

Framework for Cloud Based Document Management System with Institutional Schema of Database

¹Ahmad Y. A. Bani Ahmad, ²D. Karuna Kumari, ³Aasheesh Shukla, ⁴A. Deepak, ⁵Mr. Maulik Chandnani, ⁶Dr. Sumit Pundir, ⁷Dr. Anurag Shrivastava

Submitted: 20/08/2023

Revised: 10/10/2023

Accepted: 22/10/2023

Abstract: The majority of applications will probably continue to use relational database management systems as their preferred technology. RDBMS providers have created cutting-edge strategies to quicken the installation procedure. For good reason, relational databases have long served as the foundation of databases for a number of products. Using relational databases has a number of benefits. The market share maintained by relational database management systems is under threat from NoSQL database technologies. There are many distinct types of NoSQL databases, and they were all created specifically to do certain tasks that RDBMSs were not designed to carry out. The many database structure options accessible to them have given a renewed sense of independence to many different kinds of organizations. Despite their recent successes and their capacity to manage enormous volumes of data, NoSQL databases are not the best option for all plans, and they are not likely to unseat relational software from its position as the industry standard database any time soon. Despite their recent achievements and proficiency in handling enormous volumes of data. Depending on the goals your company has set for itself, conventional databases, NoSQL databases, or a combination of the two may be able to give the most information to your organization. Cloud database services may leverage cloud computing to achieve the objectives of ideal scalability, high accessibility, multi-tenancy, and real resource supply. But organizing these databases in a cloud computing environment is difficult since they need ACID (Atomicity, Consistency, Isolation, and Durability) exclusivity. The goal of this study is to look at the state-of-the-art regarding the need for cloud databases and the various frameworks available. In order to approach the services that are described by the cloud computing system, it defines the typical operations that are conducted on a cloud database by a cloud computing proposal.

Keywords: Cloud Database, Database framework, Document Management, Cloud computing systems

1. Introduction

A document management system (DMS) that is hosted in the cloud and combined with an institutional database constitutes a contemporary approach to the management and organizing of documents inside an organization. This system provides a centralized platform for the storage, access, and collaboration of papers, while the institutional database acts as a repository for the management of essential data. In the past, document management systems required

significant monetary expenditures for the on-premises infrastructure, hardware, and software, as well as for the ongoing maintenance of the system. However, as a result of the development of cloud computing, companies are now able to construct and deploy their DMSs on cloud platforms that are scalable and flexible to their specific needs. This indicates that businesses may be able to use the scalability and reliability offered by cloud providers as an alternative to the expensive infrastructure management that is needed [1].

When making use of a document management system (DMS) that is hosted in the cloud, documents are stored on cloud storage services such as Amazon Simple Storage Service (S3), Microsoft Azure Blob Storage, or Google Cloud Platform (GCP) Cloud Storage. These services not only provide secure document storage for an extended period of time but also facilitate the management of a significant number of documents. Document metadata such as the title, author, tags, and other essential data are stored in the institutional database, which may be constructed on a relational or NoSQL database management system. This kind of system may also be used. It's possible that the data are saved on this system. Utilizing a DMS that is housed in the cloud offers a number of benefits, the most notable of which is accessibility. Users are able to see their documents in a secure manner from any place and on any device that is connected to the internet. The capability of having several

¹Department of Accounting and Finance Science, Faculty of Business, Middle East University, Amman 11831, Jordan, 8788
aahmad@meu.edu.jo

²Assistant Professor, Department of Aeronautical, Institute of Aeronautical Engineering College, Dundigal
d.karunakumari@iare.ac.in

³Computer Engineering and Applications, GLA University, Mathura
Anuragshri76@gmail.com

⁴Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu
*deepakarun@saveetha.com

⁵Assistant Professor, Faculty of Commerce and Management, RNB Global University
Bikaner, Rajasthan
maulik.chandnani@rnbglobal.edu.in

⁶Associate Professor, Department of Computer Science & Engineering, Graphic Era Deemed to be University, Dehradun
sumitpundir1983@gmail.com

⁷Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,
Chennai, Tamilnadu
Anuragshri76@gmail.com

users work together on the same page at the same time and monitor changes using version control makes collaboration easier and increases overall productivity [2].

Security is an essential feature that must be included into a cloud-based document management system (DMS). Encryption technologies, access limits, and user authentication systems all work together to create a barrier between permitted users and documents in order to avoid data breaches and protect sensitive information. This barrier prevents unauthorized users from accessing documents. It is essential that consideration be given to ensuring full compliance with all applicable regulations, such as the “Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR)”. Search functionality is essential for any cloud-hosted document management system (DMS). Adopting technology such as Elasticsearch, which enables users to instantly search for documents based on keywords, tags, or other specific criteria, may make information retrieval more efficient and save users time. A cloud-hosted document management system (DMS) may offer businesses with a document management solution that is scalable, secure, and user-friendly. This solution can then be used in combination with an institutional database. By using cloud infrastructure and services, organizations have the opportunity to enhance their overall process efficiency, as well as their communication and document processes. [3].

It is recommended, in accordance with the "least privilege" approach, that the level of access that users have be restricted, while at the same time ensuring that the programme can function normally. For instance, application owners and administrators should only be granted access to information, programmes, and systems when it is absolutely necessary for them to carry out their responsibilities. When using this strategy, conduct is more consistent and easy to anticipate. Unauthorised users, for example, are unable to destroy privileged files on purpose or by mistake, and they also cannot halt essential activities. Because fewer rights and responsibilities need to be defined, the time it ordinarily takes to deploy programmes may be cut down significantly when the principle of least privileges is used. The principle of least privilege, on the other hand, is one of the most challenging notions to put into effect [4]. In order for companies to successfully execute the concept of least privilege, they are required to do the following:

- ❖ Data organization
- ❖ A familiarity with the placement of their private information
- ❖ Automation of the management of user access life cycles that is computerized and effective Although these concepts are applicable to the deployment of generic apps, the idea of least privilege is much more important and relevant in cloud computing since tenants will need a certain level of security isolation.

This is because renters will be responsible for their own data.

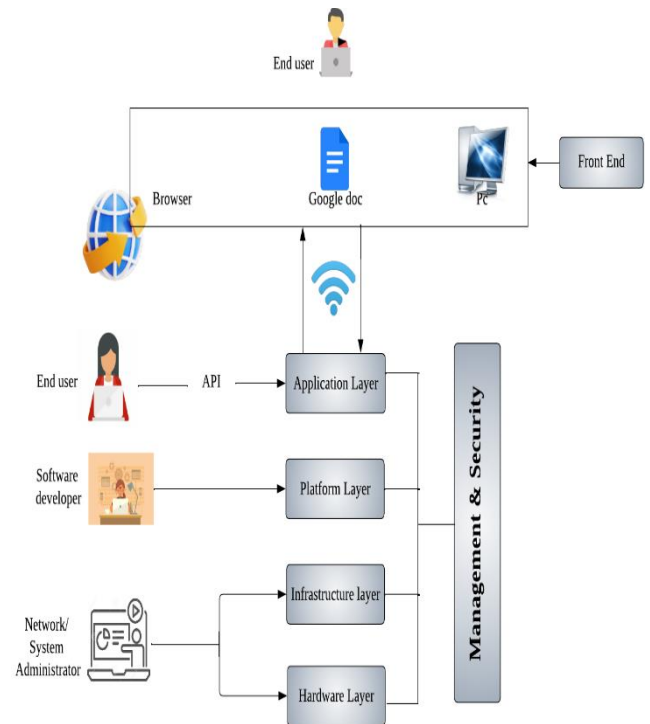


Fig.1. Graphical Display of Cloud Computing Infrastructure

The frontend and the backend, which are the two main parts of a cloud environment, may be separated into their individual parts. The frontend refers to the presentation layer that the user can see. This phrase is often used synonymously with "user interface." The data center hardware, servers, data storage systems, virtualization software, routers, bridges, load balancers, adapters, applications, and services are just a few examples of the many distinctive elements that make up the backend infrastructure that powers the cloud. There are many components that make up the backend infrastructure, and these are only a handful of them [5].

2. Literature Review

Due to the fact that educational institutions produce a large amount of information as a result of their administrative, teaching, and research activities, there seems to be a must to organize and manage this information properly. Additionally, this information has to be able to flow freely across departments. Because we don't want overlapping information, this process needs to take place while taking into account the various internal systems that are already in place. Even if not all organizations have the financial means to invest a significant amount in several software packages, open-source software is an option that is ideal for both educational and administrative tasks. Our objective is to show the benefits of enterprise content management (ECM) in the context of academic organizations, and our title for

this field of research is "content management." In the first part, several methods for enhancing process management are discussed, including electronic content management (ECM) and open-source software. In the second part, we take a look at the various components of the ECM system and discuss their potential applications in various educational contexts. In order to connect ECM to other systems that are active in educational institutions, it is recommended to implement a system architecture that is based on ECM. The case study examined [6] is a discussion of the implementation of an open-source enterprise content management system in a large organization. It demonstrates the system's usefulness and focuses on the qualities that are associated with document processing.

Relational database management systems will likely continue to be the technology of choice for the vast majority of applications. RDBMS suppliers have developed innovative approaches to accelerate the installation process. Relational databases have a long history of acting as the basis of databases for a wide variety of goods, and for good reason. There are several advantages to using relational databases. NoSQL database technologies are posing a threat to the market share held by relational database management systems. There are many different NoSQL database options, and they were all developed for the express purpose of doing particular operations that RDBMSs were not intended to handle. Many different types of organizations are experiencing a newfound feeling of freedom as a result of the abundance of database structure alternatives. Despite their recent triumphs and their ability to handle large amounts of data, NoSQL databases aren't the greatest fit for all plans, and they aren't likely to oust relational software from its position as the dominant database any time soon. Conventional databases, NoSQL databases, or a mix of the two may be able to deliver the most amount of information for your organization, but this will depend on the objectives that your firm has set for itself. Users are able to access databases on demand over the Internet from the server of a cloud database provider, while clients are able to administrate databases that are hosted in the cloud. Cloud computing may be used by cloud database services in order to achieve the goals of optimum scalability, high accessibility, multi-tenancy, and real resource supply. On the other hand, due to the fact that they need ACID exclusivity [7], which is necessary in order to approximate the services that are offered by cloud computing systems.

Because of the trend towards paperless offices and the rising prevalence of electronic information transfer, such as via emails and other web-based content, businesses now need a system that is capable of effectively managing the paper documents that they produce. A document management system that is hosted in the cloud provides users with an easy-to-use method for classifying and identifying documents, which also tags the documents with relevant

information. In today's commercial world, electronic documents are often considered to be the most valuable information assets. An end user, for instance, may use the programme to find out whether an unauthorised user "touched" his or her file in order to ensure that the file is secure [8].

3. Document Management System

Document management systems are built from the ground up with the purpose of assisting whole enterprises in better managing the creation, archiving, retrieval, and expiry of information that is included in documents. The processing of paperwork was a primary motivation for the establishment of these systems. A document management system, often known as a DMS, is one that is based on a central repository and that is designed to manage the storage of any sort of information that may be beneficial to a company and to protect that information from being lost [9]. It is not the same as the file structure on your own computer in any way. Due to the fact that the content that is stored inside a document management system (DMS) is often self-contained, it is much easier to locate information and to disseminate it when the DMS has been created correctly. This is accomplished via the use of sophisticated search engines as well as the incorporation of taxonomies or other techniques of classification within the body of the documents that have been saved [10]. The following is a list of the benefits that document management systems provide:

Reduced Storage: Due to the high cost of commercial real estate and the need to keep records for reasons like retrieval and regulatory compliance, paper-based document storage competes with workers for space inside an organisation. This competition occurs because paper-based document storage takes up more room than people do. This is because paper-based document storage relies on actual paper documents rather than electronic ones. Scanning documents and incorporating the digital copies into a document management system may make it possible to drastically cut down on the amount of physical storage space required for paper documents.

Flexible Retrieval: It takes a significant amount of time to retrieve documents that have been stored on microfilm or in hard copy form. With the assistance of a document management system (DMS), additional electronic copies of documents may be made and stored in a centralised location that is well-organized. Because it is feasible to acquire the documents without having to stand up from a workstation, the process of finding the papers takes less time. Users of the Document Management System (DMS) have the ability to acquire documents even while using other desktop-based apps at the same time.

Flexible Indexing: Paper and microfilm may be indexed using a variety of different approaches; however, each one of these approaches is arduous, costly, and time-consuming. There are several different ways to simultaneously index the photographs of documents that are stored in a document management system (DMS).

Improved, faster and more flexible search: Full text search is a function of document management systems that is not accessible with paper. This feature enables users to obtain files by searching for any term or phrase that may be found inside the document. Full text search refers to the capability of locating and retrieving files based on any word or phrase that may be discovered inside the document. A document management system, often known as a DMS, may further apply one or more taxonomies or categorizations to a document or folder in order to assist with the classification of documents.

No Lost Files: It may be expensive and time-consuming to replace documents that have been lost. Even when papers are viewed inside a document management system (DMS), the fact that they are first photographed and then stored in a single location assures that none of them will be misplaced or lost. Full-text searching capabilities make it possible to rapidly recognise and reorganise newly created documents, even in the event that they are inadvertently stored. This reduces the likelihood that new documents will be saved by mistake.

Digital Archiving: The storing of electronic documents in a format that is both native and non-proprietary, such as Microsoft Word or Excel, and the safeguarding of these documents is known as document preservation and document protection. The process of safeguarding paper documents that must still be kept by making archival copies of such documents and storing them in a document management system so that the originals may be safely discarded.

4. Research Methodology

File encryption for text: Text files, such as those that include health reports, are capable of being encrypted with the use of the advanced encryption standard (AES), which has a block size of 128 bits. In 2001, the National Institute of Standards and Technology (NIST) was the organization that first proposed using the American Encryption Standard (AES) as a method for encrypting digital data. In order to get the encrypted data, the data and the keys are first arranged in an array, after which they are subjected to a series of rotations. This process begins with the data being converted into bytes. The AES encryption standard offers the highest level of protection when it is used to the protection of textual information.

Procedure of watermarking and encryption for medical images: In order to protect the data by applying the security strategy, a watermark containing a one-of-a-kind identity is initially affixed to the digital photograph. This entails two separate actions. As part of the least significant bit (LSB) watermarking process, the patient's name, age, and gender are all watermarked in the image's low bit. LSB stands for least significant bit. Through these measures, the patient's right to privacy is protected. The patient ID that was found in the database is then used to replace the tag that had previously been linked to it. The technique of encryption that is explained in the following paragraphs is the one that we employ. After the encryption process has been completed, the DICOM data of the patient will be retrieved, and the tag that corresponds to those data will be appended. It is important that you be aware that there are a total of 16 bits that make up the watermark, and those bits are represented by the first 16 bits.

Using the strategy that has been given, it is possible to successfully encrypt blocks that have a size of 16 px2 apiece. It is possible that we will be able to retrieve the ith encrypted block if we export the block that has to be encrypted with the block BK. The components of the block BK are denoted by the letters K and S, respectively. The other component of the block is the sum of V, and the vector S serves as the vector carrier for this component. The “vector sum of K combined with the diagonal of the preview encrypted block, $D[B_{i-1}^e]$, forms the letter V”.

The blocks that are encrypted, the target block, the key (which is represented as a “vector column), the function matrix transporter, and the function matrix diagonal are denoted by the notations B^e , B, K, T, and D, respectively”. In the first block, it is important to understand that the value 0 corresponds to the Bi 1 e symbol.

Modelling and cryptanalysis: The “peak signal-to-noise ratio” (PSNR) is what is used in order to do the calculation that determines the amount of distortion that exists between an image I and its watermarked imaging $I_{w dec}$.

Table 1. The suggested image encryption method's cryptanalysis

Images	2017	2018	2019	2020	2021	2022
Original entropy in imaging	1.1708	1.2343	1.5438	1.9877	2.0976	1.0987
Entropy of encrypted imagery	7.9856	7.9934	7.9987	7.8987	7.8765	7.6556

Entropy of images that were incorrectly decrypted	7.8659	7.3421	7.8654	7.8654	7.9865	7.8965
Correlation between the unencrypted image and the original image	0.0054	0.9877	0.0013	0.0087	0.9787	0.0067
NPCR	99.87	99.99	99.86	99.89	99.87	98.76
UACI	45.67	78.96	33.56	33.45	45.98	33.87

When the encryption process is finished, the values of entropy range from 7.9987 to 7.9856, as shown in Table 1, which conducts a more in-depth analysis of the suggested encryption method's protection against attack. This demonstrates that the encrypted photos are really close to having the optimal number of eight and to having originated from a random source. Additionally, this reveals that the images have been encrypted. As a consequence of this, the approach that was described is resilient against an attack based on entropy. Every single photograph in our image test sets reveals that there is very little to no link between the images that have not had their encryption removed and the original photographs. As a consequence of this, the encrypted photos are noticeably distinct from the unencrypted images to which they correspond, which demonstrates that the method of encryption that was described is successful. Our NPCR and UACI have been computed to be 99.96 and 33.39 percent, respectively, based on the data we have. This has the immediate impact of getting the observed values for NPCR and UACI extremely close to the required values, which are 100 and 33, respectively. This is an important step towards achieving the desired results. This demonstrates how the method described here may defend itself against a range of attacks. For imaging depths that are coded on 8 bits, we have a space key of 2128, and for imaging depths that are coded on 16 bits, we have a space key of 2256. The key that we recommend is determined by the number of bits that are required to encode the picture depth. It is of the utmost importance that the critical area be of sufficient size to make it difficult to launch a direct physical attack.

5. Result and Discussions

“Logistic regression analysis” was performed to examine the influence of 2017, 2018, 2019, 2020, 2021 and 2022 on variable Images to predict the value "original imaging entropy". Logistic regression analysis shows that the model as a whole is not significant ($\chi^2(6) = 5.41, p .493, n = 6$).

The coefficient of the variable 2017 is $b = 19$, which is positive. This means that an increase in 2017 is associated with an increase in the probability that the dependent variable is "original imaging entropy". The odds Ratio of 178850903.08 indicates that “one unit increase of the variable 2017 will increase the odds that the dependent variable is original imaging entropy” by 178850903.08 times.

The coefficient of the variable 2018 is $b = -5.98$, which is negative. This means that “an increase in 2018 is associated with a decrease in the probability that the dependent variable is original imaging entropy”. The odds Ratio of 0 indicates that “one unit increase of the variable 2018 will increase the odds that the dependent variable is original imaging entropy” by 0 times.

The coefficient of the variable 2019 is $b = -56.16$, which is negative. This means that “an increase in 2019 is associated with a decrease in the probability that the dependent variable is original imaging entropy”. The odds Ratio of 0 indicates that “one unit increase of the variable 2019 will increase the odds that the dependent variable is original imaging entropy” by 0 times.

Table 2. Hypothesis

Null hypothesis	Alternative hypothesis
There is “no significant difference between the groups of the first factor 2017, 2018, 2019, 2020, 2021 and 2022” (measurement repetition) in relation to the dependent variable.	There is “a significant difference between the groups of the first factor 2017, 2018, 2019, 2020, 2021 and 2022” (measurement repetition) in relation to the dependent variable.
There is “no significant difference between the groups of the second factor Images” in relation to the dependent variable.	There is “a significant difference between the groups of the second factor Images” in relation to the dependent variable.
There is “no interaction effect between the factor 2017, 2018, 2019, 2020, 2021 and 2022” and Images	There is “a interaction effect between the factor 2017, 2018, 2019, 2020, 2021 and 2022” and Images

6. Conclusion

Because of the generation of enormous volumes of data by web-based applications, each and every component of database management systems has undergone transformation. It would seem that using cloud databases would be the most effective method for processing this kind of data. In addition, a company simply does not have the financial resources to construct an expensive infrastructure for a data centre in order to manage and monitor its own databases. As cloud databases get more and more customer attention, a new chapter in the history of database management systems is about to begin. Cloud databases can manage significant workloads from web-based databases despite the fact that they lack the ACID flexibility offered by traditional databases. These kinds of guarantees are not necessary for databases in this form. It's possible for companies to make use of the many databases that are hosted on the cloud. Despite this, every single one of them has their very unique data model, application programming interface (API), database capabilities, and query interface. This procedure has to be standardised in order to facilitate simpler future growth. Cloud computing and cloud databases are now hampered by a number of limitations; if these can be overcome, they will be in a position to dominate the global market over the next 10 years. A number of nations in Asia are providing financial support to cloud-based economic growth centres and are also using cloud services for their own companies.

References

- [1] Rajkumar Buyya et al., "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", *Future Generation Computer Systems*, Vol. 25, Issue 6, June 2009, pp. 599-616
- [2] Jiyi Wu et al, "Recent Advances in Cloud Storage", in Third International Symposium on *Computer Science and Computational Technology (ISCSCT '10)*, Jiaozuo, P. R. China, 14-15, August 2010, pp. 151-154.
- [3] Tyrvainen, P., Päiväranta, T., Salminen, A., & Iivari, J. (2006). Characterizing the evolving research on enterprise content management. *European Journal of Information Systems*, 15(6), 627-634. <http://dx.doi.org/10.1057/palgrave.ejis.3000648>
- [4] Van Rooij, S. W. (2007). Open source software in US higher education: reality or illusion? *Education and Information Technologies*, 12(4), 191-209. <http://dx.doi.org/10.1007/s10639-007-9044-6>
- [5] Vom Brocke, J., Mathiassen, L., & Rosemann, M. (2014). *Business process management*. *Wirtschaftsinformatik*, 56(4), 207-208. <http://dx.doi.org/10.1007/s11576-014-0420-8>.
- [6] B. R. Kandukuri, V, R. Pasture and A. Rakshit, "Cloud Security Issues," in proceedings of the 2009 *IEEE International Conference on Services Computing*, pp. 517- 520, September 2009.
- [7] A. Weiss, "Computing in the Clouds", *networker*, vol. 11, no. 4, pp. 16-25, December.
- [8] "Cloud Technology in the Security Management of Enterprise Document", *IEEE International Conference*, by Na Liu, vol. 30, pp. 344-360, 2011. Neha Sharma, P. William, Kushagra Kulshreshtha, Gunjan Sharma, Bhadrappa Haralayya, Yogesh Chauhan, Anurag Shrivastava, "Human Resource Management Model with ICT Architecture: Solution of Management & Understanding of Psychology of Human Resources and Corporate Social Responsibility", *JRTDD*, vol. 6, no. 9s(2), pp. 219-230, Aug. 2023.
- [9] William, P., Shrivastava, A., Chauhan, P.S., Raja, M., Ojha, S.B., Kumar, K. (2023). Natural Language Processing Implementation for Sentiment Analysis on Tweets. In: Marriwala, N., Tripathi, C., Jain, S., Kumar, D. (eds) *Mobile Radio Communications and 5G Networks*. Lecture Notes in Networks and Systems, vol 588. Springer, Singapore. https://doi.org/10.1007/978-981-19-7982-8_26
- [10] K. Maheswari, P. William, Gunjan Sharma, Firas Tayseer Mohammad Ayasrah, Ahmad Y. A. Bani Ahmad, Gowtham Ramkumar, Anurag Shrivastava, "Enterprise Human Resource Management Model by Artificial Intelligence to Get Befitted in Psychology of Consumers Towards Digital Technology", *JRTDD*, vol. 6, no. 10s(2), pp. 209-220, Sep. 2023.
- [11] Kumar, A., More, C., Shinde, N. K., Muralidhar, N. V., Shrivastava, A., Reddy, C. V. K., & William, P. (2023). Distributed Electromagnetic Radiation Based Sree Lakshmi, P., Deepak, A., Muthuvel, S.K., Amarnatha Sarma, C Design and Analysis of Stepped Impedance Feed Elliptical PatchAntenna Smart Innovation, Systems and Technologies, 2023, 334, pp. 63
- [12] Gupta, A., Mazumdar, B.D., Mishra, M., ...Srivastava, S., Deepak, A., Role of cloud computing in management and education, *Materials Today: Renewable Energy Assessment Using Novel Ensembling Approach. Journal of Nano-and Electronic Physics*, 15(4).
- [13] William, P., Shrivastava, A., Shunmuga Karpagam, N., Mohanaprakash, T.A., Tongkachok, K., Kumar, K. (2023). Crime Analysis Using Computer Vision Approach with Machine Learning. In: Marriwala, N., Tripathi, C., Jain, S., Kumar, D. (eds) *Mobile Radio Communications and 5G Networks*. Lecture Notes in Networks and Systems, vol 588. Springer, Singapore. https://doi.org/10.1007/978-981-19-7982-8_25
- [14] Rushby, N. (2013). Special issue on e-leadership: editorial. *British Journal of Educational Technology*, 44(6).

- [15] Scheepers, R. (2006). A conceptual framework for the implementation of enterprise information portals in large organizations. *European Journal of Information Systems*, 15(6), 635–647. <http://dx.doi.org/10.1057/palgrave.ejis.3000646>.
- [16] Sree Lakshmi, P., Deepak, A., Muthuvel, S.K., Amarnatha Sarma, C Design and Analysis of Stepped Impedance Feed Elliptical PatchAntenna Smart Innovation, Systems and Technologies, 2023, 334, pp. 63
- [17] Gupta, A., Mazumdar, B.D., Mishra, M., ...Srivastava, S., Deepak, A., Role of cloud computing in management and education, *Materials Today*:
- [18] Dhabliya, D. (2021). An Integrated Optimization Model for Plant Diseases Prediction with Machine Learning Model . *Machine Learning Applications in Engineering Education and Management*, 1(2), 21–26. Retrieved from <http://yashikajournals.com/index.php/mlaeem/article/view/15>
- [19] Dr. Bhushan Bandre. (2013). Design and Analysis of Low Power Energy Efficient Braun Multiplier. *International Journal of New Practices in Management and Engineering*, 2(01), 08 - 16. Retrieved from <http://ijnpme.org/index.php/IJNPME/article/view/12>