-based MANETs (CSCM). The goal of TDPC is to reduce cache redundancy, retrieval time of content and total number of cache copies.

KEYWORDS: Mobile ad hoc Network, Redundancy, Time Dependent Popularity, Content Store, Interest, Da ta

1. INTRODUCTION

Today, architecture of Internet is depending on TCP/IP model where communications are detained using devices or machines mac addresses. It is difficult to modernise today Internet architecture and additional improve its communication constraints. Moreover, the TCP/IP model has numerous weaknesses (security, location dependent and incompetent resource utilization). Now, the number of wireless devices has augmented sharply due to the rapid growth in technology. NDN approach will be principally useful in remote specially appointed systems administration environment [1].

MANET is infrastructure-less and uninterruptedly self-configuring network. The dispersed network, missing central administrative entity and comprises mobile devices which have dynamic and random movement in the network shown in the figure 1. In this network, every mobile device communicates within its range devices. Maximum routes used in MANET are multi-hop [2]. In short, MANET is an unpredictable network due to continuous dynamic nature of its mobile devices. With the development in the mobile devices, MANET is measured to be an important part of smart communication. Mobile devices exchange information with each other within their own specific range. In NDN based MANET, the joining mobiles are functioned in extremely active and challengeable atmosphere such as limited power of battery, frequency variations, sporadic connectivity, etc. NDN based MANET has been faced some serious issues which are data redundancy, inconsistency, mobility, crashes of packet, packet flooding and retransmissions of packet. These issues reduce the good results of the NDN based MANET [3].

Mobile devices can impart upheld what data they need, rather than computing a way to accomplish selection of hub. It might expressively change the implementation for the subsequent causes. In NDN the IP addresses are not assign to the nodes, it use the names of data and forward the Interest Packet base on the name of data which it is required and receive the data packet base on this name [4].



Figure 1. MANET Architecture

As a result of this data name primarily based style doesn't have routing circles, Interest Packets have been forwarded on several methods towards prospective data positions; just in case quite one way yields the demanded data, a node will measure that path provides the most effective performance and send upcoming Interest for identical data foundation in this direction solely [5]. This multi path methodology is especially useful in ad hoc networks as a result of the utilization

of multiple ways removes the vital dependency on pre computed single ways, therefore relaxed the rigorous necessities on the appropriateness of routing informs and routing state reliability between all the nodes.

We have proposed more effective networking caching scheme constructed on Time Dependent Popularity Caching (TDPC) to increase performance of cache of NDN based MANET in this paper. This approach creates filled procedure of the popularity of the data, and then selectively stores the data contents according to their popularity level by calculating the previous popularity with respect to time. This increases the usage of the store resources, and reductions rate of the data contents replacement. It evades the similar caching of content on the contiguous nodes, decreases the volume of redundant content in the network, raises the multiplicity of content and increases the hit ratio of cache. The main aim of TDPC caching scheme for the NDN based MANET to evades the similar content caching on the contiguous nodes, decreases the volume of redundant content in the network, rises the multiplicity of content, increases the cache hit ratio and to store the data locally.

The recap of this research paper is structured as follows. Segment II introduce the literature work. The complete TDPC scheme is explained in segment III and the simulation atmosphere and results are shown in segment IV. Lastly, segment V completes the paper.

2. RELATED WORK

Although several scholars have been suggested different NDN architecture for wireless networks, well-organized caching scheme of NDN based MANET are quiet challenging issues. Placement of content in cache focuses on when and where to place the copies. This section provides the summary of caching approaches in the NDN based MANET.

In literature [6], this has planned that the distributed caching schemes for ad hoc networks that either nodes have been created for caching the popular contents which permissions through or noted the forwarding route for data towards direct upcoming demands. The thought of dissimilar caching content on the neighbourhood had additionally subjugated in [7], wherever routers through alike interests and arrangements of mobility area unit classified along to enhance the hit ratio of cache. In literature [8], this has planned a caching decision scheme supported the responsiveness for data that have been stored on the neighbour nodes, targeting of distinguishing the caching data on the other nodes in network. The caching of content has been studied within the sporadically associated MANET likewise; a cooperative based caching mostly content distribution structure had planned. In literature, a cooperative caching technique was planned supported the broadcasting in environment of a radio network to boost the accommodating cache improvement. In [9], a content management design had planned for NDN based MANET, that finite the content and its physical position, and cached the contents that additional common intensively.

The construct of NDN has been introduced within the Vehicular ad hoc network (VANET) [10], discovered the social based cooperation policy between vehicles within the VANET and planned the cooperative caching supported the supplementary of corporation and messenger to boost the NDN caching performance, to enhance the consumer QoE on the multimedia system streaming facility. Considering the enough cache space, energy, vehicles, and transmission of data facility within the VANET had suggested in [11] to thrust the content to vehicles for pre-emptive caching. Within the lightweight of current improvements, caching on the corner positions in mobile cellular networks had been examined. In NDN based MANET, the devises activity as every user and router simultaneously, in this way the data demands approaching on the router envelop not exclusively the data demands delivered without anyone else anyway furthermore the boundless data demands from its neighbour routers. Hence, the prevailing caching analytical models can't be prolonged into NDN primarily based painter directly. During this case, it's essential to derive associate analytical model towards live the inspiration of caching in NDN based MANET tentatively.

To make the Named Data Network (NDN) model better support for Mobile ad hoc Networks, a caching scheme called time dependent popularity (TDPC) has been considered in this paper. Caching conclusion of TDPC drops under the complete deliberation of the data spreading in the network and the nodes cache ability. TDPC make guarantee that the data is close as the likely to the consumer nodes to reduce the data redundancy, increase hit ratio and decrease the content retrieval time.

3. TDPC CACHING SCEHEME

Decent usage of restricted cache in the Named Data Networking based MANET needs watchful calculation that which should be cached. It requires that every content should not cached equally a data cannot be cached on each device on routing path.

Table 1: Interest Packet Format

Field	Description
Name	Name of requested content
Lifetime	Expire time of the Interest which is in seconds
Nonce	Arbitrary value is used to escape repetition of Interests
Hop Count	Contains the distance of consumer to the other nodes

TDPC take two factors into account while making a caching decision on node which are distance of consumer from other nodes and calculated time-based popularity.

3.1 Distance of Consumer from Other Nodes

Separation of consumer from other devices is characterized by number of hop count checks. Complete number of jump tallies characterizes that the number of devices passed by Interest packet. We have putt one more field in the Interest packet which is hop count appeared in Table I. Customer advances Interest packet to the neighbour devices, neighbour device get packet of Interest and checks it's Content Store (CS), If mentioned information is found in content store at that point send back to the consumer device. If information isn't found in content store, neighbour device checks Pending Interest Table (PIT) section for this intrigue packet. Intrigue is essentially disposed of if PIT is as of now there. In any case section into PIT is included for the later reaction of information, increase hop count check an incentive by one and forward intrigue packet to different devices. Jump tally esteem is expanded by one at each device, when a device gets a Data packet as reaction, it will check PIT section for this information. If passage in PIT is discovered, hop count tally esteem is decremented and locale is controlled by utilizing this bounce check esteem. Information packet will be disposed of if section in PIT isn't found. We have included two extra fields in the Data packet appeared in Table II. In the event that bounce tally esteem is one, device is considered closest to shopper hub, thus information is considered for reserving by ascertaining its neighbourhood fame.

3.2 Calculating Time Based Popularity

When provider receives the Interest packet, it extracts the hop count value from packet of Interest, append to packet of Data and Data packet also contains the keywords field. When Data packets receive a relay node first it will check PIT entry in table, if not found then discard Data packet.

Table 2: Data Packet Format

Field	Description
Name	Name of requested content
Length	Requested content length
Content	Actual payload of content
Hop Count	Contains the distance of consumer to the other nodes
Keywords	Data contains keywords
Signature Type	That characterizes the category of signature such as SHA256 etc

Otherwise, extract keywords from data packet and calculate the time dependent popularity of keywords using following Equation:

$$P(i) = \left\{ \left(\frac{P(i-1)}{1+TF} \right) + IF \right\}$$

P(i) is the current time dependent popularity of keyword, which is calculated previous time-based popularity P(i-1) divided by time factor. Resultant estimation of condition will decide if the substance ought to be cached or not. On the off chance that time subordinate popularity P(i) bigger than certain limit (T), information will be stored in the CS in any case update Time subordinate Popularity Table (TPT) and forward Data packet to different devices in network by utilizing the Algorithm.

TF is the time factor which is in fact the time when the content is last accessed. IF is an impact factor of a content value of which determined by region of current with respect to requester. An area can either be local or global. If hop tally esteem is 1, it is taken a nearby locale else it is viewed as global area. The aftereffect of condition is taken a T to store information. If T is arranging as 5, information might be stored once consequence of this equation is reached to 5. Various estimations of effect are utilized for various areas. Effect factor estimation of nearby area is accepted higher as contrast with global, as we are more intrigued of information getting stored locally, which is more advantageous. We have taken IF esteem 0.25 for global district while 1 for local.

4. EXPERIMENT EVALUATION

In this section we have presented simulation results and demonstrate the performance of TDPC by evaluating it against several performance metrics such as:

- Cache Hit Ratio: Cache hit ratio is the ratio of number of requests, answered by source node to total number of requests generated within the network.
- Content Retrieval Time (CRT): CRT is defined as the time duration when an Interest packet is sent, and when the corresponding Data packet is received.
- Cache Copies: Total number of cache copies that are stored in the network.

Moreover, we have compared the performance of TDPC with A Dynamic Caching Strategy for CCN-based MANETs (CSCM) which implemented on the ndnSIM because there is no work on this research area [9].

Table 3: Simulation Parameters

Parameter	Value
Number of different content	100
Packet Size	1040 Bytes
Technology	IEEE802.11g
CS Size	50
TPT Size	100
Request Rate	20 Req/s
Replacement Policy	LRU
Simulation Time	1000s
Number of Nodes	60
Data Rate	6 Mbps
Popularity T	5
Simulation runs	10

A. Simulation Background

Results of the suggested TBPC approach is measured through using the simulation tools NS-3 and ndnSIM

[14] [15]. In MANET situation, we have taken 60 mobile nodes arbitrarily dispersed in 1000m x 1000m region. For mobility of a node we have been used arbitrary waypoint model. At the start, nodes are static and then start moving arbitrarily after some time to transformation from one point to another point position with the velocity of 10 to 25m/s. Among all mobile devices several are reserved as information supplier device although several are utilized as consumer devices rest are simply filling in as forwarder, halfway or transfer devices. In Interest packet broadcasting from consumer node to provider node we have used the best route strategy for forwarding. The requests for consents are generated by using zipf distribution [16].

All network nodes have consumed the IEEE802.11g technology constructed radio boundaries. Moreover, size of every Data packet is 1040 bytes and 6 Mbps is the data rate. The Simulation parameters are shown in TABLE III.

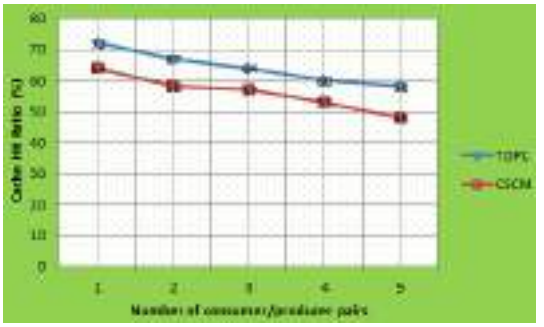

Figure 2. Hit Ratio of Cache

B. Experimentation Outcomes and Examination

Cache Hit Ratio: Figure.2 TDPC shows that hit ratio of cache is better than CSCM in all consumer producer pairs. Hit ratio decreases in both schemes with increase in the number of pairs as cache size to content sample space ratio decreases with increase in consumer producer pairs. Hit ratio in TDPC is better as compare to CSCM as in CSCM every new incoming data replaces existing data in cache, even some infrequently accessing and less popular content can replace some popular and frequently accessing contents. So, this unnecessary replacement causes lower hit ratio in CSCM.

While in TDPC only popular contents for which popularity threshold has been reached will replace existing entries in cache hence avoids unnecessary replacement and improve hit ratio.

Content Retrieval Time (CRT): TDPC CRT is less than the CSCM but increasing with the number of producer and consumer sets. Greater is the hit ratio; better would be content retrieval time as interest is fulfilled by shorter network path.

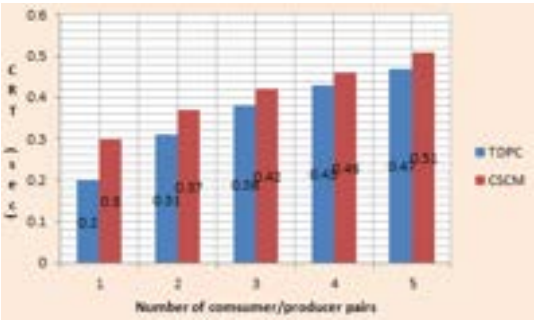

Figure 3. Content Retrieval Time

Since hit ratio is better in TDPC as compare to CSCM, this results as lower content retrieval time in TDPC shown in the Figure.3.

Cache Copies: Number of stored duplicates will increment with increment in producer/consumer sets in the TDPC and CSCM. This is on the grounds that more are producer/consumer combines particularly consumer, more devices will become in extent of correspondence henceforth more will be information duplicates stored. Anyway, TDPC consistently possesses lesser extent of store and this is again on the grounds that TDPC limits store by considering just nearby and worldwide areas as opposed to CSCM where information is stored on all devices on a correspondence way. This is clear from the diagram in Figure 4.

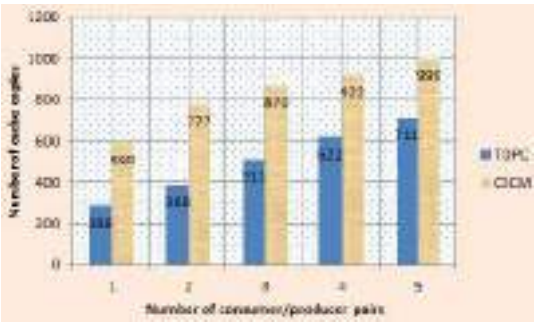

Figure 4. Number of Cache Copies

5. CONCLUSION

NDN based MANET has gain much attention in the recent past. In this paper we have purposed a novel time dependent popularity caching scheme (TDPC) for the NDN based MANET that enhance the MANET's network performance. TDPC caching the popular contents by determining their popularity based on the time. Time and areas are taken as significant factors of time depended popularity estimation. Simulation results shows that major enhancement in hit ratio of cache, retrieval interval of content and total of cache copies in TDPC. The next steps of this work maybe to enhance the network lifetime and handle mobility of the MANET, which are more important and challengeable.

REFERENCES

[1] Chen, C., Wang, C., Qiu, T., Atiquzzaman, M., & Wu, D. O. (2020). Caching in vehicular named data networking: Architecture, schemes and future directions. IEEE Communications Surveys & Tutorials, 22(4), 2378-2407.

[2] Gurung, S., & Chauhan, S. (2020). A survey of black-hole attack mitigation techniques in MANET: merits, drawbacks, and suitability. Wireless Networks, 26(3), 1981-2011.

[3] Babu, R. G., Karthika, P., & Manikandan, G. (2020). Polynomial equation-based localization and recognition intelligent vehicles axis using wireless sensor in MANET. Procedia Computer Science, 167, 1281-1290.

[4] Farheen, N. S., & Jain, A. (2020). Improved Routing in MANET with Optimized Multi path routing fine-tuned with Hybrid modeling. Journal of King Saud University-Computer and Information Sciences.

[5] Alam, T., & Benaida, M. (2019). The role of cloud-MANET framework in the internet of things (IoT). arXiv preprint arXiv:1902.09436.

[6] Q. Hu, M. Wu, and S. Guo, "Modeling content acquisition in two-dimensional content-centric MANETs," 2014 IEEE Wireless Communications and Networking Conference (WCNC), 2014.

[7] W. Dron, A. Leung, M. Uddin, S. Wang, T. Abdelzaher, R. Govindan, and J. Hancock, "Information-maximizing caching in Ad Hoc networks with named data networking," 2013 IEEE 2nd Network Science Workshop (NSW), 2013.

[8] L. Zhou, T. Zhang, X. Xu, Z. Zeng, and Y. Liu, "Generalized dominating set based cooperative caching for content centric ad hoc networks," 2015 IEEE/CIC International Conference on Communications in China (ICCC), 2015.

[9] S. Naz, R. N. B. Rais, P. A. Shah, S. Yasmin, A. Qayyum, S. Rho, and Y. Nam, "A dynamic caching strategy for CCN-based MANETs," Computer Networks, vol. 142, pp. 93–107, 2018.

[10] M. Yu, R. Li, Y. Liu, and Y. Li, "A caching strategy based on content popularity and router level for NDN," 2017 7th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC), 2017.

[11] W. Wang, Y. Sun, Y. Guo, D. Kaafar, J. Jin, J. Li, and Z. Li, "CRCache: Exploiting the correlation between content popularity and network topology information for ICN caching," 2014 IEEE International Conference on Communications (ICC), 2014.

[12] M. Amadeo, A. Molinaro, and G. Ruggeri, "E-CHANET: Routing, forwarding and transport in Information-Centric multihop wireless networks," Computer Communications, vol. 36, no. 7, pp. 792–803, 2013.

[13] A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM: NDN simulator for NS-3," NDN, Technical Report NDN-0005, October 2012.

[14] NS3, "Network simulator 3," http://www.nsnam.org.

[15] E. Chlebus and J. Brazier, "Nonstationary Poisson modeling of web browsing session arrivals," Information Processing Letters, vol. 102, no. 5, pp. 187–190, 2007.

[16] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," Proceedings of the 5th international conference on Emerging networking experiments and technologies - CoNEXT 09, 2009.

[17] M. A. Hail, M. Amadeo, A. Molinaro, and S. Fischer, "Caching in Named Data Networking for the wireless Internet of Things," 2015 International Conference on Recent Advances in Internet of Things (RIoT), 2015.



Bashir et al LGURJCSIT 2021

ISSN: 2521-0122 (Online)
ISSN: 2519-7991(Print)

LGU Research Journal of
Computer Science and IT

Vol.5 Issue1, Jan-Mar 2021

Trading Algorithm Model Based on Technical Indicators

Zulfiqar Ahmad Noor[1], Muhammad Khawar Bashir[2], Yasir Saleem[1], Zeeshan Ahmad[3], Muzammil Hussian[1], Nadia Tabassum[4]

[1] Department of Computer Science and Engineering, University of Engineering,
[2] Department of Statistics and Computer Science, University of Veterinary and Animal Sciences
[3] Department of IT, Pak Suzuki Motors
[4] Department of Computer Science, Virtual University of Pakistan.
mkbashir@uvas.edu.pk

ABSTRACT

Today the rapid proliferation of the internet provides an environment where efficient e-commerce solutions can be developed. The electronic market is gaining more attention in the global economy, it gives buyers and sellers more liberty to trade cost-effectively and allows access to an adequate amount of data for analysis. New trading agents have been developed for the best utilization of such data. These agents design strategies using financial analysis techniques such as technical indicators. Two very well-known technical indicators used to develop strategies are Convergence-Divergence (MACD) and Stochastic Oscillator (SO). This paper aims to devise a trading algorithm that combines MACD and SO in a single strategy and check the reliability of the combined signals it generates. JTAP simulation system has been used to test the proposed strategy. In this paper, we evaluated the performance of our proposed strategy when implemented on shares of Karachi Stock Exchange, Pakistan which proves improvement of strategy.

KEYWORDS: Trading strategies, technical indicators, moving average convergence-divergence, Stochastic Oscillator, genetic algorithm

1. INTRODUCTION

Initially, trading was executed in an open floor environment by the experts using the method of open outcry. The rapid growth of technology and developments in communication makes it possible to replace the open outcry system with electronic trading systems. It has transformed the nature of stock trading all over the world. Traders can purchase and sell stocks remotely using an internet-based order submission protocol. Such markets that are fully capable of electronic execution started working in the 1980s [1]. With the facilitation of remote trading in the stock market, different autonomous agents started to develop which can place buy or sell orders on their owner's behalf. Trading agents use the electronic trading markets more efficiently and act autonomously on behalf of their owners. These agents process all the related information around the clock and can decide on when and how much to buy or sell any stock. They are intelligent enough to take cognizant decisions autonomously and are equipped with sophisticated strategies for maximizing utility and profit. Such strategies are based on technical analysis which has been used in financial markets for decades. Many financial technical indicators have been used for market analysis and designing strategies. The most widely used trend indicator is MACD which shows the relationship between price and moving averages. It was introduced by Gerald Appel, in the 1970s that works by taking the difference of faster-moving average from slower moving average [2]. It indicates the traders to take benefits while they are holding the stocks for the long term or short term. Another popular technical indicator is the stochastic oscillator which was developed in the late 1950s by George C. Lane. It shows the location of the close relative to the high-low range over a set number of periods. Stochastic Oscillator neither follows the price nor follows volume but it follows the speed or the momentum of price.

In this paper, we evaluated the profitability of a trading strategy that combines MACD and SO. During experimentation, it's seen that technical indicators affect the trading strategy and can be used to improve the performance. Combining SO and MACD results in better signals because stochastic measures the relationship between a stock's closing price and its price range over a certain period, while MACD analyzes the difference between moving

averages diverging from and conversing with each other. JTAP simulation system has been used to test the proposed strategy. This system was developed by the Artificial Research Group [3].

In section 2., related work is reviewed. Section 3. presents our proposed architecture and its implementation. Section 4. presents evaluation criteria and our obtained results and in section 5., the conclusion and future work is discussed.

2. RELATED WORK

As the stock market is a complex dynamic system, so the stock prices are not only dependent on fundamentals but are highly motivated by short trading behavior. Trading agents can then design the strategies keeping such behaviors in consideration. Bloembergen, Daan, et al [4] analyze the dynamics of three different trading strategies which are based on trading behavior on the current market price only. Their result shows that it is hard to predict in advance which trading strategy will perform better. If the information is freely available, then fundamentals can work at best. After then Jose Alberto et. al. [5] defines a methodology that creates a process of modeling, developing, and evaluating trading systems. In this process, they create a multi agent system that integrates artificial intelligence techniques, allows the addition of new agents, guides the testing of different techniques, and enables the distribution of agents when computational resources are limited. However, these tools, due to the simplicity of their structure, perform poorly with adverse market conditions. Some sophisticated approaches are then designed to combine trading rules which helps in overcoming the problem [6]. These approaches were designed using genetic algorithms and genetic programs. Wooldridge [7] made a substantial contribution in establishing the generic framework of agent theory. M A H Dempster and C M Jones [8] also developed a trading system that consists of rules based on combinations of different indicators. A starting arrangement of such guidelines is selected by a genetic algorithm applied to several indicators.

Further, development focused on designing strategies for trading agents using financial analysis techniques. Technical indicators are then widely used in the financial market to predict future stock prices and market momentum which enhance trading profitability [6], [9], [10], [11]. Recent research focuses on a combination of technical indicators for precision. Genetic algorithms and genetic programs were used to combine different trading rules to overcome the poor behavior in adverse market conditions when implemented separately. This helped autonomous agents to achieve profitability in dynamic market conditions. Allen and Karjalainen [6] also used genetic programming to develop optimal trading rules which they found useful on a risk-adjusted basis. In further investigation, it is observed that genetic algorithm approaches are better when the transaction costs are zero. Besides, the indicator for technical analysis, which is based on fuzzy logic is also proposed [12]. In the article by Escobar et. al., it was concluded that agents using proposed indicators earn more profit, but the number of transactions is very high which is a disadvantage as each transaction has an associated cost.

Shah, Nabeel M [13] designed trading strategies using different financial indicators that were able to trade in multiple financial markets. Many other research studies have verified the importance of technical indicators focusing on trading strategy design and analyzing the statistical characteristics of these technical indicators [14], [15]. Technical analysis assumed that the future performance of stock markets relies on the historical performance of the market, so one can develop profitable trading rules using historical data. There has been continuous effort to investigate the performance of the technical trading analysis. Many researchers believed that technical analysis cannot generate abnormal profit [16], [17], [18]. In an article, Biondo et. al. [19] concluded that standard trading strategies and their algorithms can be successful occasionally, but it is better to use a random trading strategy for consistent profitable results. However, on the other hand, Traynor and Ferguson [20] concluded that technical analysis can produce a sizeable profit when non-public information is considered. Gunasekarage and Power [21], Chong and Ng [22] also concluded that a significant return is possible through technical trading rules. Lui and Chong [23] reported that traders who work through technical analysis significantly outperform those who do not use technical analysis. Brock, Lakonishok, and LeBaron [24] researched whether Moving Average and Trading Range Break are valid tools for trading. They tested these methods on 90 years of data from the Dow Jones Industrial Average. They found consistent results across the sub-periods in their dataset and found predictable patterns. By generalizing these two indicators they suggested that technical analysis can be a great help in predicting stock price changes.

In research at B.V. Patel Institute of BMC and IT [25], it was examined whether the technical indicators assumed any helpful part in the timing of investment at the stock market.

Test statistics are then developed to check any positive return yielded by buy and sell signals. Trend indicators used are Relative Strength Index (RSI), MACD, and Stochastic. The research concluded that MACD gives an accurate signal, better profit, and higher return compare to Stochastic Oscillator. Whereas Stochastic Oscillator gave more signals, but the accuracy of the signals was not so good as compared to MACD.

Different strategies for trading agents were proposed by Vytelingum et. al. [26] using a framework. In their framework they proposed that trading agents for designing strategies should manage their capabilities through three layers, information, Knowledge, and behavioral layer. After that Yaman et al [27] used an approach that was based on an evolutionary algorithm for modeling autonomous trading agents. In this approach collection of trading agents were generated which were using different strategies using Evolutionary Programming. Yaman concluded that the result is far better when different agents are performing concurrently.

Isao Yagi proposed [28] a trading model to suppress an increase in market volatility. They found that if the impact on market price formation depends on the minimum number of orders in a rebalancing trade.

A reinforcement learning-based algorithm was proposed by Wenhang [29] where they used two different models for theoretical analysis to extend their fundamental behavior for trading. Secondly, they analyzed cooperative and competitive behavior to design a reinforcement model. The realistic liquidation problem was solved using this methodology but using the same mechanism we can combine different models for our problem.

3. PROPOSED STRATEGY DESIGN AND ANALYSIS

The purpose of this study is to design a trading strategy for autonomous agents in financial markets that can work in dynamic market conditions. Moreover, we aim to check their performance in an autonomous simulation environment. Real data is used from the stock market in this environment so that simulation could be close to the real-time market. We used double-check strategy as a bidding strategy in which buy or sell indications have been generated by the combination of MACD and Stochastic Oscillator. For MACD, the difference of longer-term moving average (26-Day EMA) from the short-term moving average (12-day EMA). Additionally, the exponential moving average of MACD of nine days was calculated to generate a signal line. Mathematically it is written as

$$MACD = EMA\,(C, n_x) - EMA\,(C, n_y)$$

where x < y

Here C is the closing price of that specific stock for which moving averages are calculated. Position of the signal line and MACD help in determining the bullish or bearish trends. The trend is bullish if the signal is below MACD and bearish otherwise. If we combine the centerline with the MACD indicator, then the line with crossover will indicate the trading strategy that confirms the momentum of the stock. A buy signal is generated when MACD crosses the nine-day EMA of the MACD from below, while a sell signal is generated when MACD crosses the nine-day EMA of the MACD from above.

Trading signals generated by the crossover of MACD and signal line are interpreted in three different situations:

1. When MACD − Signal = 0, then it means that MACD has crossed the Signal. If we have positive MACD, it means MACD is above the signal line while if it's negative then it'll be below the signal line.
2. When EMA (C, n_x) − EMA (C, n_y) = 0, then it means that positive values increase which means upside momentum is increasing. Negative values increase as $EMA\ (C;\ n_x)$ diverges further below $EM\ A\ (C;\ n_y)$, this means downside momentum is increasing.
3. If we have equal relative price and MACD then the price graph will show higher highs or lower lows, but these will not be shown on the histogram. It also represents the strength of momentum.

Whereas, Stochastic Oscillator triggers the trading signal when the K value crosses over one-third of the moving average value known as D. It helps to identify overbought and oversold levels. The oscillator oscillates between 0 and 100. Usually, 80 taken as overbought, and 20 is an oversold indication. Mathematically it is written as:

$$K = 100\,\frac{C - L}{H - L}$$

$$D = 3 - EMA\,(K)$$

Where C is the closing price, L is the lowest price in the last n days; H is the highest price in the last n days of the stock under consideration.

Combining SO and MACD results in better signals be- cause stochastic measures the relationship between a stock's closing price and its price range over a certain period, while MACD analyzes the difference between moving averages diverging from and converging with each other. This combination evolves a strategy that confirms the trend with bullish MACD crossover and bullish stochastic crossover. It is caused whenever we have a condition that a slower moving average is left behind from a faster-moving average. As such things happen, it creates momentum in the market and prices tend to increase. Bullish of MACD depends on equilibrium and signal line. As the histogram value of MACD is above the equilibrium line or MACD line is above the signal line, MACD is bullish. The divergence of stochastic bullish happens as K value is above D that which also confirms the price turnaround.

The essential point of preference of this system is that as two different systems agreed, it will separate the signals. When we combine the two separate systems, it not only adds the advantages of both the systems

but also reduces the disadvantages that individual systems have. When MACD and Stochastic Oscillator systems are producing wrong signals about a stock at a particular time, this system will focus on the strongest signals and produces the right one. So, it will not only helpful in identifying the upward trending but also for downward trending. The focus on the strongest signals only is additionally the greatest shortcoming of this system. There will be very low chances of trading if we wait for the strongest signals only and we shall have to follow different markets to find regular signals for trading.

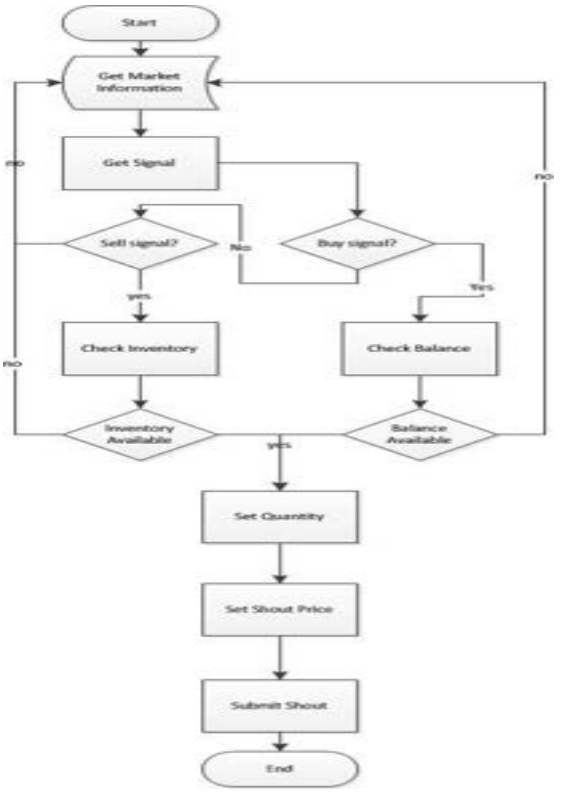

Figure 1: Flow Chart for Bidding Strategies

4. IMPLEMENTATION OF TRADING STRATEGY

For experimenting with autonomous trading agent strategy, we used simulation environment of Jacarkoo Trading Agent Platform (JTAP). The basic structure of JTAP is the client-server, and simulation duration is controlled using the same clock maintained by the server. We used the clock of 365 days to represent whole year trading. Furthermore, there are 10 rounds in each day while 500 milliseconds will make one round. JTAP used the double auction mechanism for trading. The single non-divisible asset is always there in every market that is the stocks of the market. For our experimentation, we used the stocks from the Karachi stock exchange with high and low range prices for one year of each stock. To study the outcome of different strategies in a non-sophisticated manner, we used a double-check strategy with a single stock price movement. Our strategies are based on financial indicators as we are using real-world techniques and data for simulation. We used different strategies for simulation using different settings of parameters and in this way, the optimized technique can be approached. In these simulations, four trading agents were used. To ensure a controlled environment and to perform a relative study of trading agent strategies, trading agents have the same amount of budget available at the beginning of the trade simulation. At a particular time, all trading agents will have the same price for a certain stock but on a single stock, agents can have different prices for each shout.

Input: n; currentPrice; buyQuantity; sellQuantity

Output: buy, sell, quantity, price

doubleMACD = EMA (12) – EMA(n).

signal = (MACDx0.2) +MACD$_y$x0.8

$$K = 100 * \frac{marketClosingPrice \ - \ low}{high \ - \ low}$$

$$K_1 = 100 * \frac{marketClosingPrice \ - \ low_1}{high_1 - \ low_1}$$

$$K_2 = 100 * \frac{marketClosingPrice \ - \ low_2}{high_2 - \ low_2}$$

$$D = \frac{K + \ K_1 + \ K_2}{3}$$

if (macd < signal & signal < 0 &

traderBalance > currentPrice × buyQuantity

& d > 80)

creatanewShout; Shout.setBid(false);

shout.Price = currentPrice + margin;

shoutQuantity = sellQuantity

else if (macd > signal & signal > 0 & Inventory > 0 & d < 20)

creatanewShout;

Shout.setBid(true);

shout.Price = currentPrice − margin;

shoutQuantity = buyQuantity;

This pseudo-code is an implementation of Double Check Strategy (integration of MACD and SO). In theory a 26-day EMA is subtracted from the 12-day EMA, the resulting calculation is a MACD. For performance comparison of different trading agents with different parameter settings we have replaced 26-day EMA by n.

Another indicator used is the Stochastic Oscillator (SO). SO is an indicator based on momentum and used to produce high-low ranges for n periods. This

type of indicator uses closing price for statistical analysis. After analysis, the indicator calculates a percentage value representing by k that is momentum.

To calculate d, we have used 3 periods moving average of k. To make the buying or selling decision, the value of d is very helpful. For example, if a trading agent wants to buy the stock, then needs to check i) is the MACD line below the signal line ii) is the signal in the negative range, and iii) is the value of d above 80. If all these are 'Yes' then the agent can bid for buy. For a selling event, the agent needs to check i) is MACD line above signal line ii) is signal in the positive range, and iii) is the value of d below 20. If the decision is to buy, then the further agent needs to check for available balance so that transaction could be made. An asset is deemed to be overbought once the Stochastic Oscillator approaches the 80 levels. Likewise, if the Stochastic Oscillator approaches 20, it is an indication that the asset may be getting oversold. Following are the results based on seven different shares generated using JTAP.

4.1 MCB



Figure 2: This is a graphical representation of the results generated for Double Check Strategy in comparison with MACD, Relative Strength Index (12 days), and Stochastic Oscillator during the simulation based on JTAP. This result is based on MCB's share price in the year 2004.

Table 1. Profit/loss of MCB shares

	Balance	**Share Value**	**Transactions**	**Profit/ Loss**
DC	82.3	2146.2	47	122.86 %
MACD	19.1	2146.2	192	116.54 %
RSI	978.5	0	680	-2.14%
SO	1024.8	0	311	2.83%

From table 1, we can interpret that while keeping the lowest transactions DC gives the highest profit which is122.86. Whereas MACD gives 116.54% profit and RSI gives a negative return, which is -2.14. So, gives 2.83% profit.

4.2 NBP

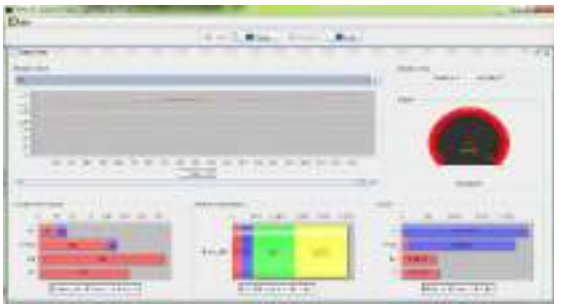

Figure 3: This is a graphical representation of the results generated for Double Check Strategy in comparison with MACD, Relative Strength Index (12 days), and Stochastic Oscillator during the simulation based on JTAP. This result is based on NBP share price in the year 2004

Table 2. Profit/loss of NBP shares

	Balance	**Share Value**	**Transactions**	**Profit/ Loss**
DC	11.5	2362.8	25	137.43 %
MACD	114.6	1999.2	101	111.39 %
RSI	634.6	0	184	-36%
SO	693.0	0	128	-30.6%

From table 2, we can interpret that DC gives 137.43% profit, MACD gives 113.39% profit whereas RSI and so gives negative returns, that is -36, -30.69% respectively. The transaction remains the lowest for DC.

4.3 NML



Figure 4: This is a graphical representation of the results generated for Double Check Strategy in comparison with MACD, Relative Strength Index (12 days), and Stochastic Oscillator during the simulation based on JTAP. This result is based on the NML share price in the year 2004.

From table 3, we can interpret that DC gives 52.24% profit, MACD gives 46.4% profit whereas RSI gives

negative returns, which is -35.19, and SO gives 6.85% profit. Transactions remain the lowest for DC.

Table 3. Profit/loss of NML shares

	Balance	Share Value	Transactions	Profit/ Loss
DC	1533.3	0	62	52.24 %
MACD	95.71	1368.2	135	46.4%
RSI	648.18	0	312	-35.1%
SO	1068.5	0	262	6.85%

4.4 OGDC



Figure 5: This is a graphical representation of the results generated for Double Check Strategy in comparison with MACD, Relative Strength Index (12 days), and Stochastic Oscillator during the simulation based on JTAP. This result is based on OGDC share price in the year 2004

Table 4. Profit/loss of OGDC shares

	Balance	Share Value	Transactions	Profit/ Loss
DC	1121.9	443.2	42	56.52%
MACD	1232.1	0	124	23.22%
RSI	528.5	0	226	-47.1%
SO	741.7	0	198	-55.6%

From table 4, we can interpret that DC gives 56.52% profit, MACD gives 23.22% profit whereas RSI and SO give negative returns, that is -47.15 and -55.68% respectively. Transactions are lowest for DC.

4.5 PSO

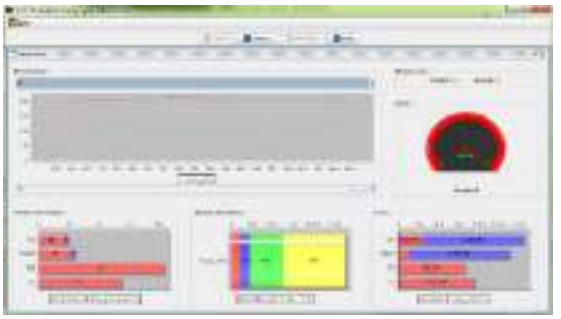

Figure 6 : This is a graphical representation of the results generated for Double Check Strategy in comparison with MACD, Relative Strength Index (12 days), and Stochastic Oscillator during the simulation based on JTAP. This result is based on PSO share price in the year 2004

Table 5. Profit/loss of PSO shares

	Balance	Share Value	Transactions	Profit/ Loss
DC	302.9	1260.8	21	56.38%
MACD	116.2	1260.8	27	37.71%
RSI	823.2	0	104	-16.78%
SO	7938.5	0	68	-6.15%

From table 5, we can interpret that DC gives 56.38% profit, MACD gives 37.71% profit whereas RSI and SO give negative returns, that is -16.78 and -6.15% respectively. Transactions remain the lowest for DC.

4.6 UBL



Figure 7: This is a graphical representation of the results generated for Double Check Strategy in comparison with MACD, Relative Strength Index (12 days), and Stochastic Oscillator during the simulation based on JTAP. This result is based on the UBL share price in the year 2004.

Table 6. Profit/loss of UBL shares

	Balance	Share Value	Transactions	Profit/ Loss
DC	1144.9	0	42	14.49%
MACD	1075.7	0	112	7.58%
RSI	490.3	0	302	-50.9%
SO	587.6	0	204	-41.2%

From table 6, we can interpret that DC gives 14.49% profit, MACD gives 7.58% profit whereas RSI and SO give negative returns, that is -50.97 and -41.24% respectively. Transactions remain the lowest for DC.

4.7 BYCO



Figure 8: This is a graphical representation of the results generated for Double Check Strategy in comparison with MACD, Relative Strength Index (12 days), and Stochastic Oscillator during the simulation based on JTAP. This result is based on BYCO's share price in the year 2004.

Table 7. Profit/loss of BYCO shares

	Balance	Share Value	Transactions	Profit/ Loss
DC	694.3	394.9	47	8.94%
MACD	14.05	956.2	192	-2.97%
RSI	974.2	0	680	-2.57%
SO	554.1	561.3	311	11.5 %

From table 7, we can interpret that DC gives 8.94% profit where MACD and RSI give negative returns, that is -2.97, -2.57% respectively SO gives 11.54% profit.

5. CONCLUSION

In this study, we have proposed an algorithm that is helpful for different trading strategies. These strategies are based on technical indicators that are used in all the financial markets for trading and analysis. At the start of each day, agents start the trading based on previous and running data to make the decisions of buying, sell or hold. Different signals are generated based on different technical indicators that are used by these agents to make the decision. We used data from the Karachi stock exchange and used different parameters to make different strategies and run them on the JTAP simulator. Based on the signals produced during the simulation we have concluded that if momentum-based strategies are combined with trend following strategies then we can have better profits. We compare the Double Check (DC) Strategy with MACD, RSI, and SO strategies applied on seven different shares. The high and low price of all shares taken for 365 days. From the result, it is evident that while keeping the lowest transactions DC gives higher profit most of the time. This shows that when we combine the MACD and SO then mostly it generates better profit. Furthermore, in volatile market conditions, we can have stable returns. In the future, these indicators can be produced using deep learning algorithms that can take raw data and direct us to make buy, sell or hold decisions.

REFERENCES

[1] D. Rime, “New electronic trading systems in foreign exchange markets,” New Economy Handbook, vol. 469504, 2003.

[2] G. Appel and M. Appel, "A quick tutorial in MACD: Basic concepts," Working Paper, Tech. Rep., 2008.

[3] “Strategic Trading Game,” http://www.jackaroomarket.org/JTAP.html, accessed: 2016.

[4] D. Bloembergen, D. Hennes, S. Parsons, and K. Tuyls, "Survival of the chartist: An evolutionary agent-based analysis of stock market trading," in Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems. International Foundation for Autonomous Agents and Multiagent Systems, , pp. 1699–1700, 2015.

[5] J. A. R. P. Sardinha, R. L. Milidiú, C. J. P. de Lucena, and P. M. Paranhos, A methodology for building trading agents in electronic markets. PUC, 2004.

[6] F. Allen and R. Karjalainen, "Using genetic algorithms to find technical trading rules," Journal of Financial Economics, vol. 51, no. 2, pp. 245–271, 1999.

[7] M. Wooldridge and N. R. Jennings, “Intelligent agents: Theory and practice,” The knowledge engineering review, vol. 10, no. 02, pp. 115–152, 1995.

[8] M. Dempster and C. Jones, “A real-time adaptive trading system using genetic programming,” Quantitative Finance, vol. 1, no. 4, pp. 397–413, 2001.

[9] L. A. Becker and M. Seshadri, “Gp-evolved technical trading rules can outperform buy and hold,” 2003.

[10] J.-Y. Potvin, P. Soriano, and M. Vallée, “Generating trading rules on the stock markets with genetic programming,” Computers & Operations Research, vol. 31, no. 7, pp. 1033–1047, 2004.

[11] F. Larsen, “Automatic stock market trading based on technical analysis,” 2007.

[12] A. Escobar, J. Moreno, and S. Múnera, “A technical analysis indicator based on fuzzy logic,” Electronic Notes in Theoretical Computer Science, vol. 292, pp. 27–37, 2013.

[13] N. M. Shah, “Trading Strategies for Autonomous Agents in Financial Markets,” Master’s thesis, School of Computing, Engineering and Mathematics University of Western Sydney, 2013.

[14] D. Zhang, K. Zhao, C.-M. Liang, G. B. Huq, and T.-H. Huang, "Strategic trading agents via market modeling," ACM SIGecom Exchanges, vol. 4, no. 3, pp. 46–55, 2004.

[15] P. R. Dasgupta and Y. Hashimoto, “Multi-attribute dynamic pricing for online markets using intelligent agents,” in Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1. IEEE Computer Society, pp. 277–284,2004.

[16] M. C. Jensen and G. A. Benington, “Random walks and technical theories: Some additional evidence,” The Journal of Finance, vol. 25, no. 2, pp. 469–482, 1970.

[17] S. N. Neftci, "Naive trading rules in financial markets and wiener- Kolmogorov prediction theory: a study of" technical analysis"," Journal of Business, pp. 549–571, 1991.

[18] M. Tanaka-Yamawaki and S. Tokuoka, “Adaptive use of technical indicators for the prediction of intra-day stock prices,” Physica A: Statistical Mechanics and its Applications, vol. 383, no. 1, pp. 125–133, 2007.

[19] A. E. Biondo, A. Pluchino, A. Rapisarda, and D. Helbing, “Are random trading strategies more successful than technical ones,” PloS one, vol. 8, no. 7, p. e68344, 2013.

[20] J. L. Treynor and R. Ferguson, “In defense of technical analysis,” The Journal of Finance, vol. 40, no. 3, pp. 757–773, 1985.

[21] A. Gunasekarage and D. M. Power, "The profitability of moving average trading rules in south Asian stock markets," Emerging Markets Review, vol. 2, no. 1, pp. 17–33, 2001.

[22] T. T.-L. Chong and W.-K. Ng, "Technical analysis and the London stock exchange: testing the MACD and rsi rules using the ft30," Applied Economics Letters, vol. 15, no. 14, pp. 1111–1114, 2008.

[23] K. man Lui and T. T. Chong, “Do technical analysts outperform novice traders: Experimental evidence,” Economics Bulletin, vol. 33, no. 4, pp. 3080–3087, 2013.

[24] W. Brock, J. Lakonishok, and B. LeBaron, “Simple technical trading rules and the stochastic properties of stock returns,” Journal of finance, pp. 1731–1764, 1992.

[25] S. N. Pinakin, "Comparision between MACD with ema and the stochastic oscillator," SIGecom Exchange, vol. 2, 2013.

[26] P. Vytelingum, R. K. Dash, M. He, and N. R. Jennings, “A framework for designing strategies for trading agents,” in Proc. IJCAI Workshop on Trading Agent Design and Analysis, Edinburgh, Scotland, pp. 7–13, 2005.

[27] A. Yaman, S. Lucci, and I. Gertner, "Evolutionary algorithm-based approach for modeling autonomously trading agents," Intelligent Information Management, vol4. 2014.

[28] Isao Yagi, Shunya Maruyama, Takanobu Mizuta, "Trading Strategies of a Leveraged ETF in a Continuous Double Auction Market Using an Agent-Based Simulation", Complexity, vol. 2020, Article ID 3497689, 7 pages, 2020.

[29] Bao, W., & Liu, X. "Multi-Agent Deep Reinforcement Learning for Liquidation Strategy Analysis.",ArXiv, abs/1906.11046. 2019.



Qadri et al LGURJCSIT 2021

ISSN: 2521-0122 (Online)
ISSN: 2519-7991(Print)

LGU Research Journal of
Computer Science and IT

Vol.5 Issue1, Jan-Mar 2021

Traffic Intensity Based Energy Efficiency Architecture for Data-Centers

Syed Shah Muhammad[1], Salman Qadri*[2], Hasnain Ahmed[1], Syed Furqan Qadri[3], Najia Saher[2], Sarfraz Ahmad[1] Tanveer Aslam[2] , Syed Ali Nawaz[2] and Muhammad Tariq Pervaiz[1]

[1]Department of Computer Science Virtual University of Pakistan, Lahore Punjab Pakistan.

[2]Department of IT, Islamia University of Bahawalpur, 63100 Bahawalpur Punjab, Pakistan.

[3]School of Computer Science and S.E, Shenzhen University, Shenzhen, China.

salman.qadri@iub.edu.pk,

ABSTRACT

The world is moving towards cost-effective and time-constrained solutions. The uses of applications and automated devices have been growing day by day. In computing, resources available in personal computers are limited due to less storage capacity and lower computation speeds. Using all applications on personal systems may not be cost-effective. Therefore, the trends of online storage and computing have become popular. On the other hand, there must be some serving end for these users. One of the major issues, due to the growth of data centers is the increase in power usage of a larger number of servers and network devices. These devices are power-hungry and consume energy even during idle hours even if there are no network traffic loads. The cost of energy used and dissipated is increased in this situation. In this paper, we have given a solution for efficient usage of energy efficiency in data center networks based on traffic loads. We have proposed a model to use traffic intensity to decide the number of machines inactive conditions so that we can save the energy consumption in data center networks. We have implemented this proposed model and simulated it to validate it.

KEYWORDS: Energy efficiency, power consumption, data centers, online storage, traffic intensity, energy aware routing.

1. INTRODUCTION

Data centers are technologies which are providing resources of information technology at the backend of a cloud. They are providing hardware, software, and infrastructure on a payment basis without any hurdle of deploying a new infrastructure according to requirements. Renowned companies like Google, Yahoo, Amazon had established their data center for services. These data centers are built on a large number of servers and network devices. In turn, all these servers and network devices use a large amount of energy. The power reduction and energy losses are major concerns of this age [1]. According to a survey energy usage for data centers is around 31 GW all over the earth [2]. Energy consumption is also continuously increasing. Only in U. S., this energy has been observed to increment up to 40% from 2007 to 2010 [3]. The energy consumed by data center networks in the US was 1.7 % to 2.2 % of total electricity usage in 2010 [4]. The major part of the data center is the data center network which plays an important role in traffic flow among resources. Datacenter networks (DCN) are supposed to work even for the maximum traffic loads. This traffic peak may not be the same all the time [5]. According to [6] a data center's average usage is 20 to 30 %. Also, resources are using energy even if there is no traffic flow [7]. Many solutions for energy efficiency in DCNs are available [8–12]. Due to the large amount of energy consumption in data center networks has become abundant [13–15]. Keeping in view all these facts there is a need to provide energy-efficient solutions for data center networks. Many solutions for data center networks had been proposed by researchers. Commonly, data centers use a 1:1 subscription ratio for communication over links but practically there are all to all scenarios as per network traffic loads. In most hours, the network devices and servers are under-utilized due to variable usage at day and night time. In most situations, network traffic flows in data centers are high during office hours and lowered at night times. Due to this the devices and links in data center networks are underutilized but consume the same energy. In the US the energy consumption was 3 billion kWh in one year [15].

This is due to fact that many of the chips, fans, line cards inside the network devices consume a large amount of energy even when not participating in traffic flows. Also, the network devices use around 30% of energy in data centers [16]. The energy-aware solutions can make efficient use of devices [17]. As the data centers are the backbone of all services provided over the cloud [18], therefore the data center networks are the major energy-consuming infrastructure of the world [19]. Some of the researchers have observed the fact that the energy consumption of data centers will be raised to 8% of total energy usage of the world by 2020, which was around 1.5% in 2011 [20,21]. Some of the authors have proposed virtualization of network resources and the consolidation in physical machines can reduce energy consumption [22]. Many of the solutions provided and the governing protocols do not take care of proportionality measures between network load and several user devices in communication. There might be a lower subset of devices that is enough for off-hours when traffic load is smaller. A solution based on bandwidth calculation for making the energy proportional network in data centers has been proposed by the researcher [23]. PTNet topology has been introduced by a researcher which provides scalability along with minimum latency, high performance, and lowest energy consumption [24]. There are some solutions but they involve long routes and some of them have high computing complexities. Also, performance issues and delivery delays have been observed. Based on these observations there is still a need for energy-efficient routing in data center networks

In this paper, energy conservation is based on routing decisions. This will be energy-aware routing for energy efficiency in data center networks. The basic idea is to use a fewer number of devices involved in routing with minimum effect on network performance. These devices can be shut down for this purpose.

First, we have formalized the problem for routing and then to prove that it is NP-hard. After this, we have proposed a routing algorithm to achieve the energy-efficient solution. This routing algorithm will focus on achieving energy-efficient routing with reduced computation complexity along with the minimum effect of the performance.

The algorithm is devised to first calculate the throughput based on the network traffic patterns which serves as a basic parameter. Also, the core nodes are determined which have a maximum load at a given time t. these nodes are kept alive for routing. These core nodes are selected based on their usage which can never be put to shut down. After that, the algorithm excludes the devices from routing paths until there is no effect on performance. In the end, these excluded switches are turned off. The major contributions are:

1. Problem formulation for energy efficiency in data center networks and its solution.
2. Finding the core nodes involved in the routing and the nodes to be excluded from routing.
3. Focusing on the energy-efficient model with lowest compromise on performance and computing complexity.
4. Implementing and evaluating the proposed solution

Simulations have been conducted to evaluate the effectiveness of this proposed energy-aware routing algorithm. Finally, the results show that this is an efficient solution for energy conservation in data center networks under different network traffic loads.

Procedures to utilize the traffic intensity to control active nodes in a data center in such a way to allow the only required number of nodes running and switch off the additional nodes. The idea is to establish a relationship between the requests and the required number of nodes to handle the incoming requests efficiently and optimize the data center performance and energy utilization At the end of this paper, we have shown the simulation results in virtual data-center using the NS3 to generate the traffic and Docker containers to act data center nodes. Before establishing the mathematical relationship, we discuss some important terminology that shall be used in this paper. The request is a formal request from a user or to an end client that needs to access resources from a server node in the virtual data-center. In real-life scenarios, this request could be transactions, video or voice session, a file transfer, etc. Requests could have certain characteristics like priority, time, and order. In our simulation, we shall consider that all requests have the same priority, completion time, and order of fulfillment as per the request arrival order. The response is the server's reply after using or deploying some resources of it. This could be simple processing. Once the request is fulfilled server generates a message the request is completed.

Request Completion Time (RCT) is the time elapsed form request generated by a user and the user received a response. RCT is the sum of the processing time plus network delay or queuing delay. The capacity of server nodes is currently inactive state corresponding to active traffic intensity. Active traffic intensity is the value reported by the network. The nodes are the servers in the data center that can fulfill the user request. Server nodes are connected in groups referred to here as a cluster. Several nodes are connected to a switch or bridge. Once all the nodes in a cluster to switched on, the relevant switch or bridge can be also turned off to save energy.

This paper refers to the request-response model having the request completion time. The general

definition of traffic intensity as defined by the ERLANG is applicable here; however, the calls and related specifications are translated to the request-response model. Originally the errant determines the optimal trunks required given PABX, but this paper will determine the number of server nodes required to fulfill the user requests with optimal performance and data center efficiency.

The rest of the paper is arranged as follows. In section 2 the latest work done in the relevant field has been elaborated. In section 3, we have formulated the problem and presented the algorithm for energy-aware routing in data center networks. In section 4, the results are discussed to validate the authenticity. In section 6, the work is concluded. Section 7 gives future directions for this research.

2. RELATED WORK

The major issue faced by data centers from a smaller scale like public and private sector organizations to very large scale data centers like Facebook, Amazon, Google, Microsoft, is the energy consumption. Due to this reason, finding the solution to optimize energy usage has always been a research-focused area. This is not only to reduce energy consumption but also to protect the environment. A large amount of work has been done to minimize energy consumption. The researchers have proposed different approaches to minimize energy consumption in different ways. In the following, we have presented some state of artworks regarding energy efficiency in data centers.

2.1. Energy-aware architectures

In category, the researchers focus to design or use such topologies or architectures for the data centers which help in saving the energy utilized. For this different energy efficient architectures have been proposed. Nano data centers [21], have been proposed in this regard. In this topology, home gateways are used to save the services and content rather than in data centers. A P2P infrastructure can be used to access this data. Some other topologies have also been proposed for the energy efficiency of wireless sensor networks [22-23]. A large amount of energy can also be saved by using transceivers with a minor energy consumption compared to switch interfaces [24].

NovaCube [19] and flattened butterfly topology [20]. A reasonable amount of energy can be conserved by making the use of such topologies, by optimizing the energy usage by nodes and network devices. Some of these above-mentioned approaches are feasible for wireless networks and others for a small network.

2.2. Power Efficient Devices

In this approach, the major motivation is to make the nodes and network devices use the energy as per network traffic. For this Dynamic Voltage (DVFS) [18] can be used. DVFS technique is used to make the power usage of devices such as CPU as traffic loads. They used the idea of voltage to frequency proportionality. This had been used to energy used by a chip proportional to the network traffic load. Researchers have used a variation on/vary off approach to implementing this idea. The workload can be distributed among a smaller subset of nodes rather than spanning it to all nodes. They proved that this could be an efficient way to conserve energy. This approach is limited to minimize the energy usage by the CPU while all other devices are using the same huge amount of energy.

2.3. Approaches for Renewable Energy

Some of the researchers showed that solar energy, wind, cool regions, water, and wind provide support to reduce a large amount of energy consumption in data centers. Net-zero networks have also been proposed Arlitt et al. in 2012 [17]. In this approach, an amount of energy can be regenerated as the amount of energy consumed by the data centers in one day. Nguyen et al. (2013) [16] focused to provide renewable energy for the data centers. They used solar and wind energy systems, along with the sensors for temperature and moisture as a testbed to implement the energy-efficient solution for data centers. They implemented their solution in different US projects and achieved fruitful results for energy efficiency. These renewable energy solutions have some limitations. First, the initial infrastructure costs and the delivery cost of renewable energy from these sources is high. Secondly, these sources depend on the weather conditions, as the solar works in sunlight. The locations of these data centers also play a role limit the efficiency of such solutions.

2.4. Virtualization

A single physical machine can be used as multiple machines by using a technology called virtualization. In this technique, more than one instance known as Virtual Machines (VM) is created using a single machine [11,25]. These VMs interact with the operating system as a real machine. Generally in this approach, we can use one machine which is acting as more than one machine thus reducing the energy used by other machines. One of the popular tools used for virtualization is VMware [26].

Some researchers [27] have user requirements and network devices usage to suggest algorithms for the placement of virtual machines. Also a comprehensive study had been conducted [28] to understand effect of two different energy consumption techniques (RSS and MNP) for different types of data center architectures.

Also GreenCloud [29] concept of virtualization in data centers. For this, the authors first investigated the network traffic and proposed the routing algorithm for the placement of virtual machines. They got results for energy efficiency along with

good network performance. They carried out this work in two parts: first by finding the network traffic loads and the placement of VMS as per requirements. The extra machines which were mapped to VMs can be turned off to optimize energy usage.

2.5. Routing level power aware approaches

Different researchers have worked on energy-aware routing to provide the solutions. In this category, energy efficiency is applied to communication devices in such a way that energy can be conserved in a considerable amount. The basic idea behind this rationale is that the devices not participating in communication consume the same amount of energy. This concept can be used to devise energy-aware solutions by merging traffic flows in such a way that only a specific group of communication devices is used. The rest of the devices are put to power off to save energy. Energy-aware routing algorithm [13], Energy optimization on-device interface approach [30], Elastic tree [9], Artificial intelligence abstraction model for Energy efficiency in DCN [2] and traffic merge [31] approaches are some examples in this domain.

2.6. Elastic Tree

One of the energy aware solutions is elastic tree [9]. In this approach, the authors have focused to find a minimum number of communication devices that can be used with a lesser impact on network performance. This approach is based on three modules: optimizer, routing module, and power controlling module. These three modules are implemented to get the power-aware routing in data center networks. The optimizer finds the lowest number of communication devices which is enough to provide the traffic flows required in the data center. In the routing part, the required routes are calculated. In the third module, the states of devices are adjusted. They proposed to use three optimizers for the calculation of the minimum number of devices which are enough to fulfill the routing requirements. This optimized solution is achieved as per traffic conditions in the network. For each flow, they calculated the routes which could be used and chose the left one. Assigning traffic flows is NP-hard and takes long time duration of hours. The scalability is also a major issue for the optimizer which can handle only up to 1000 devices. Greedy bin packing is used to enhance the scalability but still, the solution is not optimized. Also, this approach was unable to give the required quality of service as they were unable to calculate the exact routes

The communications devices specifically, switch consume large amount energy in data center networks. In traffic merging [31] technique, focused to reduce the energy consumption in data center networks. They merged the traffic from different links reaching to links by utilizing the lowest subset of ports. This merging technique was used in such a way that many of the ports are using the lowest energy consumption manner. The traffic is merged such that the number of traffic flows is reduced to use the number of minimum ports for communication. This method saved maximum energy in data centers, but this approach has a limitation that it can only save the energy consumed by the network. Another risk involved in this approach, some of the merged links or ports may lose communication.

There is another way to address the energy efficiency is energy-aware routing proposed by researchers [13]. The basic idea is to use the lowest number of switches for routing the traffic in data center networks keeping the network performance in consideration. They calculated the route and performance of putting the machines off. Then in the second phase, they put the unwanted machines to off condition keeping the performance up to a certain level. However, this approach has some issues. The computing time for the elimination of machines is incurred. Also, this saves only energy used by switches and not the energy used by the links and servers. In some cases, the performance is degraded. This could be fatal for the data center where the real-time applications are served.

A fair bandwidth allocation technique using artificial intelligence abstraction is proposed by researchers [2]. They used the blocking island (BI) standard for bandwidth allocation to energy efficiency on data center networks. They implemented this idea in two phases. First, they applied BI to efficiently use the network devices, reduced the communication flows, and then found the path which consumes minimum complexity. In the second part, they presented an energy-efficient routing algorithm. This algorithm selects routes as per traffic intensity by shutting down the unwanted routes. This approach has lowered the calculation complexity for traffic loads but still, it has some limitations. Calculating all routes and selecting the optimized path is still so much complex.

Energy saving by the optimization of interface is proposed by researchers [30]. It changes the state of the port of a switch as per the traffic queue length. They make use of traffic loads to minimize the energy consumption on switch ports. Initially, they assumed that network devices are not active. These switches are activated upon the traffic arrival. The traffic is considered as packets. The queue is observed after the arrival of each packet. The port is downgraded to a lower level when the buffer goes on a certain lower level *Town.* The port can be made up if the buffer level grows to a certain level *Tup*. In this practice, a port can have many different states. It may be in a sleep state, or some other states like 100 Mbps, 1or 100 Gbps as per buffer situations. This approach saves a reasonable amount of energy but some of the limitations of this approach are also there. The computation time is high. In some cases, network reliability is compromised and performance

is degraded. To address such issues we have proposed a new approach, which saves maximum energy with minimum impact on network performance.

3. PROPOSED MODEL

For the implementation of a traffic intensity based energy-efficient solution, we require certain additional capabilities to be included in the data center environment. These include a power monitoring module, routing algorithm, server management API, Network Management API. In addition to this network, nodes must be capable of propagating the traffic intensity information as described in the paper. The working principle of this model is elaborated in the sections below.

3.1. Monitor

The monitor is a daemon that keeps monitoring the load on nodes and determines the criteria whether new nodes need to be switched on or the existing node needs to switch off taking account of some preferences that cover the time required for start time. Consider T(s) is the node startup time and the average request serving duration is D.

Average Request Completion Time = $\boldsymbol{T_{comp}}$

Total number of requests being server = $\sum_1^N \boldsymbol{R_n}$

Estimated remaining time of completion = $\sum_1^N \sum_1^{R_N} \boldsymbol{T_r}$

3.2. New Request

Figure 1 depicts the operation of the algorithm. CC represents the current capacity in traffic intensity units and CTI refers to the current Traffic intensity value in a given interval. Request refers to a new request from a client or user that needs to reserve or use some resources from the datacenter. We consider only the computation

resources only. VDC is the virtual node controller. N-Free represents the available nodes that are not being used and have been shut down in some previous interval or were never started previously. The monitor evaluates the current capacity of the datacenter in TI units.

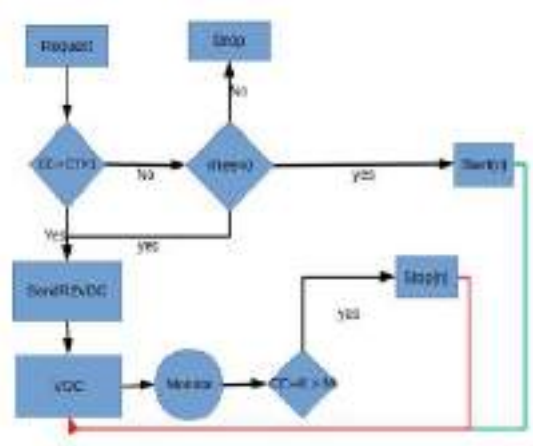

Figure 1: Operation of the algorithm of Monitor Algorithm

The datacenter nodes are started and stopped as per the CTI and CC values. K is a parameter that addresses the issues of the node start and stops time and M represents the monitored value. The start is a function that triggers the virtual node to boot up the nth node and a similar stop is a function that triggers the shutdown of the nth node.

3.3. Start Node

Nodes are started if their demand of resources and current capacity reaches the startup and shutdown threshold. All nodes in the datacenter are assumed to have the same capacity and capabilities. So starting a node n is just a matter of selection of a node from the shutdown pool. During the node boot up the incoming requests are forwarded to already running nodes having the least load. Once the new node is started new requests are forwarded to the new boot-up node. Existing requests are never transferred between the nodes.

3.4. Stop Node

A Node is required to be shut down if there is no request being served by it and all other running nodes are running with a load below the threshold. If there is no request being served by a node and other nodes are operating near to threshold or above the threshold, the stop operation is rejected and the load balancer forwards the new requests to this node. The Stop operation makes sure that there are enough resources available in the data center if the node stopped. This condition is limited only to the current interval where TI values are considered. If the interval expires algorithm recalculates the over status and decides whether the node needs to be shut downed or there will be future requests that shall need this node.

If multiple nodes are running and the current capacity is sufficiently lower than the current interval TI, the load balancer tries to move the incoming

request to specific nodes that fulfill the criteria for future operation. This means the nodes that are decided to be kept on shall receive all future requests in a given interval. This is the result of freeing up the non-interesting nodes. The decision of shutting down a node is critical as compared to starting up a node. In a later case, the random node is selected from the off-pool whereas the selection of shutdown node is done with coordination of load balancer and monitor. So if the algorithm has decided to shut down a node, there is not guaranteed that the node will be shut down due to currently running requests. The shutdown operation may be delayed depending on the load balancer. Also, it is possible that in a given interval where a shutdown has been decided by the algorithm may never actually happen in this interval.

3.5. Stop Cluster

In a datacenter, nodes are physically clustered and each cluster may have a size of C where C is an integer determined by the data center capacity planning and dimensioning teams. Each cluster has associated networking and other power equipment. Shutting down a cluster is quite interesting as it saves greater power consumption as compared to shutting down a node. When the cluster is shut downed related networking components and power components are also shut downed resulting in greater saving of power. The criteria to shut down a cluster are simple. If all the cluster nodes are down and CC << CTI in their interval the cluster will be shut downed.

```
Active_Servers[]:

Server Request Capacity = Sc
Total ServerFarm Request Capacity = Fc

Request Intensity in Current BP = R

Excess Capacity = Ec
Excess Servers = Es

Selected Servers = SS
Running Servers = RS

Foreach(bp in BP) {

        If (Fc > [R + K]) {
                Ec = Fc – R
                Es = Ec / Sc
                Select (n = Es) Servers = SS
                Transfer Load (SS,RS)
                Schedule All Future Request in BP to RS();
                }
    Else if (Fc < [R+K]) {

            Required Capacity = R – Fc +K
            Required Servers = Round (Required Capacity / Sc)
            Select OFF-SERVER[] = OFFSERVERS

            Foreach (OS IN OFFSERVERS) {
            switchOn();
            }
                Schedule All Future Request in BP to RS();
            }
}
```

Figure 1: Server Addition/Removal Algorithm

```
Foreach (as in AS) {
If (CurrentLoad(as) < 90%) { CS->addserver(as)
}

While (size(Request) > 0) {
    Foreach (r in requests) {
            Foreach (cs in CS) {
                    r->redirect(cs);
            }
    }
}
```

Figure 2: Request Allocation

This operation also takes into consideration the startup and shutdown times of the nodes as well as the networking and power equipment.

3.6. Start Cluster

Cluster start is triggered when the off-pool is exhausted and there are off-clusters. This follows the values of the monitor and CTI. If CTI >> CC and off pool = 0 then a new cluster is triggered to start. The cluster start time is also considered in this operation. All the new requests are managed by the load balancer.

3.7. Request Distribution Algorithm

In previous sections, we have discussed the basic concepts and terminology. Now this section covers the working of an algorithm to distribute the incoming requests to server nodes and select an optimal number of servers to serve the traffic load. The algorithm compares the total server farm

capacity against the capacity required to serve the TI in each BP.

It calculates the excess servers and based on the excess capacity and initiate additional server switch on or switch off. All these decisions include the K variable that accommodates the servers boot up period and request redirection time. Only new requests are scheduled for the new allocation. The scheduling algorithm works in such a way that it selects the servers with the least running load with a threshold of 90% load. If the load approaches 90% new requests for a switch on are made.

4. RESULTS

In this section, we discuss the simulation details of the comparison environment.

The simulation consists of three parts as below.

- Virtual Datacenter (VDC)
- Controller & Power Analyzer (C&PA)
- Simulated Network Environment (SNE)

VDC provides sixteen server nodes arranged in 4 clusters each containing 4 server nodes. Each node runs a specialized web application that responds to the user requests on certain TCP ports after processing some information as requested by the user. VDC is connected to a simulated environment that uses an NS3 simulator. The network topology is planned as per the capacity planning using the Traffic Intensity Principles discussed in earlier sections of this paper.

Traffic is generated from an NS3 simulated network topology following the selected traffic intensity profile. In the simulated network, users are distributed in 16 Point of Presence Points (POPs) where each POP has different traffic intensity profiles. Different types of access POPS (Max of 4) are aggregated at aggregation POP. Mixing of the different types of POPs in the aggregation is very important as it allows better utilization of the network resources

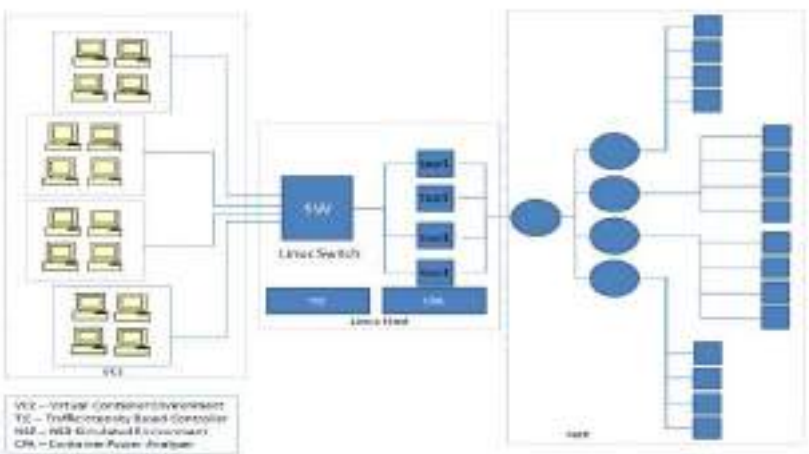

Figure 3: Simulation Environment

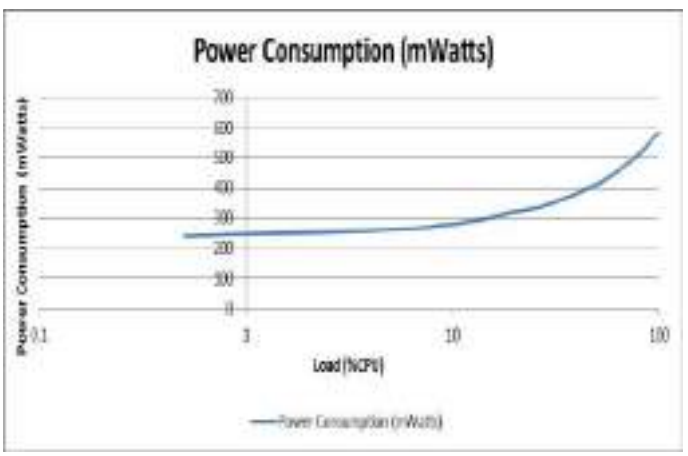

Figure 4: Power Consumption of Virtual

It also allows the performance enhancements if the traffic intensity based Packet Scheduling or Queue Management Algorithms like TIPS [R1]. Anyhow here we assume that all the network topology follows a commonly known architecture and traffic sources or users are placed randomly.

4.1 Simulation

The above algorithm was simulated in a virtual datacenter build with containers. For the request generation, we used NS3 to generate requests with the required characteristics. Each server node is run in a Linux container having a web server, MySQL database, and some other minor utilities for housekeeping and analysis purposes. The controller node is also implemented in a Linux controller. Controller and NS3 modules are fed with the traffic intensity information via a shared resource.

Total Access Switches: 16 (corresponding to 16 access POPs)

Nodes

Free Running Power Consumption = 250mW

10%load Per Percentage Wattage = 50/10 = 5W

10 to 70% = 400 – 250 = 150/70

Server load is monitored in each sampling period. Server load is analyzed for power consumption per unit of server load. The first approach uses a predefined traffic load with round-robin server allocation Second approach uses the same predefined load with a traffic intensity based server allocation algorithm. For both approaches, the server load is measured in each sampling period. Corresponding power consumption was calculated for each sampling period. Figure 6 and 7 shows the power consumption of datacenter nodes. These results are captured with the random load without considering the CTI values and TISC being not operational. Different colors of the graph represent different datacenter nodes and the Y-axis represents the power consumption and the X-axis shows the time line.

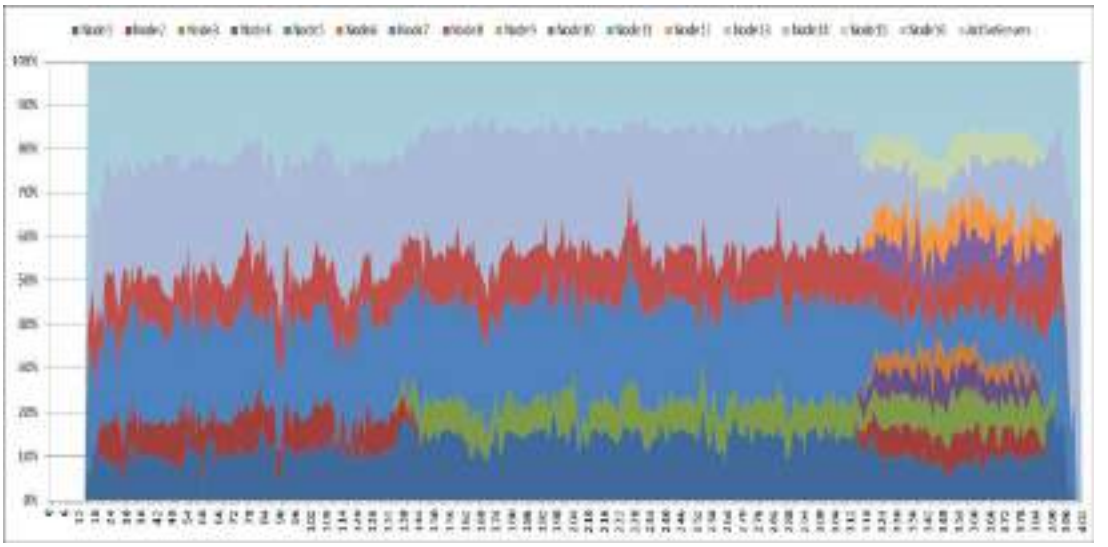

Figure 5: Power consumption of datacenter nodes

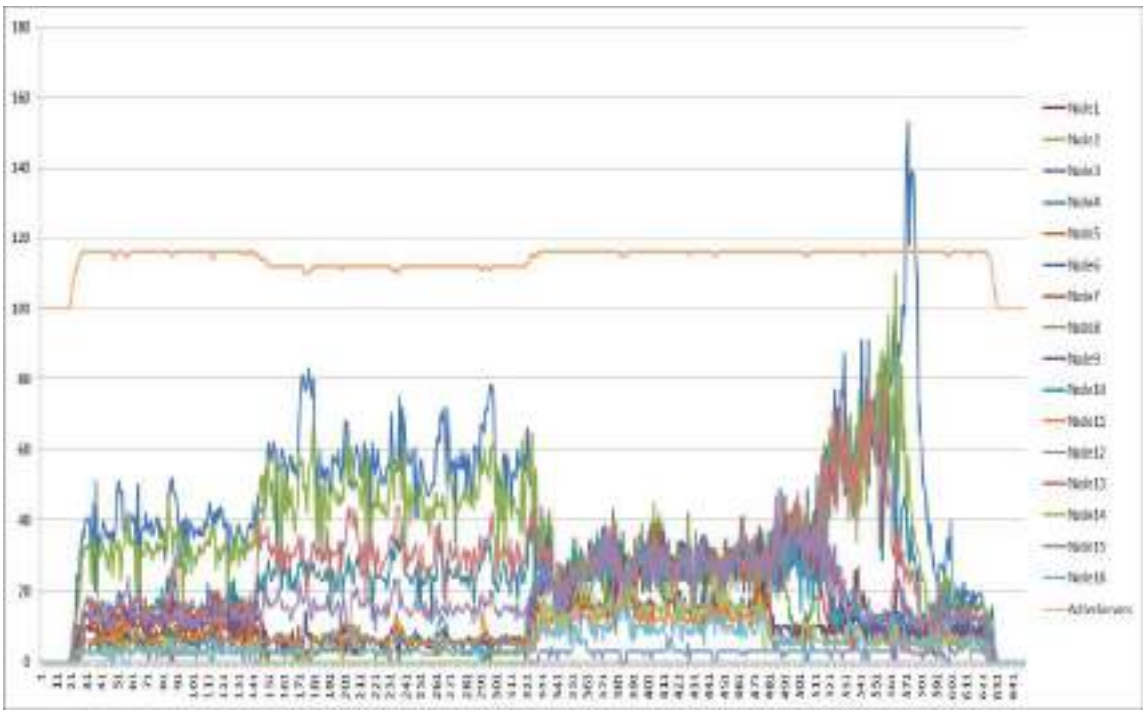

Figure 6: Power consumption of datacenter nodes with random load

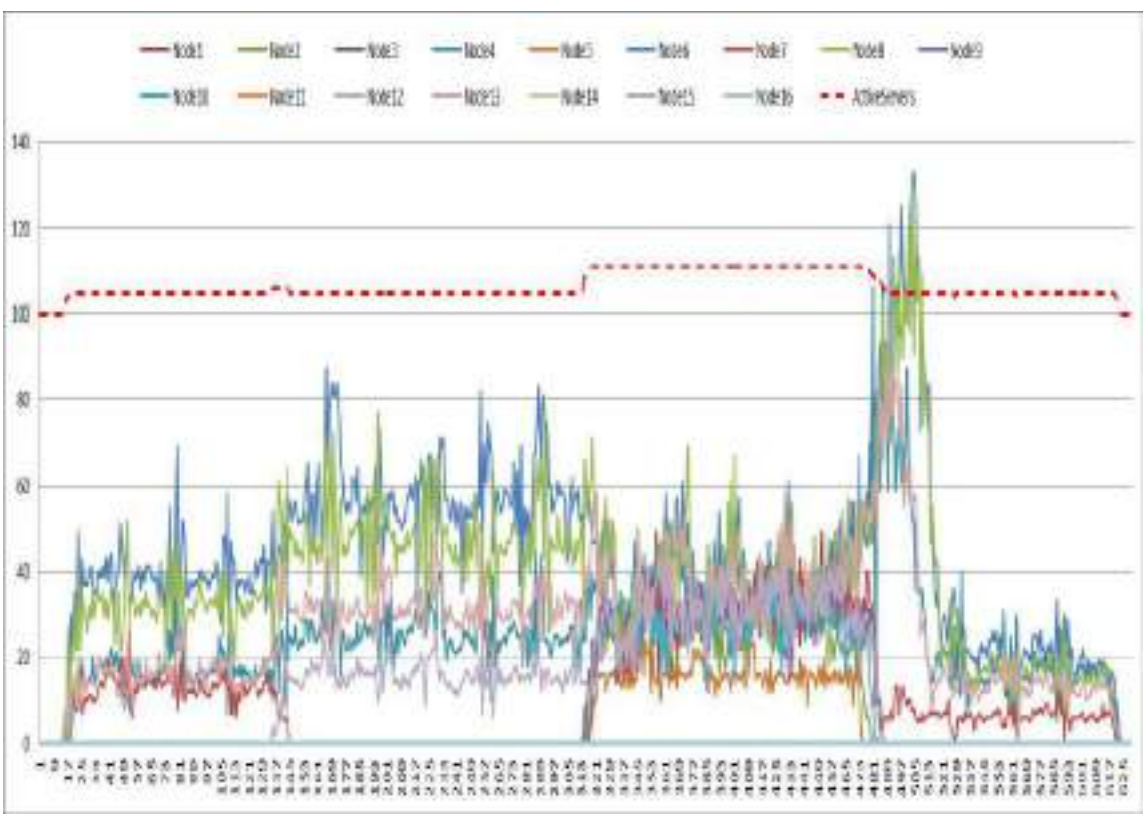

Figure 7: Comparison of the energy consumption with traffic intensity based controller and without controller

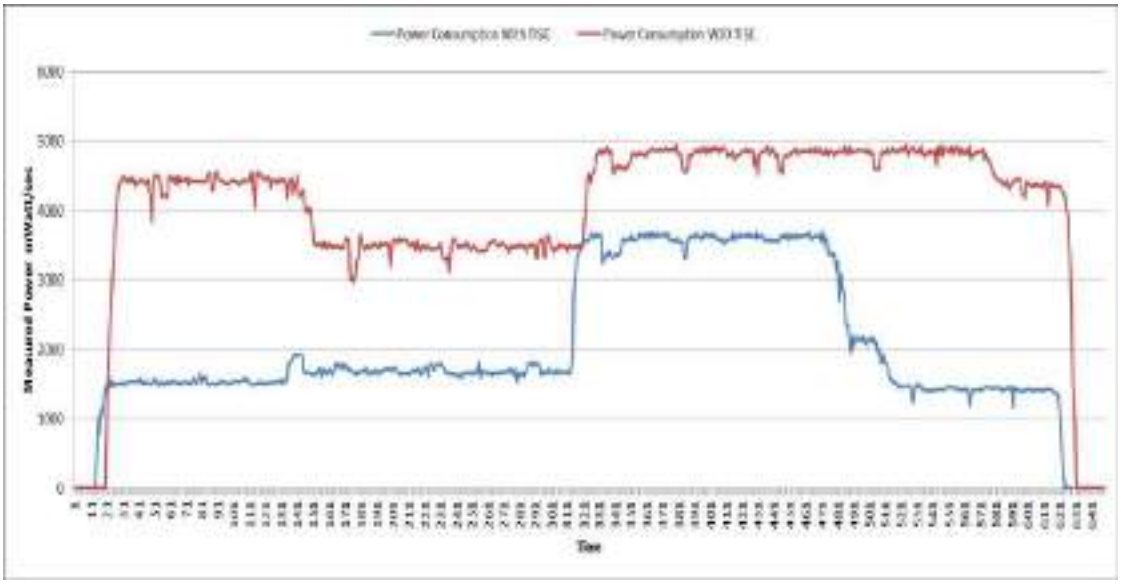

Figure 8: Comparison of the energy consumption with traffic intensity based controller and without Controller

The comparison of the energy consumption with traffic intensity based controller and without it is shown in figure 8 and figure 9. Cumulative energy utilization in the datacenter is shown for both cases in figure 10. The energy curve for the datacenter with TISC always remains below the energy curve of the datacenter without TISC. TISC results in a significant reduction in power consumption and the idle servers running. This also enables a reduction in the idle network and other related infrastructure. However, the impact of these is not included in this work.

The overall difference in energy saving is almost half of the traditional data center approaches. However, if the networking and other infrastructures are added to the analysis, it will result in large energy savings.

4.2. Virtual Container Environment (VCE)

VCE provides a data center function. The datacenter nodes are Docker-based containers that run various services including a WebApp, Database server, and related Linux packages.

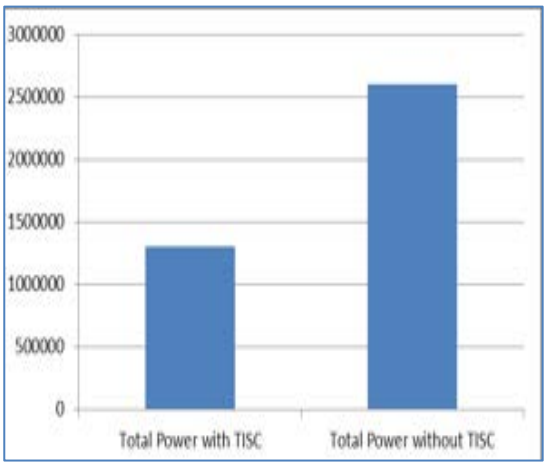

Figure 9: Energy consumption with traffic intensity based controller and without Controller

All the datacenter nodes are homogeneous i.e. all nodes provide the same functionality or same service. They have a similar CPU, Memory, and Storage Specification, so that when a user requests a service, it may be allocated to any of the servers usually having the least load.

4.3. Network Simulation Environment (NSE)

NSE is NS3 based simulation environment that provides network topology to simulate the model network and generate requests that are routed to VCE.

4.4. Traffic Intensity Based Container Controller (TIC)

TIC is a traffic intensity based container controller. Its job is to monitor the data center nodes and their load and it decides whether a node has to be switched off or switched on and hence providing a basis for saving energy.

TIC has two important functions one is to ensure that it has correct traffic profile information and the second it has correct node information and its load and estimated remaining processing time. TIC also schedules the requests and assignments to serve nodes. There are different approaches for the scheduling and allocation of requests to servers e.g. priority, RR, WRR, etc. in this research simple round-robin methodology was adopted among the least loaded and on data center nodes.

4.5. Container Power Analyzer (CPA)

CPA's role is to calculate the power consumed based on the predetermined calibration of power consumption behavior on the host system. CPA records the power consumption per minute that is father used for comparison purposes.

4.6. Traffic Model

Users connected to access switches generate requests for server resources and generate load as per user requirements. The traffic intensity profile for these nodes is randomly chosen with no predetermined rule. However, while designing the network as well as data center capacity planning and dimensioning Erlang principles are used so that the user load can be efficiently met by the network & data centers. The specific details of capacity planning and dimensioning shall not be covered in this paper rather a generic Erlang model is referred to.

Figure 11 shows the traffic model of uplinks of the access nodes of datacenters. It represents several requested generated by users corresponding to the given access node. Figure 12 shows the same values but in a different view

#/Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Node 1	10	10	10	10	0	0	0	0	0	0	18	18	18	18	18	2	2	2	2	2
Node 2	5	5	5	5	10	10	10	10	10	10	15	15	15	15	15	4	4	4	4	4
Node 3	4	4	4	4	15	15	15	15	15	15	20	20	20	20	20	5	5	5	5	5
Node 4	12	12	12	12	18	18	18	18	18	18	10	10	10	10	10	6	6	6	6	6
Node 5	15	15	15	15	10	10	10	10	10	10	5	5	5	5	5	7	7	7	7	7
Node 6	5	5	5	5	8	8	8	8	8	8	15	15	15	15	15	8	8	8	8	8
Node 7	4	4	4	4	2	2	2	2	2	2	10	10	10	10	10	10	10	10	10	10
Node 8	3	3	3	3	5	5	5	5	5	5	20	20	20	20	20	3	3	3	3	3
Node 9	2	2	2	2	6	6	6	6	6	6	18	18	18	18	18	4	4	4	4	4
Node 10	1	1	1	1	7	7	7	7	7	7	8	8	8	8	8	5	5	5	5	5
Node 11	2	2	2	2	2	2	2	2	2	2	9	9	9	9	9	4	4	4	4	4
Node 12	4	4	4	4	1	1	1	1	1	1	10	10	10	10	10	4	4	4	4	4
Node 13	5	5	5	5	2	2	2	2	2	2	15	15	15	15	15	8	8	8	8	8
Node 14	10	10	10	10	3	3	3	3	3	3	4	4	4	4	4	8	8	8	8	8
Node 15	2	2	2	2	5	5	5	5	5	5	3	3	3	3	3	2	2	2	2	2
Node 16	1	1	1	1	6	6	6	6	6	6	1	1	1	1	1	1	1	1	1	1

Figure 10: Traffic model of uplinks of the access nodes of datacenters

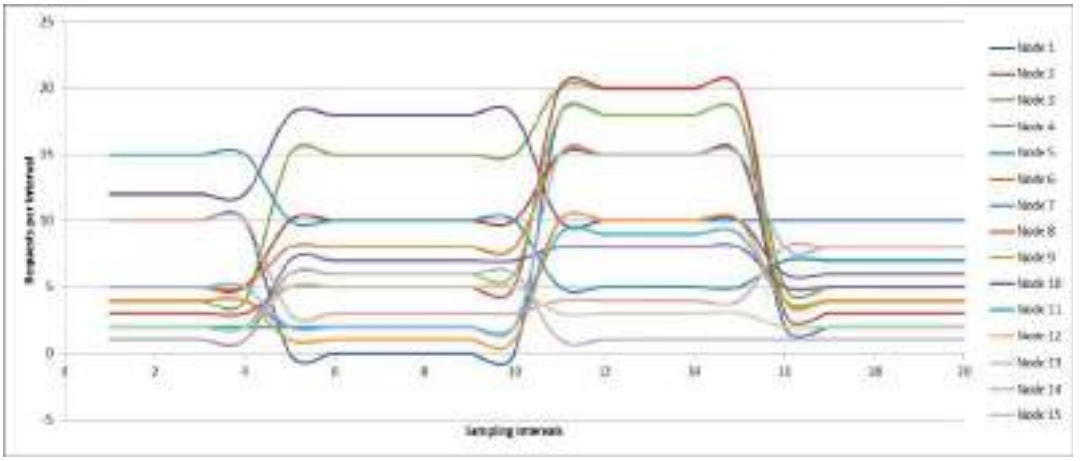

Figure 11: Graphical view of traffic model of uplinks of the access nodes of datacenters

Table 1: Traffic Intensity

Total Power with TISC	1306981	33.40%
Total Power without TISC	2606053	66.60%
Total	3913034	49.85%

4.7. Characteristics of the Traffic Model

The particular characterization of the traffic model selected was that nodes generate traffic in a randomly distributed fashion and the node distribution is also sparse. Second important thing is that traffic generated from different access nodes never exceeds the max limit assisted to the access pop. This ensures that the network capacity planning and dimensioning have been carried out in a way to carters the maximum needs of the users.

4.8. Traffic Intensity Information:

Consider is the traffic intensity of the node n and be the sampling time, be the average session time, be the actual requests sent from a node, be the request serving capacity of a single node. Then, We can define an access node, $TI_n = \frac{R_A * T_{SES}}{C_R * T_S}$,

We have selected a hypothetical request model where user requests are randomly distributed across the sampling intervals. The selected request model and the corresponding traffic intensity are given in table 1 and table 2 respectively.

Figures 11 and 12 show the traffic intensities corresponding to the traffic intensity of each access pop and it shows that some characteristics of the traffic are preserved.

5. EVALUATION METHODOLOGY

This section discusses the evaluation and comparison methodology of the simulation results. The simulation is carried out in three steps. In the first step, a predefined traffic model is used for single server single access pop. In this phase following relationships are established.

5.1. Phase -1

- requested generated by users
- Request completion time
- Respective CPU load

- Respective power consumption by for node
- Respective power consumption by the network interfaces

5.2. Phase -2

In the second phase, all the access pops generate random traffic, and requests are served in a routine procedure. During this step, the following are observed.

- The total requests generated by users
- Total requests served by the datacenter
- Average request completion time

5.3. Phase -3

In phase 3 access nodes generate traffic in a predefined pattern governed by the traffic intensity profile and the controller applies the traffic intensity-based algorithm to route the requests to nodes based on the load.

This research work focuses on an abstract level comparison of the overall energy utilization in the datacenter.

5.4. Phase -4

It calculates the energy consumption traditionally without implementing any energy optimization algorithm. In this case, all servers are switched on and incoming requests are allocated in the round-robin fashion. Some servers in the case may be overloaded and some servers may be under loaded due to the different amount of computation or request time. Since there is no way of considering the traffic intensity so requests are allocated in a simple round robin fashion regardless of the load characteristics.

5.5. Phase -5

In this phase traffic intensity based energy efficiency algorithm is implemented that tracks the current server utilization, evaluates the current TI characteristics, dynamically adds and removes the active servers. The requests are allocated in a round robin fashion among the candidate servers rather than all servers.

Both phases are run for a fixed duration and server load, power consumption, memory utilization, request completion rate, and request completion latencies are monitored and compared with each other.

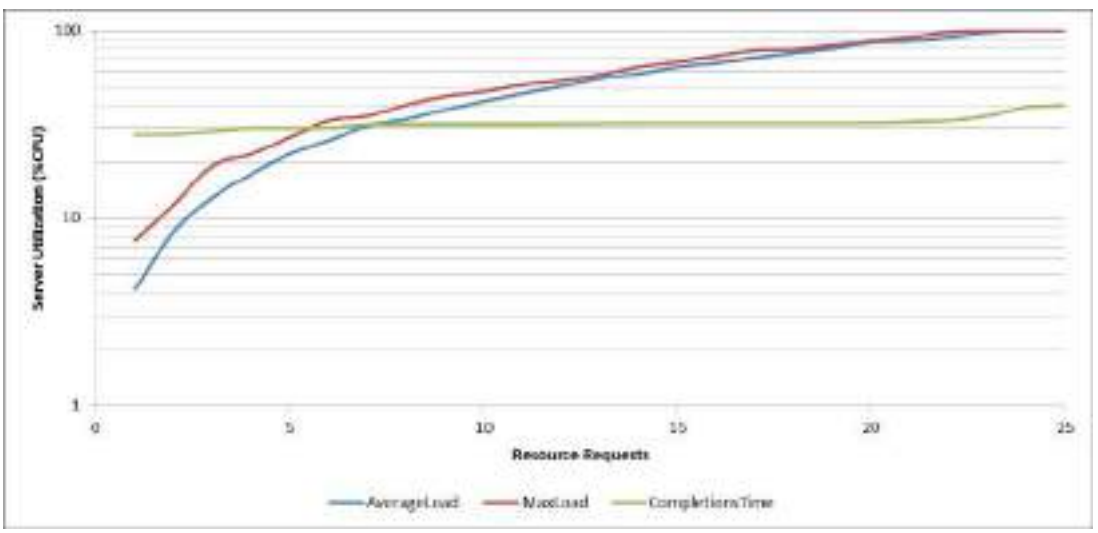

Figure 12: average request completion time with respect to server load

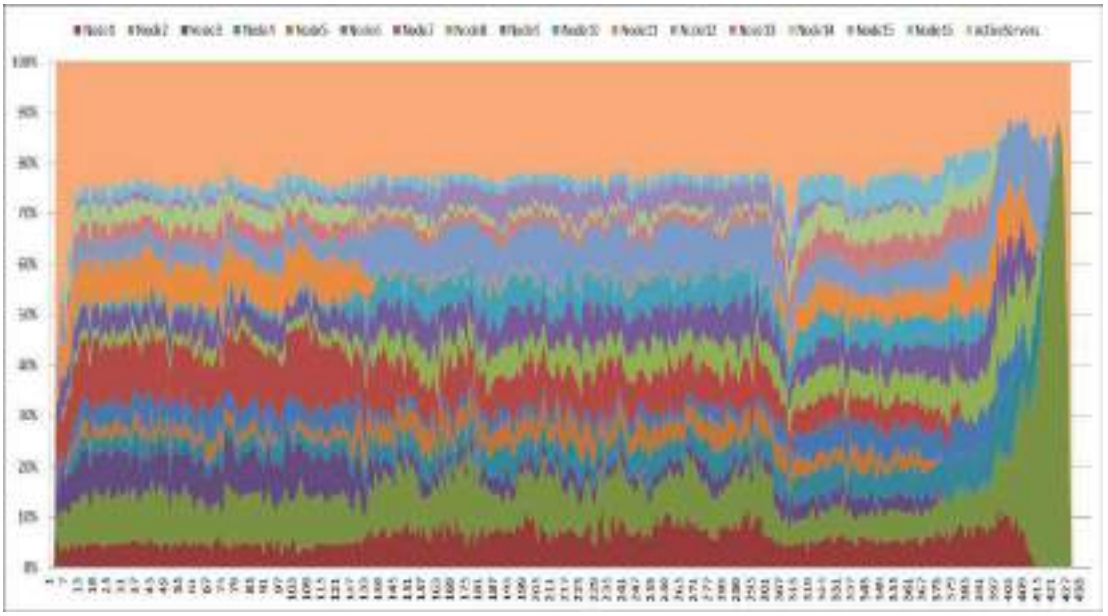

Figure 13: Servers Utilization with Random Load without Traffic Intensity Based Server Controller (TISC)

Figure 13 shows the average request completion time concerning server load. It shows that as the server load increases request completion time only changes after a certain threshold.

6. DISCUSSION

Finally, the percentage of energy consumption as observed without implementing Traffic Intensity Based Server Controller 66.60%. While, we implemented it Traffic Intensity Based Server Controller 33.40%. The energy efficiency achieved in this model is 49.85%.

7. CONCLUSION

This work analyzed the impact of Traffic intensity-based energy efficiency for a modern datacenter. All the analysis was based on a virtual data center simulated using Linux containers running various services like file transfers, voice sessions, and email and webs services. The results presented in this work show that TISC based approach in modern datacenter results in significant energy savings, esp. due to the elimination of the idle servers and idle infrastructure. Further TISC based load balancing results freeing up more servers that are underutilized. Both of these approaches result in a significant reduction in power consumption.

In this work, the analysis of the networking and related infrastructure was not considered that apparently can improve the results due to known idle components available to turn off or hibernate. Furthermore, in this work, we did not consider the hibernation on the servers that may eliminate the issues with server startup and shutdown. The restart operation on servers has its side effects.

REFERENCES

[1] Khan, S.U.; Zomaya, A.Y. *Handbook on data centers*; 2015.

[2] Heller, B.; Seetharaman, S.; Mahadevan, P.; Yiakoumis, Y.; Sharma, P.; Banerjee, S.; "McKeown, N. ElasticTree: Saving Energy in Data Center Networks." In Proceedings of the USENIX Symposium on Networked Systems Design and Implementation, NSDI 2010; USENIX Association Berkeley, CA, USA: 28-30 April 2010; San Jose, California, pp. 249–264,2010.

[3] Heller, B.; Seetharaman, S.; Mahadevan, P.; Yiakoumis, Y.; Sharma, P.; Banerjee, S.; "McKeown, N. ElasticTree : Saving Energy in Data Center Networks. *Proc". 7th USENIX Conf. Networked Syst. Des. Implement.*, 17–17, 2010.

[4] Shang, Y.; Li, D.; Xu, M. "Energy-aware routing in data center network." In Proceedings of the Proceedings of the first

ACM SIGCOMM workshop on Green networking - Green Networking '10; ACM: 30 August - 03 September 2010; New Delhi, India, pp. 1–8, 2010.

[5] Nguyen, K.K.; Cheriet, M.; Lemay, M.; Savoie, M.; Ho, B. "Powering a data center network via renewable energy:" A green testbed. *IEEE Internet Comput.*, *17*, 40–49, 2013.

[6] Arlitt, M.; Bash, C.; Blagodurov, S.; Chen, Y.; Christian, T.; Gmach, D.; Hyser, C.; Kumari, N.; Liu, Z.; Marwah, M.; et al. *Towards the design and operation of net-zero energy data centers*; 2012.

[7] David, H.; Fallin, C.; Gorbatov, E.; Hanebutte, U.; Mutlu, O. *Memory power management via dynamic voltage/frequency scaling*; 2011;

[8] Wang, T.; Su, Z.; Xia, Y.; Qin, B.; Hamdi, M. "Towards cost-effective and low latency data center network architecture". *Comput. Commun., 82*, 1–12 2016.

[9] Valancius, V.; Laoutaris, N.; Massoulié, L.; Diot, C.; Rodriguez, P. "Greening the internet with nano data centers." *Proc. 5th Int. Conf. Emerg. Netw. Exp. Technol. Conex. 09*, 37–48, 2009.

[10] Baccour, E.; Foufou, S.; Ridha, H.; Hamdi, M. *wFlatnet: Introducing Wireless in Flatnet Data Center Network*; 2015;

[11] Kim, J.; Dally, W.; Abts, D. "*Flattened butterfly:" a cost-efficient topology for high-radix networks.*; Vol. 35, 2007.

[12] Baccour, E.; Foufou, S.; Hamila, R.; Hamdi, M. A survey of wireless data center networks. *2015 49th Annu. Conf. Inf. Sci. Syst.* 2015, 1–6.

[13] Guan, X.; Choi, B.Y.; Song, S. "Energy efficient virtual network embedding for green data centers using data center topology and future migration. *Comput". Commun.*, *69*, 50–59,2015.

[14] Shin, J.Y.; Sirer, E.G.; Weatherspoon, H.; Kirovski, D. On the feasibility of completely wirelesss datacenters. *IEEE/ACM Trans. Netw.*,*21*, 1666–1679,2013.

[15] Vmware infrastructure architecture overview.

[16] Liu, L.; Wang, H.; Liu, X.; Jin, X.; He, W.; Wang, Q.; Chen, Y. GreenCloud: "a new architecture for green data center. In Proceedings of the Proceedings of the 6th International Conference Industry Session on Autonomic Computing and Communications" Industry Session, ACM, pp. 29–38,2009.

[17] Georgiou, S.; Tsakalozos, K.; Delis, A. "Exploiting network-topology awareness for vm placement in iaas clouds. In Proceedings of the Cloud and Green Computing (CGC)", 2013 Third International Conference on; IEEE, pp. 151–158, 2013.

[18] Ruiu, P.; Bianco, A.; Fiandrino, C.; Giaccone, P.; Kliazovich, D. *Power Comparison of Cloud Data Center Architectures*; 2016;

[19] Carrega, A.; Singh, S.; Bolla, R.; Bruschi, R. "Applying traffic merging to datacenter networks. In Proceedings of the Future Energy Systems:" Where Energy, Computing and Communication Meet (e-Energy), 2012 Third International Conference; pp. 1–9, 2012.

[20] Wang, T.; Xia, Y.; Muppala, J.; Hamdi, M. "Achieving Energy Efficiency in Data Centers Using an Artificial Intelligence Abstraction Model." *IEEE Trans. Cloud Comput.*, 1–1, 2015.

[21] Fang, S.; Li, H.; Foh, C.H.; Wen, Y.; Aung, K.M.M." Energy optimizations for data center network: Formulation and its solution. In

Proceedings of the Global Communications" Conference (GLOBECOM), IEEE; 2012; pp. 3256–3261, 2012.

[22] Al-Fares, M.; Loukissas, A.; Vahdat, A. "A scalable, commodity data center network architecture." *SIGCOMM Comput. Commun. Rev. 38*, 63–74, 2008.

[23] Berl, A.; Gelenbe, E.; Di Girolamo, M.; Giuliani, G.; De Meer, H.; Dang, M.Q.; Pentikousis, K. "Energy-efficient cloud computing. *Comput. "J.*, *53*, 1045–1051, 2010.

[24] Cook, G. How clean is your cloud. *Catal. an energy Revolut.* 2012.

[25] International, G. Greenpeace “likes” Facebook’s new data center, but wants a greener friendship. 2011.

[26] Koomey Growth in Data Center Electricity use 2005 to 2010. *Oakland, CA Anal. Press.* ,*34008*, 1–24, 2011.

[27] Meisner, D.; Gold, B.T.; Wenisch, T.F. PowerNap: "eliminating server idle power." *Asplos*, *44*, 205–216, 2009.

[28] Koomey, J.G. Growth in Data Center Electricity use 2005 to 2010." *Anal. Press.* 1–24, 2011,.

[29] Akella, A.; Benson, T.; Chandrasekaran, B.; Huang, C.; Maggs, B.; Maltz, D. "A universal approach to data center network design." In Proceedings of the Proceedings of the 2015 International Conference on Distributed Computing and Networking; p. 41 ,2015.

[30] Tso, F.P.; Pezaros, D.P. "Improving data center network utilization using near-optimal traffic engineering". *IEEE Trans. Parallel Distrib. Syst.*, *24*, 1139–1148, 2013.

[31] Giroire, F.; Moulierac, J.; Phan, T.K.; Roudaut, F." Minimization of network power consumption with redundancy elimination". *Comput. Commun, 59*, 98–105, 2015.



Waseem et al LGURJCSIT 2021

ISSN: 2521-0122 (Online)
ISSN: 2519-7991(Print)

LGU Research Journal of
Computer Science and IT

Vol.5 Issue1, Jan-Mar 2021

Classifying Urdu Verbs Using Rule Based Approach

Muhammad Waseem[1], Qaiser Abbas[2], Muhammad Ilyas[3], Saad Razzaq[4], Sadia Niazi[5], Ijaz Asghar[6]

[1234]Department of Computer Science & IT, University of Sargodha, Sargodha, 40100, Pakistan

[5]Department of Psychology, University of Sargodha, Sargodha, 40100, Pakistan

[6]Department of English, University of Sargodha, Sargodha, 40100, Pakistan

[1]wasibhatti27@gmail.com

ABSTRACT

To make dictionaries complete and to keep their size restricted, there is an approach in the linguistic world to equip these dictionaries with morphological information. This module of morphological information is usually known as a morphological analyzer or morphological classifier, which normally contains the complete possible linguistic information about each word for that particular language and it also describes the rules of derivations from the root of a word and its various inflections, respectively. In this work, a classifier for Urdu verbs (CUV) is proposed which is still a challenging research issue, as Urdu is a language of high inflection and derivation. The available stemmers for Urdu do not provide enough information about inflectional and derivational forms of words. Also, morphological classifiers available for Urdu are not worthy of handling various problems and delivering results that prune errors. In our work, a rule based CUV is designed which is able to classify 63 forms of Urdu verbs successfully out of 66. Available Urdu language processing tools are very rare compared to other higher inflectional languages such as German, Turkish, etc., which have competitive morphological classifiers. However, the studies related to Urdu verb morphological classification are identified and a comparative study is presented in this article. In short, this work is a positive contribution to the community, and it provides sufficient information with promising results specifically on inflectional and derivational forms of Urdu verbs.

KEYWORDS: Urdu Morphology, Urdu Verbs Classification, Rule Based Model, Verb Analyzer, CUV Algorithm.

1. INTRODUCTION

The main purpose of morphological investigation or analysis is to study the change from surface to the lexical portrayal. On the other hand, the converse of morphological investigation is known as a morphological generation. The fundamental approach to do this is the development of transducers or finite state machines which are implicitly bi-directional and hence can be used for both the mentioned purposes e.g., analysis and generation. This is due to the coding of rules that we have to do for once which then can be used for analysis and also for a generation. Morphological analysis or classification can be performed through probabilistic classifiers like Naïve Bayes, K Nearest Neighbors, Decision Trees, Support Vector Machines, etc., but in our work, we present the rule-based approach (length based or affix stripping) to Urdu verb classification. Morphological studies for various languages are discussed in Section 2 of the literature review.

Urdu is an Indo-European language and is gotten predominantly from Arabic, Persian and Turkish dialects [1]. As per a guess, there are around 450 million individuals who communicate in Urdu around the world. In Pakistan, this language has a national status while in India, it is considered an official language. Due to Persian and Arabic inheritance, it uses the same right-to-left Perso-Arabic script for writing text. Urdu has free orderliness and rich in terms of morphology [2] and can obtain words from different dialects promptly. Urdu takes over the majority of the jargon from Arabic [3] and Persian. Hindi has a different script, however, Urdu and Hindi both have the same structures in terms of phonology, morphology and syntax. A few etymologists thought about them as two distinct tongues of one language. The primary contrast between Hindi and Urdu is that Hindi uses Devanagri drought composed from left to right while Urdu uses Perso-Arabic drought composed from right to left in Nastaleeq script. Due to the cursive nature of the Nastaleeq script, the

letters in a word normally connect with each other. Like other scripts, context-sensitivity is present in it and is normally considered to be a complicated system of writing. In Urdu Zabta Takhti 1.01 [4], Urdu consists of 57 letters but according to Siddiqui Urdu has 52 letters[1]. The difference of opinion is due to some disputed number of letters in some words and vowels. Urdu character set is divided into three parts, which are alphabets, vowels and other symbols.

Urdu is a morphologically rich language and is able to borrow words from other languages readily. Urdu morphology is a complex phenomenon. Many regional languages besides Arabic, Persian and Hindi have a deep impact on Urdu formation and development. Derivational affixes also regularly occur in Urdu. Words generally have no spaces between them but for computational purposes separated by spaces. Morphological studies in Urdu are quite low in number, however, the studies that have some similarity are discussed in Section 3 of related work. Then we portray aspects of Urdu morphology shortly for the better understanding of a reader in Section 4.

In this work, our focus is on the single-word verb. Inflection of single verb classification is the multiple of tense, aspect, gender, number and mood, summing up to 66 different classified forms which we attempted to generate using rule-based approach presented in Section 5. For representing correct tense, aspect and mood, many verb auxiliaries are also used with these verbs. These verb auxiliaries do inflect with respect to tense, aspect, mood, gender and number. A regular verb is basically represented in three basic stem classifications, which are root, causative one and causative two. Morphology applied to the root form can also be applied to the causative one and causative two forms. The algorithm designed can classify the form of verbs and is presented in Section 5 along with the interface and running example (Section 5.1). Results are critically analyzed in comparison to related work in Section 6.

2. LITERATURE REVIEW

In [5], Koskenniemi's proposed a two-level morphological model. This was the first model in computational linguistics for the investigation of morphologically rich language. However, still, only two well-known strategies are existing yet with respect to our work for the investigation of morphology. One strategy is stemming, and we can observe a number of stemmers exist for languages around the globe with their level of accuracy. This strategy of stemming was spoken to by Porter in [6] and another strategy is morphologically parsing [7]. This method of parsing was first portrayed by Koskenniemi in his work on two-level morphology [5].

One of the two strategies discussed above is Porter's algorithm [6]. It is comprised of two phases. The first phase deducted common endings from inflected words known as de-suffixing, and the second phase added a potential consummation component to the string acquired in the former point. This is known as a re-encoding phase. The two phases could perform successively or simultaneously. The algorithm's distinctive feature was that it didn't utilize a dictionary, which made it suitable for morphological investigation.

Another strategy from above is the Koskenniemi model [5]. It contemplated morphological investigation and morphological change in various situations with respect to phonological information. It used a finite state transducer (FST) to encrypt correspondence. This correspondence was between the root/surface form of the word and the word itself, for example, a word form: f l y + s from root/surface form to f l i e s.

KIMMO by Karttunen [8] was the first morphological analysis system that had adopted the Koskenniemi model. This system contained two sections. The first section was about rules and the second section was about lexical features. These lexical features included morpheme construction and morphosyntactic constraints. Techniques of machine learning (ML) were used for obtaining morphological rules discussed in the first section of KIMMO, while supervised and unsupervised methods of ML could usefully classify the two components in the second section of KIMMO. The supervised learning methods [9-10] could exploit the morphological information from the lexical database e.g., CELEX2 was one of those vocabulary databases. On the other hand, unsupervised learning methods [11-13] used a words-list with or without frequency along the way.

3. RELATED WORK

In 2004, Sara Hussain [14] purposed a finite state morphological analyzer for Urdu. It was based on a directed acyclic morphological structure and was named MORPH. Sara analyzed frequently used Urdu verbs first, and then she focused on frequent nouns, and finally studied affixes and presented closed-class words. An online available lexicon[2] was used for comparative analysis. Words with higher frequency (as per use) were chosen for analysis only, hence a limited study. It could not handle all aerabs[3] (diacritics) given in a text file and hence all were ignored during the preprocessing step.

[1] https://www.dawn.com/news/919270/controversy-over-number-of-letters-in-urdu-alphabet

[2] http://www.lc-star.org/docs/LC-STAR_D1.1_v1.3.doc)

[3] Aerabs are short symbol which change the sound of word in Arabic and Urdu as well.

In 2006, Humayon et al [15] proposed work on Urdu morphology, orthography and lexicon extraction. Humayon chose Urdu functional morphology [4] because Russian, Italian, Spanish, Swedish and Latin functional morphology was successfully implemented before this work. He described that morphological analysis could be done on a single word or words in combination. In his work of single word analysis, each word in the sentence was analyzed separately. In combination, words with postpositions and auxiliaries were analyzed. This study focused mainly on noun clitics. It further shed light on limited verb-auxiliaries, and finally, composition of adjective clitics was presented.

Tina Boegel, et al [16] purposed a finite state morphological analyzer (FSMA) for Urdu/Hindi. A number of issues were addressed for building an FSMA. Among the issues, potential ambiguity and the non-concatenative morphological issues were particular. Via a cascade of finite-state transducer approaches, the work allowed similar treatment for both Urdu and Hindi languages. Transducer translated both Urdu and Hindi scripts to a common ASCII transcript. The transliteration system was based on Xerox finite-state technology tools. No compatibility issue arose in transliteration. However, the major issue in designing a transliteration system was that both Hindi and Urdu were written in different scripts. This issue was resolved by processing both scripts to a common script.

Studies presented are not exactly according to the line that we have chosen in our work which is affix stripping or length-based approach. These related studies have somehow similarity with respect to encoding of rule and the existing accuracy for the length-based models is 79.63% [31-30].

4. URDU MORPHOLOGY ASPECTS

In morphology, words belong to a particular lexical category, which can be nouns, verbs, adjectives, prepositions, adverbs, etc. Some are discussed as follows.

4.1. Noun Morphology

Urdu nouns are categorized by gender, number, and case. Nouns can be masculine or feminine with respect to gender, for example, لڑکا laRkA/boy is a masculine noun whereas لڑکی laRkI/girl is a feminine noun. In case of number, nouns can be singular and plural, for example گھڑی gHaRI/watch is a singular noun and گھڑیاں gHaRIYAN/watches is a plural noun. The noun has three cases, which are direct, oblique and vocative, for example, لڑکا laRkA/boy is the direct case while لڑکوں laRkoN/boys is an oblique case of noun لڑکا laRkA/boy. The vocative case has two more forms, which are singular and plural. The singular vocative form is لڑکے laRkE/boy and plural vocative case is لڑکو laRko/boys of noun لڑکا laRkA/boy. Nouns may be further divided into two classes based on declension or group, which are called type-I (marked) and type-II (unmarked) nouns. The basic difference between these two categories is that the former has characteristic terminations in the direct singular while the latter does not. Nouns marked for gender are called marked nouns while nouns that have no special gender suffix are called unmarked nouns. Noun ending on masculine or feminine suffix is called marked noun for example لڑکا laRkA/boy and بچہ bacah/child are masculine marked nouns whereas لڑکی laRkI/girl and بچی bacI/baby-girl are feminine marked nouns. Nouns that do not end on gender suffix are called unmarked nouns, for example, کام kAm/work and کتاب kItAb/book are unmarked nouns.

4.2. Pronoun Morphology

There is no gender distinction for Pronouns. Same Pronouns are used for both masculine and feminine. Pronouns are categorized with respect to number as singular and plural in all persons. For example, first-person single pronoun is میں mEN/I and first-person plural pronoun is ہم Ham/we. Gradation of acquaintance or regard is used in second and third persons, for example تو tU/you and تم tum/you pronouns are used for acquaintanceship while آپ aAp/you pronoun is used for regard in second-person. In third-person distance from speakers is also involved, for example وہ woh/that and یہ yEh/this. Pronouns take direct, oblique and vocative cases in the same manner as nouns can take. Nominative pronoun most commonly occurs as the subject of verb except for transitive verbs in perfect tenses where subject takes نے /nE. Similarly, when a pronoun is trailed by postposition کو 'ko', کا 'kA', میں 'mEN'/I and سے 'sE'/for, it occurs in an oblique case. The second person recognizes three degrees of acquaintanceship that reveal connection among speaker and individual tended to or speaker disposition towards that individual.

4.3. Adjective Morphology

Adjectives are categorized as marked/stamped and unmarked/plain adjectives. Plain adjectives have an exposed stem in the illustration. Unmarked or plain adjectives have no special suffix and do not change to show agreement, for example, دلچسپ کہانی 'dilcasp kahAnI'/interesting-story, تازہ آم 'tAzah aAm'/fresh-mango, etc. On the other hand, marked adjectives have a variety. It can change its inflection with the gender of the following word, for example, اچھا کام 'acHA kAm'/good-work, اچھی بات 'acHI bAt'/good-talk, etc. Here adjective اچھا 'acHA'/good and اچھی 'acHI'/good are marked for the masculine and feminine property of the following words. Stamped adjectives concur with nouns and have the ability to

[4] Functional morphology is the analysis of the mechanical and evolutionary relationship of anatomical form to organismal behavior and dynamics

adjust itself with the gender, number and case with a little arrangement of inflectional postfixes. In contrast, plain adjectives have just single invariant structure. In conclusion, the inflection of adjectives depends on the inflection of nouns [17-18].

4.4. Adverb Morphology

Urdu verbs have two different groups of words, one which are original adverbs called underived adverbs and another is originally other class of verbs but used as adverbs. According to Schmidt [19], Urdu adverbs are divided into five classifications, which are called adverbs of time, place, manner, degree and modal. Adverb of time describes the time where an event is happening, for example, ہمیشہ 'HamESah'/always, کل 'kal'/yesterday, etc. Adverb of place describes the location where an event happened, for example, باہر 'bAhir'/out, اندر 'andar'/in, etc. Adverb of manner describes how an event happened, for example, یوں کرو 'yUN karo'/do-like-this, کیجیے یوں 'yUN kIjIyE'/do-like-this, etc. Adverb of degree describes the intensity of happening of event, for example, بڑا آدمی 'baRA aAdmI'/big-man, زیادہ گندہ 'ziyAdah gandah'/more-dirty. Modal adverb [20] modifies a verb or adjective. For example, صرف 'sirf'/only, شاید 'SAhId'/Perhaps, etc.

Adverb does not change the form with the sentence structure. Mostly derivational information is learned from the adverb form. It shows what components the word has been shaped. If an adverb appears in a text, then it is searched in the dictionary as it appears in the text. Most of the adjectives and adverbs fall into the following categories including the concept of proximity like near and far, demonstratives pointing to some object, interrogatives having the concept of questions and relatives.

4.5. Verb Morphology

Urdu verb is more perplexing when contrasted with other word classes [21]. The verb can appear in different forms to show variations in the happening of activities or a particular event. The verb can be classified as intransitive, transitive and di-transitive with respect to a number of object phrases and subject phrases taken by these verbs [22]. Intransitive verbs are those verbs, which take zero verb phrases. Transitive verbs are those, which take at least one verb phrase with respect to the subject phrase. Di-transitive is those, which take at least two or more verb phrases with respect to subject phrases. Verb or predicate can form complex predicates and can have regular or irregular morphology. Transitive and Intransitive verb forms have regular morphology and are called regular verb forms [23]. Di-transitive forms are called irregular forms having irregular morphology. Different morphological classifications that show variation in verb meaning [24] are root, perfective, imperfective, imperative, subjunctive, infinitive, causative one and causative two forms.

Urdu has a rich morphological verb. Verb forms are commonly classified into five groups which are called infinitive, perfective, repeated, subjunctive and imperative. Urdu also has three stem forms, classified as base, one causative and two causative forms. By adding these three basic forms to different morphemes, they are further divided into gender, number, person and mood which establishes 60 classifications of verb in respect of forms. Such 60 classifications of verb don't have all the characteristics of the verb. With these sixty types, many verb auxiliaries and light verbs are used which present different characteristics of verbs. Combining verb auxiliaries is appropriate at the syntactic level. It is important to combine auxiliaries at the syntactic level of the verb. If the auxiliaries are lumped at the lexical level with verb forms, then the requirement for complex agreement is avoided.

As per Urdu verbs, adding certain morphemes creates various inflected forms or classifications. Different auxiliaries and suffixes are combined with these five forms. With this combination, tense and aspect inflections are formed in a wide scope. Tense is indicated through auxiliary verbs in Urdu. As for as the agreement in verbs is concerned, number and person of the argument structure of the sentence is important which plays a vital role in the inflection of verbs.

In this work our focus is on single-word verb. The inflection of single verb expands in terms of tense, aspect, gender, number and mood. For representing correct tense, aspect and mood, many verb auxiliaries are also used with these verbs. These verb auxiliaries do inflect with respect to tense, aspect, mood, gender and number. A regular verb is basically represented in three basic stem forms, which are root, causative one and causative two. Morphology applied to the root form can also be applied to the causative one and causative two forms.

5. METHODOLOGY

In building Classifier for the Urdu Verbs (CUV), our approach is completely rule based without any addition of training datasets, databases or dictionaries, which were practiced in related work [14-15], and others as discussed in Section 3. We use length-based suffix removal approach for cutting suffixes in descending order. More lengthy suffixes are checked and removed before short length suffixes. By doing this, the correctness of the approach is reasonable. The basic rules finalized are implemented in the design of the CUV algorithm, which describes the procedural model [25] to solve our problem of morphological classification or analysis. The pseudocode of the algorithm is as follows.

CUV ALGORITHM: Inputs a verb and provides its 66 different classifications.

1. //Input a verb for the required information
2. Input ←single word x
3. //Initialization.
4. Base_form ←null
5. Inflect_form ← null
6. Verb_type ← null
7. Index ← null
8. //Check that input verb is not null.
9. If x ≠ null
10. {
11. // if x be in exception list then its morphological information is picked up from that list.
12. If x is in exception_list V
13. {
14. index← null
15. Baseform ← V.baseform
16. Inflectform←V.inflectform
17. Verbtype ← irregular verb
18. }
19. // Check for valid verb inflection with respect to standard rules defined
20. Else If name ends with any of inflection form
21. {
22. Index ← indexOf(inflectionform)
23. Baseform ← substring (0, index)
24. Inflectform ← substring(index)
25. Verbtype ← form out of verb form
26. }
27. Else
28. {
29. // if no verb inflection identified then put the input verb as root.
30. baseform ← name
31. inflectform ← null
32. verbtype ← Root Form
33. }
34. // Morphological information of the input word is printed.
35. Output: inputword, rootword, baseform, inflectform, verbtype
36. // First check the baseform and then Print all the possible 63 inflection forms of the input word.
37. Output: If baseform← !null
38. {
39. Print baseform +infinitive form
40. Print baseform + imperfective form
41. Print baseform + Perfective form
42. Print baseform + subjunctive form
43. Print baseform + imperative form
44. Print baseform + subjuctive form
45. Print base form + causative one form + Infinitive form
46. Print base form + causative one form + Imperfective form
47. Print base form + causative one form + Perfective form
48. Print base form + causative one form + Subjunctive form
49. Print base form + causative two form
50. Print base form + causative two form + Imperfective form
51. Print baseform + causative two form + perfective form
52. Print baseform + causative two form + subjunctive form
53. }
54. // Do not print anything if baseform is null.
55. Else
56. Output: no word
57. }

Procedural description of the algorithm along with an example is presented next. For the extraction of rules, a list[5] available at CRULP is studied. This list contains 1540 verbs which are collected from the Online Urdu Dictionary.[6] Due to morphological issues that list is pruned to 1299 verbs that we finally used in our proposed work. A detailed discussion is available in Section 6.

5.1 RUNNING EXAMPLE

Evaluation of CUV algorithm is best described by running it on a concrete example. If word کھیلنا

[5] [5] It is currently available at http://www.cle.org.pk/software/ling_resources/urduverblist.htm
[6] http://oud.cle.org.pk/

‘kHElnA’/play runs on the system CUV then it first checks the word in the exception list from line 11 to 18 in the algorithm. If word is found, then output detail is displayed from that except list. Urdu has some verbs which have different inflections than the root form like جا‘jA’/go has گیا‘gyA’/went. So, there is an exception list added into the algorithm to handle such cases. If word is not found in the exception list then CUV checks its inflection and finds out its root word, inflection and category of the verb. Next our system displays morphological detail of that word according to line 19 to 26 in the CUV algorithm. After displaying morphological detail of the word, all other possible 63 classifications with detailed morphological information are evaluated by the CUV algorithm as depicted from line 36 to 53. The interface of our purposed system while executing the Urdu verb کھیلنا ‘kHElnA’/play is presented in Figure 1 along with the output screen in Figure 2 shortly due to large view.



Figure 1: Input Screen of CUV



Figure 2: Short view of output screen

6. RESULTS AND DISCUSSION

The following formula is used for calculating the accuracy of the purposed system as did in related studies [30].

$$Accuracy = \frac{Words_Count_With_Correct_Morphology}{Total_Number_Of_Words_Observed} \; X \; 100$$

Our test data contains 1299 verbs of Urdu. After running all verbs on CUV, 1043 are observed as correct according to morphological information evaluated by the CUV successfully and hence achieved 80.29% of accuracy.

Among the remaining 256 failure verbs, some are depicted below in Table 1, which cover up the issues of almost all the failure verbs. These failures are the output of our proposed CUV on 25 different verbs.

Table 1. Failure Verb List

چھانا	کھانا	لوانا	آونا	بوانا	جانا
نوانا	اپنانا	بھانا	پختانا	فرائ	فراش
نوشتن	نواختن	بداوت	ڈریس	بھگانا	للچانا
چوانا	گنا	گا	لینا	مد	منا
مس					

At present our purposed system is limited to verbs only but in future it will be enhanced up to other classes like noun, adjective, pronoun, adverb, etc. Moreover, the size of the Urdu verb list will also be increased accordingly.

Related work discussed in Section 3 is the case of finite state transducers (FST) technology used to study the Urdu verb morphology. In [16][15] and [14], FST based approaches used lexicon heavily to generate morphology and hence not suited to be compared. Further, they didn’t present quantifiable results. Morphological analyzers can be divided majorly in two branches: linguistic and statistical-based approaches. In a linguistic based approach, strategies applied include affix stripping method or length-based method (related to our work), template-based design, and lexicon-based models. Similarly, in statistical approaches, NLP and ML based probabilistic models are devised. The accuracy achieved till to date in length-based approaches is 79.63% [31-30]. Our approach does not rely on lexicon. Rules are hardcoded in the programming language used and the scope of the work is also narrow with limited test data of 1299 words, which becomes the reason that we are getting 80.29% of accuracy. This percentage of accuracy is better in other designing approaches like statistical.

Inflection and irregular forms are the main issues of failure in detecting the verb morphology. At present, we are successful in identifying 63 out of 66 inflection forms completely (depicted in Figure 2) and the goal is still at distant. CUV has the advantage

of speed and accuracy on limited data in comparison to models discussed in [14-16].

7. CONCLUSION

Only few morphological classifiers are available which focus on encoding rules. The finite state transducer is composed of set of rules and lexicon. Finite state transducer could not handle dynamic inflections too. Our work focuses only on single verb and does not handle complex predicates. Our system is totally rule-based and does not depend upon any lexicon. During design issues like over stemming, most famous irregular verbs, irregular singular masculine perfective form and invalid verbs, were solved first. We tested CUV on verb list, which was designed by CRLUP[7]. Original list is prone to errors. After careful analysis, list is pruned, and more than sufficient results are achieved by testing the updated list on CUV. This hard-coded rule-based approach has the advantage of speed and accuracy on limited data as compared to related models discussed in Section 3. We have further analyzed that this approach could not be successful for large data set until we have intelligent searching mechanism based on some training data along with the use of some statistical model [26-28]. So, a large cushion is available for the verb classification using statistical models.

REFERENCES

[1] Abbas, Q. Ahmed, S.M. Niazi, S., "Language Identifier for Languages of Pakistan Including Arabic and Persian.", International Journal of Computational Linguistics (IJCL), Vol. 01(03), ISSN 2180-1266, P 27-35, 2010.

[2] Abbas, Q., "Morphologically rich Urdu grammar parsing using Earley algorithm, Natural Language Engineering", (NLE), Vol.21(2), PP.1-36, ISSN: 1351-3249, DOI: 10.1017/S1351324915000133, Cambridge University Press, 2015.

[3] Bajwa K. B., Nawaz W., Abbas Q., Almuzaini A., Khan R.. "A. Sentiment Classification and Analysis on Arabic using Semi-Supervised Approach. Journal of Emerging Technologies", and Innovative Research (JETIR), Vol. 5(8), 2018.

[4] Hussain, S., & Afzal, M., "Urdu computing standards: Urdu Zabta Takhti (UZT) 1.01. In Proceedings. IEEE International Multi Topic Conference", IEEE INMIC 2001. Technology for the 21st Century. (pp. 223-228). IEEE, 2001.

[5] Koskenniemi, K. Two-level morphology: "A general computational model for word-form recognition and production.", Ph.D. dissertation, University of Helsinki, 1983.

[6] Porter, M. An algorithm for suffix stripping. . In: Program 14, 130–137, 1980.

[7] Munir, S., ABBAS, Q., Jamil, B., "Dependency Parsing using the URDU.KON-TB Treebank.", International Journal of Computer Applications, Vol. 167(12), pp. 25-31, 2017.

[8] Karttunen, L. K. "A general morphological processor." In: Linguistic Forum 22, 163-186, 1983.

[9] Bosch, A. V. "Memory-based Morphological Analysis. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics,", ACL'99, University of Maryland, USA, 285-292, 1999.

[10] Wicentowski, R. "Multilingual Noise-Robust Supervised Morphological Analysis using the WordFrame Model. "In: Proceedings of Seventh Meeting of the ACL Special Interest Group on Computational Phonology (SIGPHON), 70-77, 2004.

[11] Kurimo, Mikko." An Introduction and Evaluation Report. In: Proceedings of Unsupervised segmentation of words into morphemes -- Challenge 2005." Helsinki University of Technology, Helsinki, Finland, 2005.

[12] Mathias, C. and KrisIn, L. "Unsupervised Morpheme Segmentation and Morphology Induction from Text Corpora Using Morfessor 1.0." Publications in Computer and Information Science, Report A81, Helsinki University of Technology, 2005.

[13] Goldsmith, J., "Unsupervised Learning of the Morphology of a Natural Language. In: Computational Linguistics", Vol 27, 153-189, 2001.

[14] Hussain, S. Finite state morphological analyzer for Urdu. Master thesis CRULP Lahore, 2004.

[15] Humayoun, M., Hammarström, H. and Ranta, A., 2Urdu morphology, orthography and lexicon extraction. Chalmers tekniska högskola", 2006.

[16] Bögel, T., Butt, M., Hautli, A. and Sulger, S., "Developing a finite-state morphological analyzer for Urdu and Hindi. Finite state methods and natural language processing", p.86, 2007.

[17] Zia, T, Abbas, Q, Akhtar, M. P. Evaluation of "Feature Selection Approaches for Urdu Text Categorization, International Journal of Intelligent Systems and Applications", (IJISA), Vol.7, No. 6, PP.33-40, DOI: 10.5815/ijisa.2015.06.03, MECS Publisher, HK, 2015.

[7] It is currently available at http://www.cle.org.pk/software/ling_resources/urduverblist.htm

[18] Zia, T, Akhtar, M. P., Abbas, Q., “Comparative Study of Feature Selection Approaches for Urdu Text Categorization, Malaysian Journal of Computer Science (MJCS)”, Vol.28, No. 2, PP.93-109, University of Malaya, 2015.

[19] Schmidt, R. L. Urdu: An Essential Grammar. London: Rutledge, 1999.

[20] Abbas, Q. Zia, T. and Khan A.N. , “Syntactic and Semantic Analysis of Urdu Modal Verbs using XLE Parser.”, International Journal of Computer Applications, Vol. 107(10), pp. 39-46, US, 2014.

[21] Abbas Q. and Raza G. , “A Computational Classification of Urdu Dynamic Copula Verb. International Journal of Computer Applications”, (IJCA), Vol. 85(10), P 1-12, ISSN 0975 - 8887. Published by Foundation of Computer Science, New York, USA, 2014.

[22] Abbas, Q. Building, “A Hierarchical Annotated Corpus of Urdu: The URDU.KON-TB Treebank.” Lecture Notes in Computer Science (LNCS). Vol. 7181(1), P 66-79, ISSN 0302-9743, Springer-Verlag Berlin/Heidelberg, 2012.

[23] Rizvi, S. M., “Development of algorithms and computational grammar of Urdu. Master Thesis”, 2009.

[24] Abbas, Q. Semi-Semantic Annotation: “A guideline for the URDU. KON-TB treebank POS annotation. Acta Linguistica Asiatica”, 6(2), 97-134, 2016.

[25] Abbas, Q. Hussain, Q. Zia, T. Mansoor, “A. Reduced Solution Set Shortest Path Problem: Capton Algorithm With Special Reference to Dijkstra's Algorithm.”, Malaysian Journal of Computer Science (MJCS), Vol. 31(3), PP 175-187, University of Malaya, 2018.

[26] Abbas, Q., “A Stochastic Prediction Interface for Urdu, International Journal of Intelligent Systems and Applications”, (IJISA), Vol.7, No.1, PP.94-100. DOI: 10.5815/ijisa.2015.01.09, 2015a.

[27] Abbas, Q. Karamat, N. Niazi,S., “Development of Tree-bank based probabilistic grammar for Urdu Language. International Journal of Electrical & Computer Science (IJECS).”, Vol. 09(09), P 231-235, ISSN 2077-1231, 2009.

[28] Abbas, Q., Iqbal, M. M., Niazi, S., Noureen, M., Ahmad, M. S., Nisa, M., Malik, M. K. Mango , “Classification Using Texture & Shape Features.”, International Journal of Computer Science and Network Security, Vol. 18(8), 2018.

[29] Mateen, A. Malik, K. M. Nawaz, Z. Danish, H. M. Siddique, H. M. Abbas, Q., “A Hybrid Stemmer of Punjabi Shahmukhi Script. International Journal of Computer Science and Network Security.”, (IJCSNS), 17(8), 2017.

[30] Jabbar, Abdul, et al. "A comparative review of Urdu stemmers: Approaches and challenges." *Computer Science Review* 34 (2019): 100195.

[31] Jabbar, Abdul, et al. "A survey on Urdu and Urdu like language stemmers and stemming techniques." *Artificial Intelligence Review* 49.3 (2018): 339-373.



Naz et al LGURJCSIT 2021

ISSN: 2521-0122 (Online)
ISSN: 2519-7991(Print)

LGU Research Journal of Computer Science and IT

Vol.5 Issue1, Jan-Mar 2021

Configuration-Free Systems for Wifi Sensing based Smart Home using the Smart Remote Controller

Binish Naz[1] ,Muhammad Rizwan[2], Fahad Ahmad[3], Aysha Shabbir

[1,2,3,4]Department of Computer Science,Kinnaird College for Women,Lahore, Pakistan

mynazmughal@gmail.com

ABSTRACT

The Smart Home invention offers us extreme supervision over our home by automating the lighting structure, the dimming, the screens, electric machines, the sound and safety frames. Main technologies that provide connectivity to smart home facilities, WiFi is one of them. In traditional households, where the household appliances will increase, the remote controls to manage them and the interference between them will also increase. This makes the system configuration dependent and troublesome for users to manage them that increases their burden. Recently, some systems were developed to manage multiple household appliances through one interface. However, the matter is once the dataset will increases the interface gets sophisticated and every home appliance or controller wants a special device to connect with it. In this paper, we introduced FreeGesture in the DeepRemote controller, which is a gesture recognition scheme without a device that uses preferred computer vision algorithms, particularly deep learning, to recognize numerous devices and manage them via IR or network. It simplifies deployment and makes systems without configuration. We are going to consider it as "Smart Remote" in this article. Smart Remote consists of four buttons, a camera, an Inertial Measurement Unit (IMU), a WiFi component, an Infrared (IR) transceiver, and a speaker. The popularity accuracy of smart Remote for 5 varieties of home appliances from completely different places is hyperbolic from 81.07% to 95.8%.

KEYWORDS : *Smart home, WiFi sensing, Deep learning, FreeGesture, Smart Remote.*

1. INTRODUCTION

Smart home appliances that have multiple functionalities made our daily routine advanced and rich. The interconnections of plenteous devices fastened in a very good household appliance with sensors and actuators for mechanization facilities. Current advances have discovered in wireless technology that to capture natural dynamics, WiFi signals are quite delicate, so they can be used to detect persistence. The three levels involved for the smart home integrator that can be a blend of various forms of technologies are a level of application, networking, and perception. The level of perception can be loaded to gather the knowledge of the event from the configuration and functions, as it is the limit for individuals to work with associated objects. The need for a friendly interface to use the interface has driven the progress of widespread and new detecting approaches [1].

There are the main benefits of used WiFi sensing in smart homes.

1) Cost effective. It is very effective by price to prepare sensing tasks on current infrastructures. It is doable by WiFi sensing, typical interior WiFi transceivers.

2) Suitable placement. Place one or many pairs of WiFi transceivers within the place of concentration to build a smart home supported hardware. Unplanned placement isn't cheap by literary works [2-3]. For suitable placement, the software has been developed in the community. There is no need to do extra effort for the development of a device driver. Therefore, we can get an appropriate deployment.

3) Non-invasive sensing. The Energetic or inactive WiFi detection removes the belief in direct contact and uses hidden radios. User knowledge is not mandatory for the achievement of detection responsibilities. Therefore, it is practical to "recognize" and "understand" user behavior [1].

We assume that on different platforms CSI (Channel State Information) can be easily achieved. There ought to be configuration-free applications that manage CSI signal because it is prone to interventions like multipath, media competition, and different magnetic attraction noise [4]. It suggests that systems are regardless of placement and no need for training. The system is not independent of

configuration because of such intrusions. A strict setup is needed for keystroke recognition [5] and respiration monitoring [6]. In this situation, the transmitter and also the receiver is devoted pretty close to accomplishing high SNR, preventive its applied application. Therefore, the mockups that follow training-based approaches aren't promptly acceptable across totally different things, and in new locations, additional training would be required. So, it's challenging and difficult to create configuration free WiFi sensing-based systems.

A basic interface used to manage home appliances is an infrared (IR) remote controller. As in fig. 1, the number of remote controllers increases when we install a new appliance in our home because each home appliance needs one controller to operate him. The user needs to be learned about how to use each controller which actually could be a burden for him. It also increases the management cost and makes the system configuration dependent. We need a system that reduces the user's burden and management cost by operating many home appliances at one interface to simplify the deployment.



Figure 1: Increasing home appliances with many IR remote controllers

CSI (Channel Status Information), which is shown as a matrix that is a dimension record with the flow of completely different carrier data that can be considered different sensors. CSI (Channel Status Information) is a considerably better metric of the RSSI (Received Signal Strength Indicator) and most systems are based on it. This is somewhat like employing a detector array to seizure data with a matrix output, which is extremely the same as computer vision knowledge [7]. Therefore, computer vision favored algorithms, particularly deep learning [8] is also useful. Once WiFi signal meets deep learning, it may bring a lot of chances for WiFi empowered systems, which make simpler the placement and makes the systems independent of configuration. Deep learning (also called deep structured learning or hierarchic learning) is a subfield of a wider field of machine learning approaches supported learning information representations, as hostile task-specific algorithms.

The first solution to the above-mentioned problem is the association of interfaces for monitoring diverse household appliances. Unification of domestic gadgets and IR remote-controllers are two strategies for interface merger.

Following the first strategy [9-12], the domestic machines that are very much coordinated with the management IR and using the controller it can be monitored. The domestic device's IR signals can be recorded and monitored by a controller, which can work utilizing the internet and display. If the controller quits working appropriately then it becomes troublesome to manage the IR signals of many domestic appliances, the regulation of home machines can be done by learning the remote-controller's IR signals.

For the second method, i.e. unification of the home appliances [13-15], connecting the home appliance to the internet or home network, as shown in fig 2, is needed. Such type of home appliances is referred to as information appliances. Smartphones or net applications can be controlled information appliances through the network.

In the unification of remote controllers, one interface is used to regulate multiple household appliances. This approach, however, cannot remove the complication of management hence takes some effort and time to select and pick a target home appliance out of multiple appliances. So, a method or procedure for choosing a target home appliance based on intuitions is needed.

A common approach for picking an objective device is to join particular appliances or markings to machines that encourage to recognize that appliance, to which the remote controller is pointing out. It may use QR code, IR transmitters, IR LED (Infrared Light Emitting diode) as particular markings/gadgets. Using the monitoring device, it is feasible to choose a particular appliance by perceiving the specific marking/appliance. The major benefit of this system is to acknowledge any courses of action of users and domestic appliances. However, it takes the expense to associate a marking/gadget to each appliance.

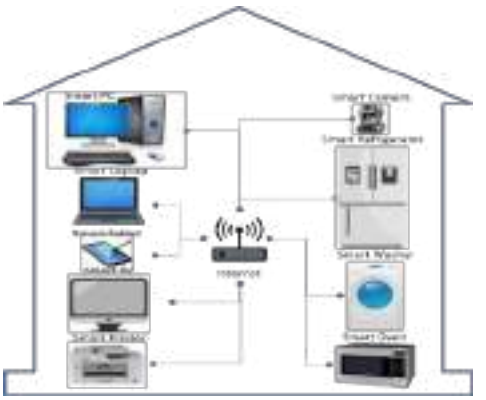

Figure 2: Household appliances connected to the internet or home network

To choose devices for house use, while not combining uncommon gadgets to family unit apparatuses, numerous techniques have been envisioned (utilizing voice, vision, and so on). In case of voice oriented selection, clients, for instance, "on the ground floor, turn on the television " or "Set the cooling inside the parlor to the 26 levels of warming." In case of vision oriented selection, clients will pick family machines with instinctive administration like advise to a particular domestic device. The device utilizing a wearable-glass will comprehend this technique. Nonetheless, it is inconvenient for the framework upheld wearable appliances to utilize IR because of the territory for the management of IR, the wearable device's mounting on the transmitter is precluded. Consequently, we will in general accept that a controller that is hand-held, has high applicability for reasonable use [16].

In the light of the content stated over, the following paper introduces a synopsis of the smart remote controller in WiFi sensing based smart home-related configuration and multiple controls' issues, and the possible methodologies to overcome these issues in the field are proposed. The purpose of our paper is to explain a technique for functioning all of the appliances in a home from one remote controller supported deep learning approaches and FreeGesture recognition scheme to make a system configuration free. The association of this paper is as per the following. Literature review presents in the section 2. Section 3 presents the new proposed solution. Section 4 presents the discussion of results. The paper concludes in section 5.

2. LITERATURE REVIEW

2.1 Unification of remote controllers

Two approaches that are used to affiliate more than two remote controllers to minimum one interface include: using information appliances and engaging a remote controller with varied IR signals of household appliances. At the same time when most household appliances can be controlled by IR signals, systems and frameworks for joining the IR controller have been introduced in the past. A framework for connecting heritage household appliances to the network have introduced by Nakumara *et al.,* [12]. IR remote controllers, that can be used via a network and remotely regulator home appliances, have functions of learning IR signals and connecting to a home network. IRKit [9] and eRemote [10] are two types of IR remote controllers. HUIS remote controllers are built to control every home appliance, including TV, air-conditioner, lights and audio systems with one device remarkably. HUIS [11] is Associate in the Nursing IR remote controller that provides nursing electronic paper. Users can choose a favorite interface without any trouble.

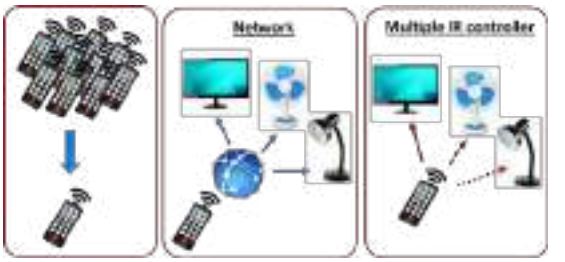

Figure 3: Unification of remote controllers

There are some standardized communication protocols like ECHONET [15], UPnP (Universal Plug and Play) [13] and SEP2.0 that are used to control information appliances. Fast power line communication gets users' consideration widely as an associate for communication of home machines. UbiREMOTE, developed by Kiyokawa *et al.,* [14] is one of the proposed systems for controlling information appliances. This is achieved via interactions in a three-dimensional virtual space. This system uses UPnP standardized protocol for most powerful home appliances. The joining of multiple and functionally varying remote controllers can develop a malfunction and make it difficult to find a target appliance.

2.2 Selection of home appliance with special devices attached

The choice of a target appliance becomes easier if we attach the appliance to a special device. Wii Remote's gestures, a system component that was implemented and checked by Neelrath *et al.,* [17] is used to select a target home appliance. In this system, IR LED is connected to a home appliance and the LED (Light Emitting diode) flashes specific patterns which are further received and analyzed via Wii Remote. Ubi-Finger, developed by Tsukada *et al.,* [18] is a small finger-worn device that makes the selection and control of a target appliance much easier by gestures. An IR transmitter place next to the IR receiver to make it possible. The IR transmitter is attached to the Ubi-Finger' tip while the IR receiver is connected to the home machines.

Ujima *et al.,* [19] advised setting a depth camera and a projector on each home appliances. Thus, the home machine is displayed on the user's body and the signal control it using the user's two arms. Another framework, created by Komeda *et al.,* [20] utilizes a heading of the controller, in light of the greatest precision indoor situating framework, to call attention to the target machine. Remote-touch, developed by Ullah *et al.,* [21] makes selection convenient via reading the QR code. Installation of the device or an indoor situating framework is necessary for the selection of home appliances by these methods.

2.3 Selection of home appliance without attaching special devices

Researchers have found a couple of mechanisms to select electronics without using any dedicated tool. GeeAir which is proposed by Pan *et al.,* [22] selects electronic via speech recognition. It controls that electronic device through IR signals. Then there's Xiao *et al.,* [23] who says we can select the electronic device using electromagnetic noise patterns. Kong *et al.,* [24] on the other hand, make utilization of smart glass and environmental information to recognize the electronic device. Smart glass takes a picture of the device. Environmental information includes WiFi whereabouts, sound or glare. The pictures from a smart device are recognized using neural network algorithms. Yes, the special devices are not necessarily needed to be engaged physically with electronics but still, these methods make the process difficult to understand and it's tough to differentiate appliances.

Hase *et al.,* [25] study heavily makes use of deep learning algorithms. His programs learned from a much larger dataset of images i.e. around 10,000 images per appliance. Also, his system was only controllable with Infrared signals. Here's how Takahashi *et al.,* [16] fill these gaps. They developed DeepRemote and use transfer learning, hence the much smaller number of images were required for the training process. The actual neural network was trained with ImageNet [26]. They also evaluate performance, accuracy and show the response time of DeepRemote in a real environment. Our research is most relevant to Takahashi's study. Here's how our study differs from his study though. We introduce a device-free Gesture recognition [27] scheme in DeepRemote to increase its recognition accuracy more than previous and we will consider it as Smart Remote in this paper. Already conducted experiments have proven that FreeGesture recognition achieves more gesture recognition accuracy.

3. PROPOSED MODEL

3.1. System Overview

Deep learning (identifies appliances by visual features) is used in a Smart Remote controller which controls and organizes multiple home appliances. The developed system can actually function in the absence of any direct device connection to the appliance. Altering distance and direction of the home appliance does not leave a marked effect on control and recognizing power hence this controller is practical, compatible and easily operate able in a normal daily life routine environment.

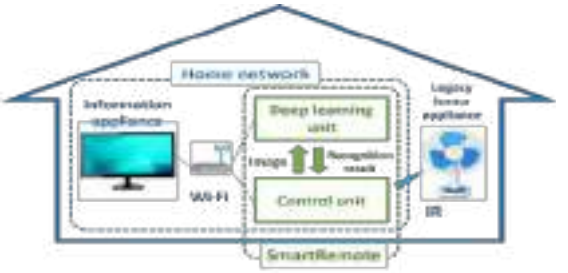

Figure 4: Proposed System model

3.2. The flow of the system

Fig. 4 expressions the model of the system. Two units, a deep learning unit, and the control unit, connected via the home network, make up this system. The former one is used for the identification of home appliances utilizing an image while the later one acts as a controller with an attached camera. IR signals and WiFi are used for controlling home appliances through the Smart Remote. Most of the existing controllers can only control homogenous home appliances, but unlike these controllers, our system can recognize and control diverse electronic.

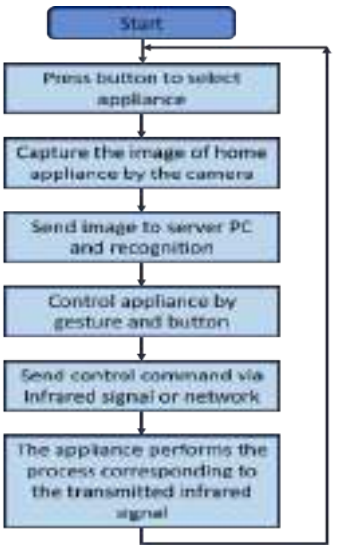

Figure 5: Control flow for the proposed smart remote controller

Fig. 5 demonstrates the control flow of the household appliance utilizing Smart Remote. In the preliminary step, a user direction the regulator to select a target appliance which is followed by pressing recognition key on the control unit. Pressed button captures the image of a home appliance using the camera attached through the control unit. The image is forwarded to the deep learning unit through WiFi connection, which in turn identifies the home appliance and sent the final result (according to the image) back to the Control Unit. A user can hear the audio feedback result. Following recognition, controlling the

appliance by using an IMU sensor (an inertial measurement unit) and 4 buttons (one power button, the other three buttons for controlling appliances) on the control unit becomes possible. To enhance the accuracy of this identification, in the measurement unit, a FreeGesture recognition scheme is used. There are multiple kinds of commands hence we introduce a function that ultimately changes the assigned operations of the button. For instance, when controlling a television, channel, and volume by these buttons if two tasks are allowed. The user can change the button assignment via gestures (e.g. by rotating the control unit, the user can choose a key to control volume or channel) if more than two operations are enabled. The identification/recognition button is always available regardless of appliance identification status. Recognition of the home appliance is important and necessary in order to control it [16].



Fig. 6. Control unit of Smart Remote [16]

3.3. Control Unit

The control unit is shown in fig. 6. Consists of 4 buttons & speaker, an infrared transmitter, a camera, a battery, and an onboard computer (OCT). USB and GPIO are used for communication between the device and the computer. A 3D printer was used for this controller. Since Raspberry-Pi-3, an OCT used a WiFi component. The camera is astride the control unit and acquires a 640 × 480 image. MPU-9150 is used for the IMU sensor. For gesture recognition, it acquires data from the 20 Hz turn sensor. The revolution of the right and the revolution of the left can be recognized as signs. The control unit judges that the user has made a signal, if the threshold value becomes greater than the rotation sensor's middling value (within a time-period).

3.4. Unit for Deep Learning (UDL)

- Environmental Monitoring: Because the deep learning processing load is excellent, the computer required high processing power. In the proposed system, in light of the fact that the control unit isn't so incredible in the figuring, top to bottom the learning is dealt with on a server PC (a UDL). Through the domestic network, it can be associated with the control unit. For communication between the control unit and the UDL, the ROS network can be utilized [28].
- Recognition model: We use VGG16 [29] in the deep learning unit in our system. VGG16 is one of the models of the convolutional neural system (CNN) effectively formed by an expansive arrangement of image data, ImageNet. It comprises 13 convolution layers and 3 totally associated layers. 1000 units of the output layer and input material can be characterized into 1000 modules.
- Fine modification: VGG16 generates the characterization aftereffect of 1000 classes. Leave classes to integrate a fan, a TV and pointless classes like a container. In this way, we changed VGG16 on the model that had some expertise in recognition of home machines with the tuning. The most poweful approach to train Convolutional Neural Network is to Tune-up [30]. The parameters of the predetermined levels can be formed, and the configuration of completely associated levels can be changed using Fine Tuning (keeping the parameters of the rest of levels). The setup has two advantages: lessening the time mandatory for training and reducing the data set for training.

3.5. Algorithm

The working of our proposed framework is clarified through the description of the following algorithm.

STEP 1: On the target device, the user points the controller and (on the control unit), the recognition-switch is pressed.

STEP 2: The image is captured, with the camera prepared with the control unit.

STEP 3: Send the image to the deep learning unit by means of WiFi.

STEP 4: In the image, device recognition is done using the UDL.

STEP 5: The recognition results are sent to the control unit.

STEP 6: Inform the user through audio comments on the result of recognition.

STEP 7: The user controls the recognized device via four power buttons and an IMU (Inertial Measurement Unit).

STEP 8: The device performs the action based on the command of the transmitted infrared signal.

4. RESULTS DISCUSSION

We define the performance evaluation of the smart remote control performed by [16] in this segment. First, they describe the training data and test environment, then examine the precision of this model and introduce user learning.

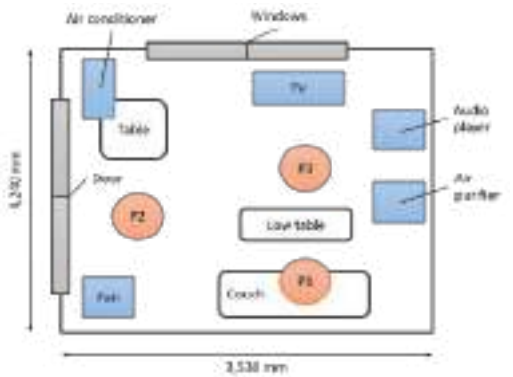

Figure 7: A design of the experimental environment

Fig. 7 demonstrates the design of the experimental environment. From three different positions P1, P2 and P3, they captured 20 photos for each device as appeared in Fig. 7. Overall, three hundred photos were composed for training.

Using fine-tuning with 300 pictures of 5 classes, they trained a model. They build up the decline of the stochastic slope (SDG) as the streamlining agent and the clear-cut cross-entropy as an element of the misfortune and occasions at 200. It is affirmed that the precision increments and the misfortune decline as the number of cycles build (times). Accuracy got to about 1.0 in the around 100th period in this training. Along these lines, it is inferred that the training was effective. The training time was around 90 minutes using the Core-i5 processor, on the computer.

Precision-rate in each location is very different, with an accurate precision-rate (90.45 percent) was P2. In all places, the average precision-rate is 81.07 percent.

When the acknowledgment button is pushed on the control-unit, the response time is assessed, until the moment that the detection outcomes are returned. The highest time was 3.07 sec, the base time was 1.72 sec, and the general response time was 1.95 sec. Delay in communication was the principal cause behind the deferral. It is feasible to see the image, without the GPU, and using the up to down learning [16].

From the above experiment, we can say that if we introduce the FreeGesture recognition scheme in its measurement unit, then the recognition accuracy could be increased from 81.07% to 95.08% (almost 15% increase).

5. CONCLUSION

We introduced a device-free gesture recognition scheme called “FreeGesture” in the DeepRemote controller and we consider it as a smart remote control in our paper that uses algorithms in favor of artificial vision, particularly deep learning, which is used to recognize different devices and operate them through IR or network. We use in-depth learning to simplify deployment and make the system without configuration for smart home based on WiFi detection. We use Smart Remote to intuitively control equipment through image recognition through in-depth learning. Our framework includes a deep learning unit, which explains how to perceive a home device that users must use and a control unit that can control household machines with switches (4 buttons) and signs. Our framework perceives a device with an accuracy of 15% higher than the past arrangement from 81.07% to 95.08% when we present the new FreeGesture recognition scheme with smart remote control. Utilize the pre-prepared VGG16 arrange by adjusting just 300 pictures. At the point when the framework is utilized with this exactness, the user should endeavor to accurately perceive 1.23 times. Likewise, with a recognition time of 1.97 seconds, our framework can work without a GPU and switch the device on for under 5 seconds when it is correctly recognized.

REFERENCES

[1] Jiang, H., et al., "Smart home based on WiFi sensing: a survey." IEEE Access,. **6**: p. 13317-13325, 2018.

[2] Virmani, A. and M. Shahzad., "Position and orientation agnostic gesture recognition using wifi. in Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services.", 2017.

[3] Wang, W., et al., "Understanding and modeling of wifi signal based human activity recognition. in Proceedings of the 21st annual international conference on mobile computing and networking.", 2015.

[4] Abdelnasser, H., M. Youssef, and K.A. Harras. Wigest:, "A ubiquitous wifi-based gesture recognition system. in Computer Communications.", (INFOCOM), 2015 IEEE Conference on. 2015.

[5] Ali, K., et al., "Keystroke recognition using wifi signals. in Proceedings of the 21st Annual International Conference on Mobile Computing and Networking.", 2015.

[6] Wang, H., et al., "Human respiration detection with commodity wifi devices: do user location and body orientation matter?", in Proceedings

of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016.

[7] Rajagopal, N., P. Lazik, and A. Rowe., "Visual light landmarks for mobile devices. in Information Processing in Sensor Networks", IPSN-14 Proceedings of the 13th International Symposium on. 2014.

[8] Wang, T., et al., "Deep learning for wireless physical layer: Opportunities and challenges. China Communications", **14**(11): p. 92-111, 2017.

[9] “Irkit - open source wifi connected infrared remote controller.” [Online]. Available: http://getirkit.com/en/.

[10] “eremote.”[Online].Available:http://linkjapan.co.jp/product/eremote/

[11] “Huis remote controller.” [Online]. Available: http://huis.jp/

[12] Nakamura, M., et al. Adapting legacy home appliances to home network systems usingweb services. in Web Services, 2006. ICWS'06. International Conference on. 2006.

[13] Kim, D.-S., et al., Design and implementation of home network systems using UPnP middleware for networked appliances. IEEE Transactions on Consumer Electronics, **48**(4): p. 963-972, 2002.

[14] Kiyokawa, K., et al., UbiREMOTE: framework for remotely controlling networked appliances through interaction with 3D virtual space. in Proceedings of the first annual ACM SIGMM conference on Multimedia systems. 2010.

[15] Saito, S., N. Ishikawa, and Y. Tsuchiya. Development of ECHONET Lite-Compliant Home Appliances Control System Using PUCC Protocols from Smart Devices. in Computer Software and Applications Conference (COMPSAC), 2015 IEEE 39th Annual. 2015.

[16] Takahashi, Y., et al. DeepRemote: A smart remote controller for intuitive control through home appliances recognition by deep learning. in 2017 Tenth International Conference on Mobile Computing and Ubiquitous Network (ICMU). 2017.

[17] Neßelrath, R., et al., A gesture based system for context–sensitive interaction with smart homes, in Ambient Assisted Living., Springer. p. 209-219,2011.

[18] Tsukada, K. and M. Yasumura, Ubi-finger: A simple gesture input device for mobile and ubiquitous environment. Journal of Asian Information, Science and Life (AISL),. **2**(2): p. 111-120, 2004.

[19] Ujima, K., A. Kadomura, and I. Siio. U-Remo: projection-assisted gesture control for home electronics. in CHI'14 Extended Abstracts on Human Factors in Computing Systems. 2014.

[20] Komeda, J., et al. Intuitive appliance control system based on a high-accuracy indoor positioning system. in 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence & Computing and 2014 IEEE 11th Intl Conf on Autonomic & Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom). 2014.

[21] Ullah, A.M., et al. Remote-touch: Augmented reality based marker tracking for smart home control. in Computer and Information Technology (ICCIT), 2012 15th International Conference on. 2012.

[22] Pan, G., et al., GeeAir: a universal multimodal remote control device for home appliances. Personal and Ubiquitous Computing, **14**(8): p. 723-735, 2010.

[23] R. Xiao, G. Laput, Y. Zhang, and C. Harrison, “Deus em machina: Ontouch contextual functionality for smart iot appliances,” in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ser. CHI ’17. New York, NY, USA: ACM, , pp. 4000–4008,2017.

[24] Kong, Q., et al. Selecting home appliances w,ith smart glass based on contextual information. in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016. ACM.

[25] Hase, T. and T. Sakao. Integrated remote controller distinguishing home appliances by deep learning. in Consumer Electronics, 2016 IEEE 5th Global Conference on. 2016.

[26] Deng, J., et al. Imagenet: A large-scale hierarchical image database. in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. 2009.

[27] Zou, H., et al. WiFi-enabled Device-free Gesture Recognition for Smart Home Automation. in 2018 IEEE 14th International Conference on Control and Automation (ICCA). 2018.

[28] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs,

R.Wheeler,andA.Y.Ng,“Ros:anopen-sourcerobotoperatingsystem,” in ICRA workshop on open source software, vol. 3, no. 3.2. Kobe, p. 5,2009.

[29] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” CoRR, vol. abs/1409.1556, 2014.

[30] P. Agrawal, R. B. Girshick, and J. Malik, “Analyzing the performance of multilayer neural networks for object recognition,” CoRR, vol. abs/1407.1610, 2014.

Editorial Policy and Guidelines for Authors

LGURJCSIT is an open access, peer reviewed quarterly Journal published by LGU Faculty of Computer Science. The Journal publishes original research articles and high quality review papers covering all aspects of computer science and information technology.

The following note set out some general editorial principles. A more detailed style document can be downloaded at http://lgurjcsit.lgu.edu.pk

All queries regarding publications should be addressed to editor at email lgurjcsit@lgu.edu.pk . The document must be in word format, other format like pdf or any other shall not be accepted.

The format of paper should be as follows:

- Title of the study (center aligned, font size 14)
- Full name of author(s) (center aligned, font size 10)
- Name of Department
- Name of Institution
- Corresponding author email address.
- Abstract
- Keywords
- Introduction
- Literature Review
- Methodology
- Data analysis/Implementation/Simulation
- Results/ Discussion and Conclusion
- References.

Heading and sub-heading should be differentiated by numbering sequences like, **1. HEADING** (Bold, Capitals) 1.1 ***Subheading*** (Italic, bold) etc. The article must be typed in Times New Roman with 10 font size 1.5 space, and should have margin 1 inches on the left and right. Length of paper should not be longer than 15 pages, including figures, tables, exhibits and bibliography. Table must have standard caption at the top while figures below with. Figure and table should be in continuous numbering. Citation must be in according to the IEEE 2006 style.

LAHORE GARRISON UNIVERSITY

Lahore Garrison University has been established to achieve the goal of excellence and quality education in minimum possible time. Lahore Garrison University is situated in the Punjab metropolis city of Lahore is an important milestone in the history of higher education in Pakistan. In order to meet global challenges, it is necessary to touch the highest literacy rate while producing skillful and productive graduates in all fields of knowledge.

VISION

Our vision is to prepare a generation that can take the lead and put this nation on the path to progress and prosperity through applying their knowledge, skills and dedication. We are committed to help individuals and organizations in discovering their God-gifted potentials to achieve ultimate success actualizing the highest standards of efficiency, effectiveness, excellence, equity, trusteeship and sustainable development of global human society.

MISSION

At present, LGU is running Undergraduate, Graduate, Masters, M.Phil. and Ph.D. programs in various disciplines. Our mission is to serve the society by equipping the upcoming generations with valuable knowledge and latest professional skills through education and research. We also aim to evolve new realities and foresight by unfolding new possibilities. We intend to promote the ethical, cultural and human values in our participants to make them educated and civilized members of society.

Contact: For all inquiries, regarding research articles submission and correspondence, kindly contact at this address:
Phone: +92- 042-37181823
Email: lgurjcsit@lgu.edu.pk
Copyright @ 2021, Lahore Garrison University, Lahore, Pakistan. All rights reserved.

Published by: Faculty of Computer Science, Lahore Garrison University Lahore, Pakistan.