IJMRSETM

e-ISSN: 2395 - 7639

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

Volume 11, Issue 5, May 2024

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.802

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.802 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 11, Issue 5, May 2024 |

Designing a Deep Learning-Based Architecture for Efficient Big Data Collection and Organization: Challenges and Opportunities

Chandrakant Maharji, Dr. Sumit Bhattacharjee, Dr. Pravin Adivadekar

Research Scholar, Singhania University, Rajasthan, India

Research Guide, Singhania University, Rajasthan, India

Research Co-guide, Singhania University, Rajasthan, India

ABSTRACT: This paper explores the integration of deep learning into big data analytics to enhance data collection and organization efficiency. The study aims to develop an environment that combines big data management frameworks with deep learning systems to improve traditional data analysis methods. By leveraging the power of deep learning, the study seeks to extract complex patterns from large datasets, enabling the use of simpler linear models for classification and prediction tasks. The importance of high-quality data for deep learning models is emphasized, highlighting the need for accurate, diverse, and timely data. The paper discusses the challenges and benefits of using deep learning in big data analytics, addressing issues such as data quality, security, and computational intensity. Through a comprehensive review of literature and methodologies, the study aims to contribute novel algorithms and approaches to advance the understanding and application of deep learning in the context of big data analytics. Overall, the research aims to pave the way for more efficient and effective data analysis and decision-making processes in the era of big data.

I. INTRODUCTION

Big data analytics and deep learning are two interconnected fields that have revolutionized the way data is processed, analyzed, and utilized in various industries.Big data analytics and deep learning are two critical components of modern data science that have transformed how organizations process, analyze, and extract insights from vast amounts of data. Big data refers to the massive volume of structured and unstructured data generated by various sources, such as social media, sensors, and transactional systems. Traditional data analytics techniques may not be sufficient to handle The volume of data, which is where deep learning comes in. Deep learning models are designed to learn and extract patterns from large datasets, making them well suited for big data analytics tasks such as image recognition, natural language processing, and speech recognition.

Deep learning models are particularly effective in handling the complexities and nonlinear patterns present in big data, enabling organizations to extract valuable insights and improve their operations. The use of deep learning in big data analytics has led to significant improvements in accuracy and efficiency, enabling organizations to make better decisions based on insights derived from their data. Despite the challenges posed by the scale and complexity of big data, the integration of deep learning techniques offers promising solutions for enhancing data analysis, prediction accuracy, and overall efficiency in leveraging the vast amounts of data generated in today's digital age.

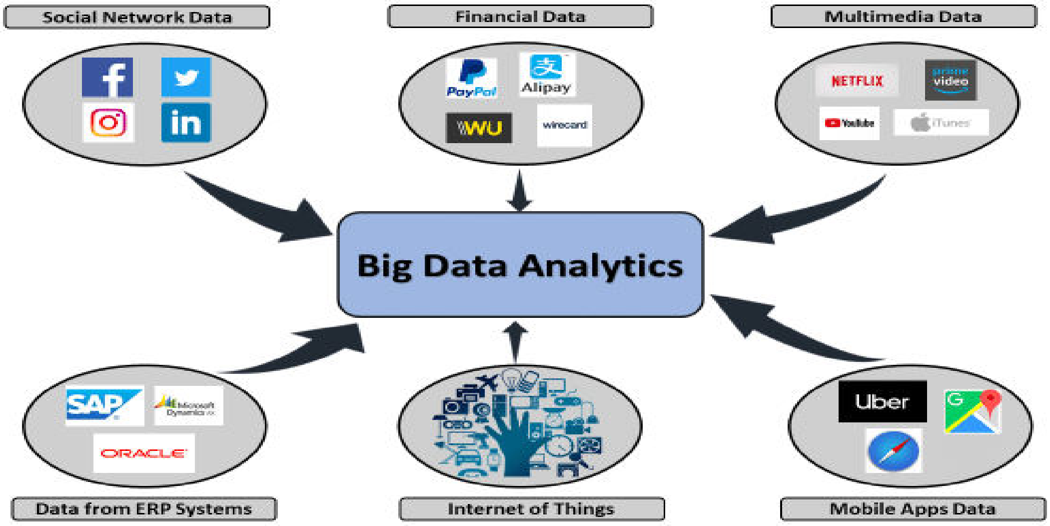


Figure 1: Big data analytics data sources.

The paper aims to explore how deep learning can be used for dealing with several crucial issues in big data analytics, such as extracting complex patterns from substantial volumes of data, semantic indexing, data tagging, quick info retrieval, and simplifying discriminative activities. The study also aims to identify areas that require further exploration to integrate particular problems launched by big data analytics, such as streaming data, high dimensional data, scalability of models, and distributed computing. By presenting insights to relevant succeeding works, the study aims to pose questions that can guide future research in this area.
The paper aims to contribute to the development of novel algorithms and approaches for big data analytics, leveraging the power of deep learning. The findings of this study can potentially enhance traditional big data analysis methods and provide insights into the effectiveness of the system through the application of tools to various deep learning applications. Overall, this study seeks to advance the understanding and application of deep learning in the context of big data analytics, paving the way for more efficient and effective data analysis and decisionmaking processes.
The intersection of big data analytics and deep learning has led to significant advancements in data analysis and decision making processes. Deep learning models are particularly effective in handling the complexities and nonlinear patterns present in big data, enabling organizations to extract valuable insights and improve their operations. Despite the challenges posed by the scale and complexity of big data, the integration of deep learning techniques offers promising solutions for enhancing data analysis, prediction accuracy, and overall efficiency in leveraging the vast amounts of data generated in today's digital age.
Big data analytics poses challenges for traditional machine learning and feature engineering algorithms due to the complexity and nonlinear patterns often observed in big data. Deep learning, on the other hand, enables the extraction of such features, allowing for the use of simpler linear models for tasks like classification and prediction. This is particularly important when dealing with the scale of big data.

The objectives of the study are:
1. To develop an environment for integrating big data management and processing frameworks with deep learning systems.
2. To improve traditional big data analysis methods by leveraging the power of deep learning.
3. To evaluate the effectiveness of the system through the application of tools to various deep learning applications.

The study aims to contribute to the development of novel algorithms and approaches for big data analytics, leveraging the power of deep learning. The findings of this study can potentially enhance traditional big data analysis methods and provide insights into the effectiveness of the system through the application of tools to various deep learning applications. Overall, this study seeks to advance the understanding and application of deep learning in the context of big data analytics, paving the way for more efficient and effective data analysis and decision making processes.

II. LITERATURE REVIEW

The papers chosen for review cover a wide range of topics related to big data analytics, machine learning, and deep learning. They discuss the challenges and opportunities in handling and analyzing large scale data, present various algorithms and techniques for big data processing and analysis, and highlight the potential of deep learning in various

applications. The literature review covers both theoretical and practical aspects, providing insights into the state of the art methods and tools for big data analytics.
The paper [1] discusses the potential of deep learning techniques for big data analytics. It highlights the advantages of deep learning over traditional machine learning methods in handling large and complex datasets. The authors provide an overview of deep learning architectures like deep neural networks, convolutional neural networks, and recurrent neural networks, and their applications in various domains.
The survey paper [2] explores the use of machine learning techniques for big data processing. It covers different machine learning algorithms, including supervised, unsupervised, and reinforcement learning, and their applications in big data analytics. The paper also discusses the challenges and opportunities in integrating machine learning with big data processing.
The paper[3] provides a comprehensive review of deep learning applications and challenges in big data analytics. It discusses the potential of deep learning in various domains, such as computer vision, natural language processing, and speech recognition. The authors also highlight the challenges associated with deep learning, including the need for large labeled datasets, computational resources, and interpretability.
The survey paper [4] focuses on different platforms and tools available for big data analytics. It provides an overview of various open source and commercial platforms like Apache Hadoop, Apache Spark, and MongoDB, and their suitability for different types of big data analytics tasks.
The paper [5] surveys open source tools for machine learning with big data in the Hadoop ecosystem. It covers tools like Apache Mahout, Apache Spark MLlib, and Apache Singa, and their capabilities for implementing machine learning algorithms on large scale datasets.
The paper [6] proposes a deep learning approach for traffic flow prediction using big data. It demonstrates the effectiveness of deep neural networks in capturing the complex patterns and dependencies in traffic data, leading to improved prediction accuracy.
The paper [7] presents a modular software architecture for processing big geospatial data in the cloud. It discusses the challenges associated with handling and analyzing large scale geospatial data and proposes a scalable and flexible solution based on cloud computing.
The paper [8] introduces a community activity prediction model based on big data analysis. It leverages deep learning techniques to analyze large scale social network data and predict user activities and interests within communities.
The paper [9] proposes a distributed single linkage hierarchical clustering algorithm using MapReduce for big data applications. It addresses the scalability challenges of traditional hierarchical clustering methods and demonstrates the efficiency of the proposed algorithm on large datasets.
The paper [10] presents a crossindustry study of MapReduce workloads in big data systems. It analyzes the characteristics and performance of interactive analytical processing tasks and provides insights into optimizing big data analytics systems.
The paper [11] focuses on spatiotemporal data mining in the era of big spatial data. It discusses the challenges and opportunities in analyzing large scale geospatial and temporal data and presents algorithms and applications for various spatiotemporal data mining tasks.
The paper [12] introduces Jet, an embedded domain specific language (DSL) for high performance big data processing. It provides a concise and expressive syntax for data parallel programming, enabling efficient execution on various parallel platforms.
The paper[13] presents a comparative review of state of the art commercial visual analytics systems for big data. It evaluates the capabilities and features of different systems in areas like data handling, analytics, and visualization, providing insights for choosing the appropriate tool for specific use cases.
The paper[14] introduces Starfish, a self tuning system for big data analytics. It addresses the challenges of configuring and optimizing big data systems by automatically adapting the system's configuration based on workload characteristics and resource constraints.
The paper [15] provides an understanding of machine learning with deep learning, including architectures, workflows, applications, and future directions. It discusses the fundamental concepts, techniques, and recent advancements in deep learning for various domains.

Overall, the literature review highlights the importance of integrating big data analytics and deep learning, as well as the challenges and opportunities of this integration. The studies and papers reviewed provide valuable insights into the applications and limitations of deep learning in big data analytics, as well as the need for further research in this area. By leveraging the power of deep learning, organizations can improve their big data analysis methods, leading to more efficient and effective decision making processes.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.802 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 11, Issue 5, May 2024 |

III. METHODOLOGY

The methodology used in the study to design a deep learning based architecture for efficient big data collection and organization involves several steps. First, the study conducts a literature review to understand the current state of the art in big data analytics and deep learning. The literature review covers various topics, including the meaning of big data, the concept of deep learning, the relation between big data and deep learning, and the benefits of big data analysis. The literature review helps to identify the gaps in the current research and the potential areas for improvement.

Next, the study defines the objectives of the study, which include developing an environment for integrating big data management and processing frameworks with deep learning systems, improving traditional big data analysis methods, and evaluating the effectiveness of the system through the application of tools to various deep learning applications. The study also emphasizes the importance of high quality data for deep learning models and the role of artificial intelligence in supporting big data architecture, efficient data management, enhanced analysis, and impactful results.

The study then proposes a methodology for designing a deep learning based architecture for efficient big data collection and organization. The proposed methodology involves several steps, including data collection, data preprocessing, data analysis, and model training. The data collection step involves gathering large volumes of data from various sources, including social media, sensors, and transactional systems. The data preprocessing step involves cleaning, transforming, and organizing the data into a format suitable for deep learning models. The data analysis step involves using deep learning models to analyze and extract insights from the data. The model training step involves training deep learning models using large datasets to improve their accuracy and efficiency.

The study also highlights the challenges of deep learning models, including the need for large amounts of data to train effectively, the computational intensity of the training process, and the difficulty in interpreting the models' decisions. To address these challenges, the study proposes using advanced hardware, such as graphics processing units (GPUs) and hardware accelerators, to improve the computational efficiency of deep learning models. The study also proposes using advanced training algorithms to improve the accuracy and efficiency of the models.

Finally, the study evaluates the effectiveness of the deep learning based architecture for efficient big data collection and organization by applying it to various deep learning applications. The evaluation step involves measuring the accuracy, efficiency, and scalability of the system and comparing it to traditional big data analytics methods. The study also identifies areas for further research, such as defining data sampling criteria, domain adaptation modeling, defining criteria for obtaining useful data abstractions, repairing semantic indexing, semi supervised learning, and active learning.

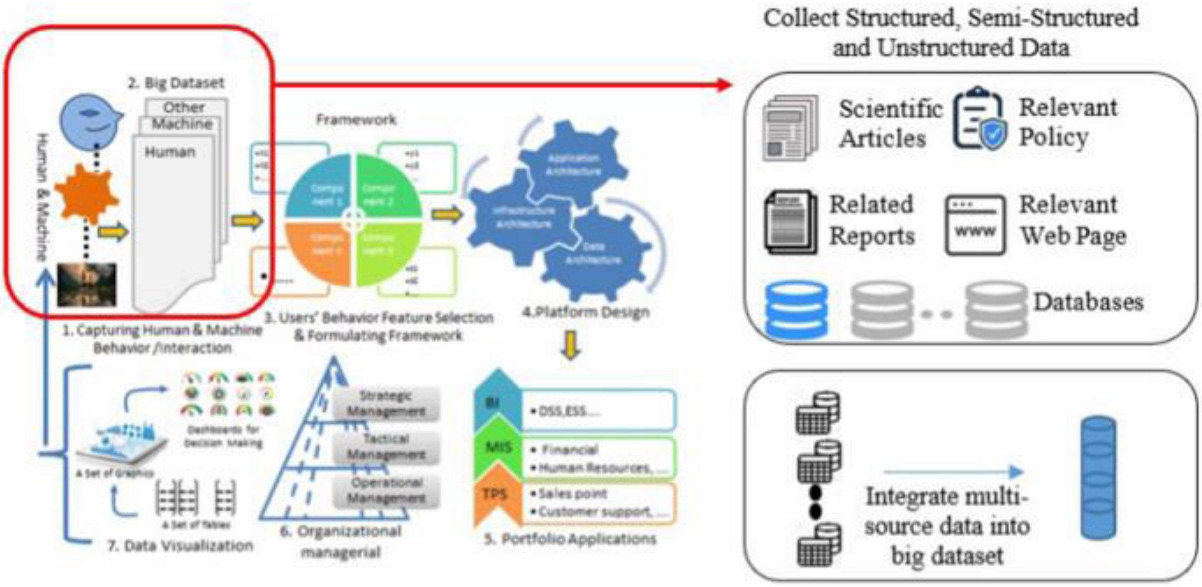


Figure 2: Dataset for big data analytics.

Figure 2 depicts a framework for collecting, integrating, and analyzing structured, semistructured, and unstructured data from various sources to create a large, multisource dataset for big data analytics.

The framework consists of the following key components:

1. Exploring Domain & Method: This step involves identifying relevant data sources and methods for data collection. The data sources can be human generated (e.g., social media, surveys) or sensor generated (e.g., IoT devices, scientific instruments).
2. Big Dataset: This component represents the collection of structured (e.g., databases), semistructured (e.g., scientific articles, related reports), and unstructured (e.g., social media, web pages) data from various sources as shown in figure 2.
3. Framework: This is the core component that facilitates the integration and processing of the collected data. It consists of several subcomponents:
   a. Data Management: Responsible for storing, organizing, and managing the collected data.
   b. Data Analytics: Includes techniques and tools for analyzing and extracting insights from the data.
   c. Process Management: Oversees the various processes involved in data collection, integration, and analysis.
   d. Resources: Represents the computing resources (e.g., storage, processing power) required for big data processing.
   e. Visualization: Tools and techniques for visualizing and presenting the analyzed data and insights.
4. Platform Design: This step involves designing and developing the platform or infrastructure required to support the big data analytics framework, including hardware, software, and networking components.

The ultimate goal of the framework is to integrate the multisource big dataset, consisting of structured, semistructured, and unstructured data, to enable comprehensive data analysis and gain valuable insights.

Figure 2 also highlights the importance of considering relevant scientific articles, related reports, web pages, and databases as potential data sources for building a comprehensive big data analytics solution.

IV. RESULTS

The study on designing a deep learning based architecture for efficient big data collection and organization has yielded significant findings regarding the benefits and challenges of using Big Data analytics and Deep Learning. Here are the key results:

1. Benefits of Big Data Analytics and Deep Learning:
   i. Competitive Advantage: Organizations can gain a competitive edge by effectively analyzing big data using deep learning models. This enables them to identify new opportunities, enhance customer experience, and optimize operations.
   ii. Improved Decision Making: Big data analytics, coupled with deep learning, provides valuable insights that empower organizations to make data driven decisions, leading to better outcomes based on factual information rather than intuition.
   iii. Enhanced Customer Experience: By leveraging big data analytics and deep learning, organizations can better understand customer needs and preferences, resulting in improved customer satisfaction and loyalty.
2. Challenges of Big Data Analytics and Deep Learning:
   i. Data Quality: Ensuring high quality data is crucial for accurate analysis and decision making. Poor data quality can lead to incorrect insights and decisions.
   ii. Data Security: Handling large volumes of sensitive data in big data analytics poses security risks that need to be addressed to protect data privacy and integrity.
   iii. Technical Expertise: Big data analytics and deep learning require specialized technical skills and expertise, which can be a challenge for organizations lacking the necessary resources and knowledge.

Overall, the study highlights the immense potential of combining Big Data analytics with Deep Learning to unlock valuable insights, improve decision making processes, and enhance customer experiences. While there are challenges to overcome, such as data quality, security, and technical expertise, the benefits of leveraging these technologies for data analysis and prediction are substantial. The findings underscore the importance of integrating advanced analytics techniques like deep learning into big data processing to drive innovation, efficiency, and competitiveness in today's data driven landscape.

V. DISCUSSION

Based on the results, deep learning has been shown to be effective in analyzing large datasets, particularly in extracting complex patterns from unstructured data. Deep learning models can learn and extract patterns from large datasets, making them well-suited for big data analytics tasks such as image recognition, natural language processing, and speech recognition. The use of deep learning in big data analytics has led to significant improvements in accuracy and efficiency, enabling organizations to make better decisions based on insights derived from their data.

However, there are also challenges associated with using deep learning for big data analytics. One of the main challenges is the data-hungry nature of deep learning models, which require large amounts of high-quality data for training. This can be a challenge for organizations that lack sufficient data or have data quality issues. Additionally, deep learning models can be complex and require significant computational resources, which can be a barrier for some organizations.

To address these challenges, future research could focus on developing more efficient deep learning models that can handle large datasets with less computational resources. Additionally, there is a need for more research on data quality issues and how to address them in the context of big data analytics.

Another area for future research is the integration of deep learning with other big data analytics techniques, such as machine learning and artificial intelligence. By combining these techniques, organizations can gain even more insights from their data and make more informed decisions.

Thus, deep learning has the potential to revolutionize big data analytics, but there are also challenges that need to be addressed. Future research should focus on developing more efficient deep learning models, addressing data quality issues, and integrating deep learning with other big data analytics techniques. By doing so, organizations can unlock the full potential of their data and make better decisions based on data-driven insights.

V. CONCLUSION

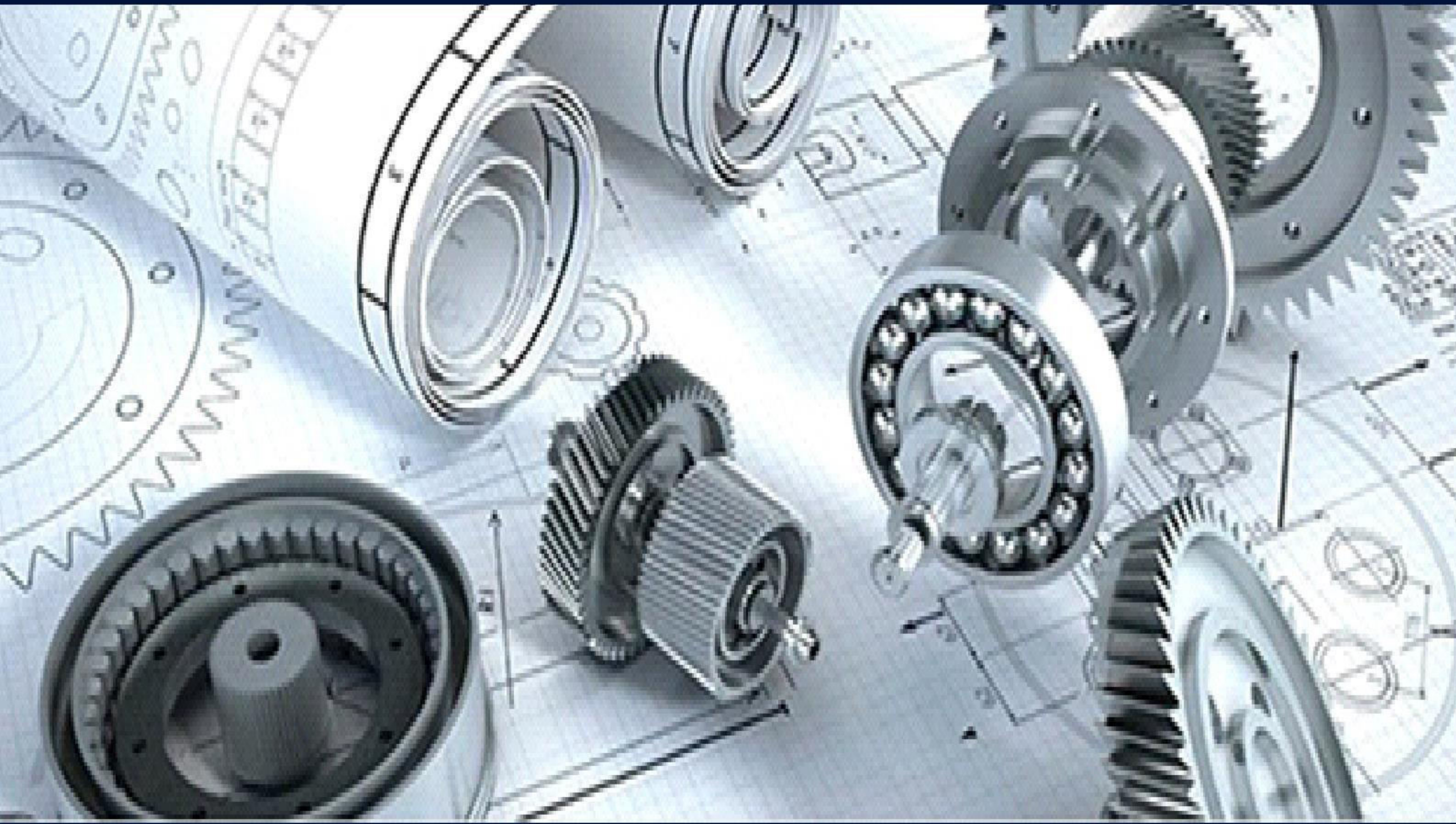
The study aims to explore the concept of big data and delve into the concept of deep learning. The objectives of the study include developing an environment for integrating big data management and processing frameworks with deep learning systems, improving traditional big data analysis methods, and evaluating the effectiveness of the system through the application of tools to various deep learning applications. Big data analytics poses challenges for traditional machine learning and feature engineering algorithms due to the complexity and non-linear patterns often observed in big data. Deep learning, on the other hand, enables the extraction of such features, allowing for the use of simpler linear models for tasks like classification and prediction. This is particularly important when dealing with the scale of big data. To achieve the objectives of the work, it is crucial to design a big data architecture that can handle high volumes, large variety, and high-velocity data. This architecture should provide the logical and physical capability to ingest, process, and analyze the data effectively. Additionally, the study emphasizes the role of artificial intelligence (AI) in supporting big data architecture, efficient data management, enhanced analysis, and impactful results. Furthermore, the study highlights the importance of high-quality data for deep learning models. Data quality, including volume, variety, velocity, and accuracy, should not be compromised to ensure the success of the integration of big data management and deep learning systems. By addressing these objectives, the study aims to contribute to the development of novel algorithms and approaches for big data analytics, leveraging the power of deep learning. The findings of this study can potentially enhance traditional big data analysis methods and provide insights into the effectiveness of the system through the application of tools to various deep learning applications. Overall, this study seeks to advance the understanding and application of deep learning in the context of big data analytics, paving the way for more efficient and effective data analysis and decision-making processes.

REFERENCES

[1] Nagwa M. Elaraby, Mohammed Elmogy, Shereif Barakat (2016) “Deep Learning: Effective Tool for Big Data Analytics” International Journal of Computer Science Engineering (IJCSE) Vol. 5 No.05 Sep 2016, ISSN : 2319-7323 pp:- 254-262

[2] Yuhua Xu and Shuo Feng ,Junfei Qiu, Qihui Wu, Guoru Ding, “A survey of machine learning for big data Processing”, EURASIP Journal on Advances in Signal Processing (2016) 2016:67

[3] Najafabadi, Maryam & Villanustre, Flavio & Khoshgoftaar, Taghi & Seliya, Naeem & Wald, Randall & Muharemagic, Edin. (2015). Deep learning applications and challenges in big data analytics. Journal of Big Data. 2. 10.1186/s40537-014-0007-7.
[4] Singh, D.; Reddy, C.K. A survey of Platforms for Big Data Analytics. J. Big Data 2015, 2, 8
[5] Landset, S.; Khoshgoftaar, T.M.; Richter, A.N.; Hasanin, T. A survey of Open Source tools for machine learning with big data in the Hadoop ecosystem. J. Big Data 2015, 2, 24.33
[6] Lv, Y., Duan, Y., Kang, W., Li, Z. and Wang, F.-Y. (2015), 'Traffic flow prediction with big data: a deep learning approach', IEEE Transactions on Intelligent Transportation Systems 16(2), 865–873.
[7] Kramer, M. and Senner, I. (2015), 'A modular software architecture for processing of ¨ big geospatial data in the cloud', Computers and Graphics 49, 69–81.
[8] Zhang, Y., Chen, M., Mao, S., Hu, L. and Leung, V. (2014), 'Cap: Community activity prediction based on big data analysis', Ieee Network 28(4), 52–57.
[9] Jin, C., Patwary, M. M. A., Agrawal, A., Hendrix, W., Liao, W.-k. and Choudhary, A. (2013), 'Disc: A distributed single-linkage hierarchical clustering algorithm using mapreduce', work 23, 27.
[10] Chen, Y., Alspaugh, S. and Katz, R. (2012), 'Interactive analytical processing in big data systems: A cross-industry study of mapreduce workloads', Proceedings of the VLDB Endowment 5(12), 1802–1813.
[11] Vatsavai, R. R., Ganguly, A., Chandola, V., Stefanidis, A., Klasky, S. and Shekhar, S. (2012), Spatiotemporal data mining in the era of big spatial data: algorithms and applications, in 'Proceedings of the 1st ACM SIGSPATIAL international workshop on analytics for big geospatial data', ACM, pp. 1–10. 34
[12] Ackermann, S., Jovanovic, V., Rompf, T. and Odersky, M. (2012), Jet: An embedded dsl for high performance big data processing, in 'International Workshop on End-to-end Management of Big Data (BigData 2012)', number EPFL-CONF-181673.
[13] Zhang, L., Stoffel, A., Behrisch, M., Mittelstadt, S., Schreck, T., Pompl, R., Weber, S., Last, H. and Keim, D. (2012), Visual analytics for the big data era comparative review of state-of-the-art commercial systems, in 'Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on', IEEE, pp. 173–182.
[14] Herodotou, H., Lim, H., Luo, G., Borisov, N., Dong, L., Cetin, F. B. and Babu, S. (2011), Starfish: A self-tuning system for big data analytics., in 'Cidr', Vol. 11, pp. 261– 272.
[15] Taye, M.M. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. Computers 2023, 12, 91. https://doi.org/10.3390/computers12050091.
[16] Chen, M., Mao, S. & Liu, Y. Big Data: A Survey. Mobile Netw Appl 19, 171–209 (2014). https://doi.org/10.1007/s11036-013-0489-0
[17] Alharbi and Alghamdi, Enhancing Performance of Educational Data, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 19 (2019) pp. 3814-3819
[18] Lazer, D.M., Kennedy, R., King, G., Vespignani, A., 2014. The parable of Google Flu: traps in big data analysis. Science 343, 1203–1205.
[19] Zongijjan. A survey on deep learning for big data. https://www.researchgate.net/publication/326099136_A_survey_on_deep_learning_for_big_data. [Accessed August 30, 2022]
[20] Yunchuan Sun, et al, Deep Learning for Big Data Analytics: A Review" by. (2019)
[21] Shubhashish G. et al, Survey of Deep Learning Technologies, DOI: https://doi.org/10.21203/rs.3.rs-665584/v1
[22] Sharma A, Virmani T, Pathak V, Sharma A, Pathak K, Kumar G, Pathak D. Deep Learning Architectures for Big Data Analytics: A Review. Biomed Res Int. 2022 Jul 6;2022:7205241. doi: 10.1155/2022/7205241. PMID: 35845955; PMCID: PMC9279074.
[23] Ahmed, Muhammad & Rizvi, Syed & Hussain, Rashid & Ahmed, Jawwad,. Deep Learning Architecture for Big Data Analytics: A Comprehensive Review,. International Journal of Business Information Systems. 42. 371. 10.1504/IJBIS.2023.129718.
[24] China Internet Network Information Center, 2014a. Research Report on Social Media User Behaviors 2014.
[25] China Internet Network Information Center, 2014b. Statistical Report on Internet Development in China.
[26] Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M.W., Zook, M., 2013. Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. Cartogr. Geogr. Inf. Sci. 40, 130–139. http://dx.doi.org/ 10.1080/15230406.2013.777137.
[27] de Montjoye, Y.-A., Hidalgo, C.A., Verleysen, M., Blondel, V.D., 2013a. Unique in the crowd: the privacy bounds of human mobility. Sci. Rep. 3. http://dx.doi.org/ 10.1038/srep01376.
[28] de Montjoye, Y.-A., Quoidbach, J., Robic, F., Pentland, A., 2013b. Predicting personality using novel mobile phone-based metrics. In: Greenberg, A., Kennedy, W., Bos, N. (Eds.), Social Computing, Behavioral–Cultural Modeling and Prediction. Springer Berlin Heidelberg, pp. 48–55.
[29] Duggan, M., Brenner, J., 2013. The Demographics of Social Media Users – 2012. (accessed 16.12.15). Einav, L., Levin, J., 2014. Economics in the age of big data. Science 346, 1243089. http://dx.doi.org/10.1126/science.1243089.



INNO SPACE
SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL
OF MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

+91 99405 72462 +91 63819 07438 ijmrsetm@gmail.com