Exploring Deep Insights from Vast Data: an Overview of Deep Learning Techniques for Big Data

1st Sarhad Baez Hasan
Computer Science Department
Soran University
Faculty of Science
Soran, Kurdistan-Iraq
Sarhad.hasan@soran.edu.iq

2nd Shahab Wahhab Kareem
Department of Information System Engineering
Erbil Polytechnic University
Department of Computer Engineering
College of Engineering
Knowledge University
Erbil, Kurdistan-Iraq
shahab.kareem@epu.edu.iq

3rd Didar Dlshad Hamad Ameen
Computer Science Department
Faculty of Science
Soran University
Soran, Kurdistan-Iraq
didar.dlshad@soran.edu.iq

2024 International Conference on Electrical Engineering and Computer Science (ICECOS) | 979-8-3503-6825-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICECOS63900.2024.10791131

***Abstract*—Big data has emerged very fast, and this has brought both opportunities and problems that are related to the application of deep learning. This paper explores how deep learning can be implemented using big data and in particular the Apache Mahout library. Unlike most other machine learning tools, Mahout can perform distributed computation, especially while using the MapReduce frameworks including Apache Hadoop and Apache Spark. However, this study does not only focus on the explanation of the need for Mahout's integration and employment but also carries out a high-fidelity argument on how deep learning algorithms can be incorporated in the Mahout framework to optimize the manipulation of large and complicated datasets. Mahout-based contributions include the comparative analysis of key deep learning paradigms including distributed processing, parallel processing and transfer learning. This paper also includes examples of the Mahout deep learning integration use and its application in real contexts. This article describes the challenges that currently organizations face in handling big data and indeed, the potential threats these challenges pose to the success of deep learning techniques when used in organizations' big data processing and analysis.**

Keywords*— Big data, deep learning, Hadoop, spark, data analysis, data processing

I. Introduction

As the digital age proceeds, the number of data forms and the rate at which data is produced increases therefore the need for more refined means of analyzing this data. Of all the mentioned methods, deep learning is an innovative method that utilizes a mechanism to derive highly implicit relationships in big and diverse data sources. Nevertheless, the integration of deep learning with big data comes with a series of issues, some of which are the following: Inability to scale up the processing power and inadequate mechanisms for data organizing. Linear models used in conventional approaches to machine learning are appropriate less and less for the huge loads of data produced every day. This shortcoming calls for the need to adopt deep learning approaches which give specific techniques for handling of large and complex data sets. However, the use of these techniques in real-world big data problems comes with issues of volume, variety, velocity and veracity of big data.

Although countless efforts have been made to explore the challenges highlighted above and advance various big data frameworks, there is little understanding of how a number of these big data frameworks, particularly Apache Mahout, can be utilized all the more to enhance the deep learning ones. Mahout is one of the frameworks that can easily support distributed computing and to the best of our knowledge, has not been fully tested with the integration of deep learning mechanisms.

Due to the application of big data in nearly all fields nowadays there is a problem of big data deep learning models scalability, efficiency and accuracy. This need must therefore be met to sustain organizational competitiveness in the new age of analytics.

This paper tries to fill in this gap by investigating the application of deep learning methods under the Apache Mahout system. Specifically, the study will:

Assess the effectiveness of using deep learning methodologies like distributed computing and/or parallel processing and/or transfer learning in Mahout. [1,2,8,11,13].

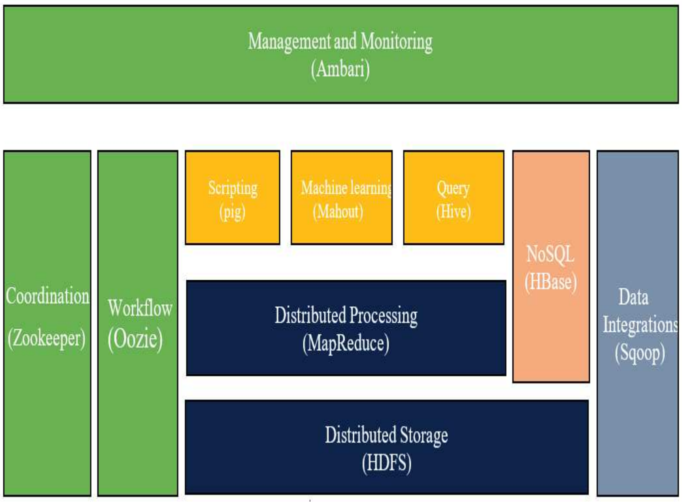

Fig. 1. Apache Hadoop Ecosystem [2]

A. Big Data Techniques in Deep Learning

In recent years, deep learning has become increasingly popular for handling and analyzing massive amounts of data. Traditional machine learning algorithms are becoming less effective as data generation grows daily, whereas deep learning algorithms are designed to manage large, complex datasets, making them ideal for Big Data analytics [3,7,8,10,17].

Several Big Data techniques can enhance the performance and scalability of deep learning:

- Distributed Computing: This involves distributing computing tasks across multiple machines, significantly improving processing speed and scalability for large datasets.
- Parallel Processing: By dividing a task into smaller tasks processed simultaneously by different processors, this technique speeds up data processing and model training.
- Data Compression: Reducing data size without losing crucial information helps decrease storage requirements and improve processing speed.
- Transfer Learning: Utilizing a pre-trained model for a new task saves time and resources, enhancing accuracy and speed by leveraging existing knowledge.
- Ensemble Learning: Combining multiple models to improve prediction accuracy and robustness, this technique reduces overfitting and enhances generalization performance.
- Feature Selection: Selecting the most important features from a dataset reduces dimensionality, improving processing speed and accuracy.
- Online Learning: Updating models with new data in real-time allows for the processing of large datasets without needing to store all data in memory, enabling real-time predictions.

These techniques optimize deep learning algorithms to handle large and complex datasets, improve processing speed and scalability, and make more accurate predictions. In today's data-driven world, these strategies help organizations leverage Big Data to gain insights, make informed decisions, and maintain competitiveness [17,18,19,22].

B. *Characteristics of Big Data*

The 7Vs: vast volume, high velocity, diverse variety, questionable veracity, and intrinsic value of data provide obstacles for urban computing. Every one of these traits presents formidable obstacles that necessitate specific methods in data fusion approaches [2,6].

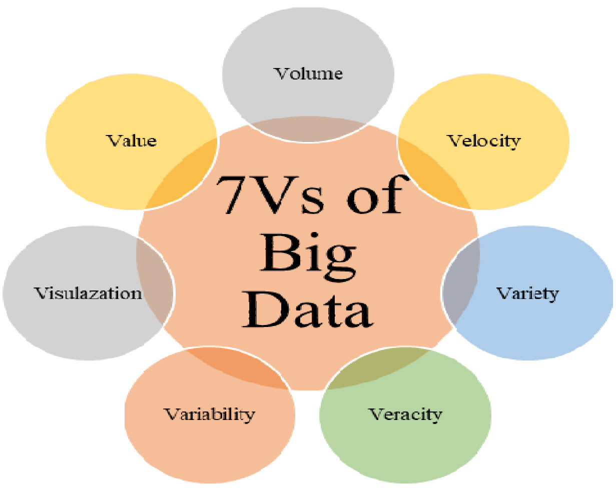

Fig. 2. Big data characteristics [2]

Volume: Handling the enormous amount of data produced in city settings.

Velocity: Making timely decisions by managing data streams in real-time.

Variety: Coherently handling a range of data sources and types.

Veracity: Resolving ambiguities and errors in urban data.

Value: Obtaining valuable insights from urban datasets through extraction.

Variability: This term describes data whose meaning changes over time in dictionaries.

Visualization: Without the aid of some sort of representation, such as a word cloud, dashboard, or graph, one cannot comprehend the insights contained in the data.

C. *Challenges Big Data in deep learning*

Deep learning has emerged as a powerful tool for processing and analyzing large amounts of data. However, the challenges associated with big data can significantly impact the performance and effectiveness of deep learning models. Some of the major challenges in big data for deep learning include: [2,5,11,13,15]

- Data Volume: Big data can be overwhelming in its sheer magnitude, and deep learning models need a lot of data to train well. Data processing, management, and storage problems may result from this.
- Data Variety: Big data derives trendy several dissimilar arrangements, including structured, semi-structured, and unstructured data. This can make it difficult to process and analyze the data using traditional methods, and can also impact the performance of deep learning models.

TABLE I. CATEGORIZATION OF BIG DATA [2]

Type	Examples
Structured	RDBMS (Oracle, DB2, Teradata, SAP, JDE, JDA)
Unstructured	Facebook, Twitter, Instagram, Audio, Video, Geo-Spatial
Semi-structured	XML, CSV, E-Commerce, IMS, E-Mail, EDI

- Data Velocity: Big data is often generated at a high speed and applying deep learning models to it for processing and analysis should also occur in real-time. This can be rather difficult, especially in the case of streaming data for which the funnel may not be readily defined.
- Data Veracity: The accuracy and correctness of big data could differ and it has an influence on the deep learning models. This is important in order to significantly build a strong foundation for the attainment of the desired results in the overall deep learning process.
- Data Security: Big data as a rule concerns highly important and confidential information, and, therefore, it is crucial to guarantee that this information will be protected against unauthorized access or usage.
- Model Complexity: Deep learning models can be intricate and this can cause a problem when it comes to the interpretation of the results of the analysis.

- Computational Resources: Deep learning models are computationally intensive that is in terms of processor usage and memory needed. It can be extremely tough to do this especially in the context of big data.
- Lack of Expertise: A market scarcity exists in both big data and deep learning and that is why the affairs often prove to be difficult when implementing them.
- Integration: One of the main drawbacks – the integration of big data and deep learning solutions to legacy systems and infrastructure may turn into a real problem.
- Ethical and Legal Considerations: And there are legal and ethical concerns as to the application of the big data and deep learning such as the question of privacy and consent as well as the ownership of the data.

Advanced deep learning methods such as distributed computing and parallel processing, data compression, transfer of learning, ensemble learning, feature selection and online learning could further improve big data analytics to accommodate large dataset flows, increase analytic speed, capacity and precision that could improve organization competitiveness [14,16,20].

II. Literature Review

The research focuses on big traffic and environmental data and social media data in cities and describes its possibilities for city planning and decision-making even if there are problems in management and data analysis. It also analyses deep learning technqiues of processing these datasets with a special emphasis on the feature that can learn the inputs and incorporate the heterogeneous sources. DL-output-based, DL-input-based, and DL-double-stage based fusion approaches are compared at different layers of deep learning. The study carried out reinforces the direction of future research effort in model interpretability, data quality, and privacy [1].

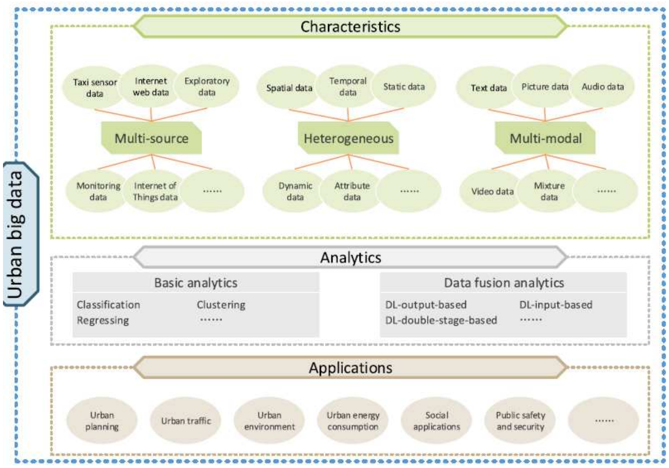

Fig. 3. Urban big data [1]

The study compares three popular big data processing frameworks: Hadoop, Spark, and Storm. The study makes the case that although all three-support processing of enormous amounts of data, their underlying systems and applications differ. Big data frameworks are categorized by the research into three categories: hybrid, batch-only, and stream-only. In the study Spark outperforms Hadoop and Storm in all categories. The study also discusses the characteristics of big data, including volume, velocity, variety, variability, veracity, visualization, and value [2]. The study explores the potential of deep learning for managing large data, highlighting its ability to deliver predictive analytics solutions. It discusses popular architectures like Deep Belief Networks (DBNs) and Convolutional Neural Networks (CNNs), which can handle high-level abstractions in data. Challenges in deep learning for big data include developing efficient algorithms, scalable architectures, and effective regularization techniques. Solutions like parallel processing, distributed computing, and cloud computing are suggested to enhance performance [3]. The study also explores potential applications in industries like healthcare, finance, and transportation. A method for elderly health monitoring using big data and deep ensemble learning is introduced, using the Hadoop MapReduce framework to process data from wearable sensors.[4]

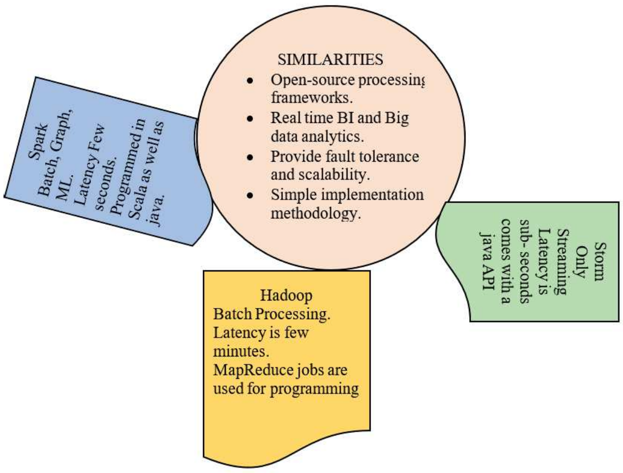

Fig. 4. Similarities between three frameworks [2]

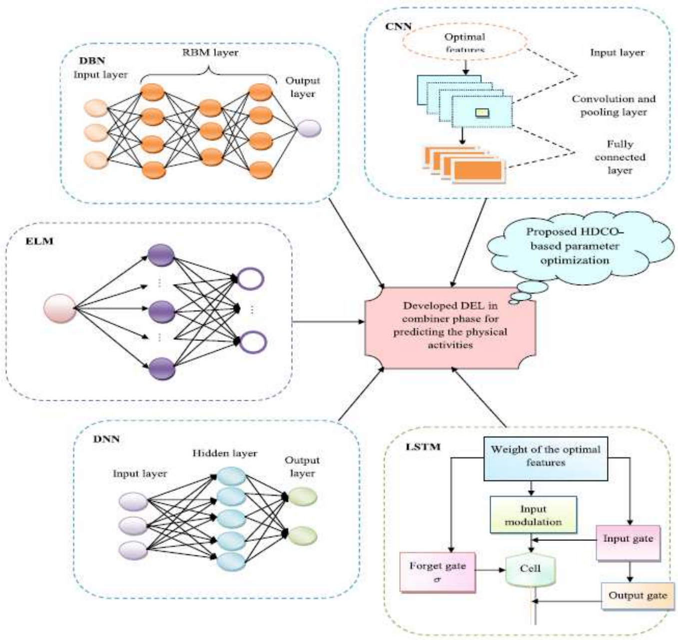

Fig. 5. Developed a categorization system for physical activities in the health monitoring system based on DEL [4].

The proposed system, using publicly available datasets, outperforms existing methods in accuracy and efficiency. However, it faces challenges like large data and computational resources. The system can improve elderly healthcare services by enabling real-time monitoring and early detection of health issues, leading to better outcomes and reduced healthcare costs [5]. The study provides an overview of big data system architecture and its challenges in analytics, defining big data from three perspectives: attributive, comparative, and architectural. It compares big data analytic approaches like Hadoop, HBase, MapReduce, and Hive based on attributes like scalability, flexibility, partition tolerance, schema usage, SQL interface, and costs. [6] The study introduces a 5Ws density model for analyzing large datasets, classifying BigData attributes into five dimensions. It measures flow patterns across datasets and displays densities using parallel coordinates, improving BigData visualization accuracy and reducing data clutter.

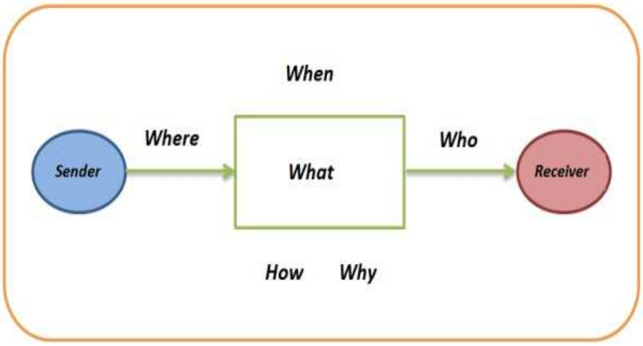

Fig. 6. 5Ws data pattern on each incident [6]

The proposed BigData visualization approach may not be suitable for all data types and lacks computational complexity, potentially impacting large datasets. Despite this, it provides a novel and effective method for large datasets.

The study introduces the Boosting Stochastic Newton Descent (BSND) algorithm for minimizing calibrated risks in big data classification, improving accuracy and speed on the ImageNet dataset, and highlighting the need for efficient algorithms [7].

TABLE II. ACCURACY AND PROCESSING TIME FOR BSND VERSUS SGD TECHNIQUES [7].

method	preprocessing time (s)	# of passes	training time (s)	total time (h)	top-1 accuracy	*top-5 accuracy*
SGD	-	200	132000	36.7	32.2%	55.34%
SGD-P	100	200	132000	36.8	34.1%	56.5%
SGD-QN	-	100	90000	25	32.9%	56.6%
BSND	100	10	7900	2.2	36.23%	59.06%

The study presents a method for modeling and optimizing big data using category theory and functional programming [8]. It uses basic concepts like objects, morphisms, functors, and natural transformations to model data structures as functors and transformations. The study also presents methods for modeling graph representations as functors and using natural transformations for efficient data structure transformations. The Dynamic Data Encryption Strategy (DDES) model enhances privacy and security in big data preservation using a cipher policy [9]. The GRAND framework, a graph neural network, enhances diagnostic resolution in post-silicon debug by identifying true/false candidates from a failing chip's diagnosis callout and employing transfer learning [10].

An investigation into deep learning methodologies for large data processing highlights the limitations of traditional methods and the need for advanced algorithms. The study compares deep learning techniques, including convolutional neural networks and deep belief networks, discussing their performance and challenges in handling large datasets.[11]

AI, deep learning, machine learning, and big data techniques are being used to optimize the design, production, operation, fault detection, and status monitoring of superconducting devices, such as flux pumps and electric machines. This highlights the potential for AI-driven intelligent condition monitoring and high-precision predictions for superconducting devices. [12,13]

III. Discussion and Analysis

When choosing Apache Mahout for this study the factors considered were the strong distributed computing feature of Mahout and compatibility with the big data frameworks such as Hadoop and Spark. Due to these features, Mahout can easily be used to scale deep learning algorithms to handle big data. Nevertheless, there were several challenges, including high computational resource requirements; more specifically, when training deep learning models on humongous datasets. This problem was solved in the current work by fine-tuning the data preprocessing phase as well as by employing the principles of parallelization that involved the distribution of the computation across the nodes in Hadoop. Moreover, the decision to use Mahout was also based on a greater compatibility with other Apache projects, making its use as a tool more diverse in terms of data storing and analysis.

The table 3 presents a comparison of thirteen studies on deep learning methods for big data analysis. The table provides information on the aim and objective of each study, the methods used, and the results obtained.

TABLE III. COMPARING THE STUDIES MENTIONED IN THE PROMPT.

Ref.	Methods	Results
1	Deep learning, feature-based, stage-based, semantic meaning-based, and stage-based data fusion techniques	Deep learning-based urban big data fusion techniques enhance air quality, traffic flow, and crowd flow forecasts, demonstrating potential in integrating data from multiple sources.
2	Comparing the processing times, CPU use, latency, execution times, and output of the task for Hadoop, Spark, and Storm experimentally	Spark performed better than Hadoop and Storm in terms of processing time, CPU usage, latency, execution time, and job performance.
3	Predictive sparse decomposition, unsupervised CNN training, gradient descent algorithm, mean squared-error loss function, deep learning, deep belief networks, convolutional neural networks, deep stacking networks, T-DSNs, backpropagation	Explores the challenges and opportunities of deep learning in big data analytics, highlighting its potential in handling noisy, incomplete, and diverse data sources.

	techniques.	
4	Deep Ensemble Learning (DEL), Hybrid Dingo Coyote Optimization (HDCO), Hadoop MapReduce	Compared to ELM, CNN, LSTM, DBN, DNN, and HealthFog, respectively, the created HDCO-DEL achieved 13.66%, 16.01%, 17.33%, 13.6%, and 14.01% greater accuracy on the second dataset. Generally, quite accurate in predicting movement.
5	Hadoop, Hive, MapReduce, NoSQL, NewSQL, BigData analytics techniques	The study explores the challenges and design of big data systems, comparing Hadoop with other frameworks like Hive, NoSQL, NewSQL, MapReduce, and HBase for data analytics.
6	5Ws data dimensions, 5Ws density model, 5Ws density parallel coordinates visualization	The new model can efficiently handle multiple dimensions and large datasets. The experiment shows that this model can be used for BigData visualization and analysis.
7	Stochastic Newton Descent, Boosting (BSND)	BSND, despite being faster than the advanced SGD algorithm, significantly enhances the accuracy of the SGD baseline on the vast ImageNet and Higgs dataset.
8	Category theory, functional programming language	Sample program demonstrated the effectiveness of using category theory and functional programming in the context of BigData
9	Dynamic Data Encryption Strategy (DDES), Dynamic Encryption Determination (DED) Algorithm, Privacy Weight Values (PWVs), Pairs Matching Collision (PMC)	Experimental findings show the proposed DDES model outperforms optimal encryption techniques in terms of privacy weights, execution time, and superior protection.
10	Graph Neural Networks transfer learning	When comparing diagnostic resolution with current diagnostics techniques, improvements range from 4.51x to 5.98x.
11	Deep learning techniques	The result shows that a variety of approaches combined with supervised and unsupervised training techniques may be used to build deep learning systems.
12	Map-Reduce algorithm, BigData analytics techniques	Improved data analysis and prediction process in cloud environment, Increased number of customers in cloud environment
13	AI, ML, DL, and Big Data techniques	Examines how AI, ML, DL, and Big Data approaches may be used to solve problems and offer the best results for a number of superconducting devices, such as electric machines, flux pumps, magnets, and other large-scale devices.

Liu et al. (2020) present deep learning-based fusion approaches categorized into DL-double-stage-based fusion, DL-output-based fusion, and DL-input-based fusion, enhancing the accuracy and efficiency of urban computing activities. Singh et al. (2023) compare Hadoop, Spark, and Storm, demonstrating Spark's superior performance in processing time, CPU usage, latency, execution time, and task performance. Chen and Lin (2014) discuss deep learning's potential for big data analytics, emphasizing strategies for training deep learning models on big data and its applications across various domains.

Abidi et al. (2023) proposes a deep ensemble learning-based system for healthcare monitoring, achieving high accuracy in predicting physical activities using Hadoop MapReduce and HDCO. Malhotra et al. (2017) discuss various big data analytic approaches and their challenges, comparing tools like Hadoop, HBase, MapReduce, and Hive based on scalability, flexibility, and other attributes. Zhang et al. (2014) introduce a method for visualizing big data using parallel coordinates with a density approach, classifying attributes into five dimensions and using parallel coordinates for visualization.

D'Ambrosio et al. (2014) propose the Boosting Stochastic Newton Descent algorithm for minimizing calibrated risks in big data classification, demonstrating its accuracy and efficiency. Thiry et al. (2018) introduce a method for modelling and optimizing big data using category theory and functional programming, leveraging formal frameworks and practical implementations. Jadhav et al. (2023) present the Dynamic Data Encryption Strategy concept, enhancing security and privacy in big data preservation using cipher policy and privacy classification approaches.

Wang et al. (2023) introduce GRAND, a graph neural network framework for enhancing diagnostic resolution in post-silicon debugging, leveraging transfer learning and GNNs. Jan et al. (2019) provide a comparative analysis of big data analytics deep learning methods, discussing their performance and limitations. Ramamoorthy and Rajalakshmi (2013) discuss the use of MapReduce and big data analytics techniques for optimizing data analysis in cloud environments, aiming to reduce complexity and improve efficiency. Finally, Yazdani-Asrami (2023) explores the potential of AI, deep learning, machine learning, and big data techniques to address challenges in superconducting devices.

IV. Conclusion

In the light of growing availability of big data, enhanced by the progress of deep learning techniques, there are both opportunities and threats. This paper was concerned with a range of deep learning techniques such as distributed computing, parallel processing, data compression, transfer learning, ensemble learning, feature selection, and online learning within the big data environment. Out research indicates that these techniques improve scalability, time and accuracy of large scale deep learning models on large datasets to noticeable extent. Notably, the combination of Apache Mahout with Hadoop and spark big data frames was discovered to fit best in the management of big data processing issues. Nonetheless, there are challenges that are still evident and which include the issues of big data, the number of data, speed, data quality, and security. Such issues demonstrate the imperative of further research and development of more sophisticated methods and, in particular, algorithms and frameworks that would complement the work with the big data more efficiently.

A. Future Research Directions:

- Optimization of Deep Learning Algorithms: More investigations should be carried out in the enhancement of deep learning algorithms, more especially when handling large and miscellaneous data types. This may in particular refer to considering new areas like, for example, combining different learning paradigms into one model.

- Integration with Emerging Technologies: It is concerning to explore the integration of deep learning to the enhance new generation big data computation tools like Apache Flink.
- Scalability and Resource Management: The same should also be studied in relation to deep learning methods and other methods for scaling up deep learning solutions especially in environments that are limited in terms of resource endowment. This could include data preprocessing break down, and more efficient method of training the model or bringing about the model.

Thus, addressing these areas, the further development of the field of big data analysis and its application in organizations, based on the effective use of deep learning by following the proposed best practices, can be expanded.

REFERENCES

[1] Liu, J., Li, T., Xie, P., Du, S., Teng, F. and Yang, X., 2020. Urban big data fusion based on deep learning: An overview. Information Fusion, 53, pp.123-133.

[2] Singh, A., Sandhu, R., Mehta, S., Giri, N.C., Kuziakin, O., Leliuk, S., Saprykin, R. and Dobrozhan, A., 2023, October. A Comparative Study of Bigdata Tools: Hadoop Vs Spark Vs Storm. In 2023 IEEE 4th KhPI Week on Advanced Technology (KhPIWeek) (pp. 1-5). IEEE.

[3] Chen, X.W. and Lin, X., 2014. Big data deep learning: challenges and perspectives. IEEE access, 2, pp.514-525.

[4] Abidi, M.H., Umer, U., Mian, S.H. and Al-Ahmari, A., 2023. Big Data-based Smart Health Monitoring System: Using Deep Ensemble Learning. IEEE Access.

[5] Malhotra, S., Doja, M.N., Alam, B. and Alam, M., 2017, May. Bigdata analysis and comparison of bigdata analytic approches. In 2017 International Conference on Computing, Communication and Automation (ICCCA) (pp. 309-314). IEEE.

[6] Zhang, J., Meng, Z. and Huang, M.L., 2014, November. BigData visualization: Parallel coordinates using density approach. In The 2014 2nd International Conference on Systems and Informatics (ICSAI 2014) (pp. 1056-1063). IEEE.

[7] D'Ambrosio, R., Belhajali, W. and Barlaud, M., 2014, October. Boosting stochastic newton descent for bigdata large scale classification. In 2014 IEEE International Conference on Big Data (Big Data) (pp. 36-41). IEEE.

[8] Thiry, L., Zhao, H. and Hassenforder, M., 2018, July. Categorical models for BigData. In 2018 IEEE International Congress on Big Data (BigData Congress) (pp. 272-275). IEEE.

[9] Jadhav, D.A., Thatipudi, J.G., Thakur, V., Nagendar, Y., Lakshmi, K.M. and Sharma, S., 2023, October. Enhanced Privacy and Security Maintenance to Preserve BigData using Cipher Policy. In 2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 218-224). IEEE.

[10] Wang, H., Zhang, Z., Xiong, H., Zou, D., Chen, Y. and Jin, H., 2023. GRAND: A Graph Neural Network Framework for Improved Diagnosis. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.

[11] Jan, B., Farman, H., Khan, M., Imran, M., Islam, I.U., Ahmad, A., Ali, S. and Jeon, G., 2019. Deep learning in big data analytics: a comparative study. Computers & Electrical Engineering, 75, pp.275-287.

[12] Ramamoorthy, S. and Rajalakshmi, S., 2013, July. Optimized data analysis in cloud using BigData analytics techniques. In 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.

[13] Yazdani-Asrami, M., 2023. Artificial intelligence, machine learning, deep learning, and big data techniques for the advancements of superconducting technology: a road to smarter and intelligent superconductivity. Superconductor Science and Technology, 36(8), p.084001.

[14] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), pp.436-444.

[15] Ayhan, S., De Oliveira, Í.R., Balvedi, G., Costas, P., Leite, A. and De Azevedo, F.C., 2023, October. Big data-driven prediction of airspace congestion. In 2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC) (pp. 1-8). IEEE.

[16] Zhang, Q., Yang, L.T., Chen, Z. and Li, P., 2018. A survey on deep learning for big data. Information Fusion, 42, pp.146-157.

[17] Goodfellow, I., Bengio, Y. and Courville, A., 2016. Deep learning. MIT press.

[18] Gheisari, M., Wang, G. and Bhuiyan, M.Z.A., 2017, July. A survey on deep learning in big data. In 2017 IEEE international conference on computational science and engineering (CSE) and IEEE international conference on embedded and ubiquitous computing (EUC) (Vol. 2, pp. 173-180). IEEE.

[19] Jan, B., Farman, H., Khan, M., Imran, M., Islam, I.U., Ahmad, A., Ali, S. and Jeon, G., 2019. Deep learning in big data analytics: a comparative study. Computers & Electrical Engineering, 75, pp.275-287.

[20] Wilamowski, B.M., Wu, B. and Korniak, J., 2016, June. Big data and deep learning. In 2016 IEEE 20th Jubilee International Conference on Intelligent Engineering Systems (INES) (pp. 11-16). IEEE.

[21] Faker, O. and Dogdu, E., 2019, April. Intrusion detection using big data and deep learning techniques. In Proceedings of the 2019 ACM Southeast conference (pp. 86-93).

[22] Chae, S., Kwon, S. and Lee, D., 2018. Predicting infectious disease using deep learning and big data. International journal of environmental research and public health, 15(8), p.1596.

[23] Peng, L., Peng, M., Liao, B., Huang, G., Li, W. and Xie, D., 2018. The advances and challenges of deep learning application in biological big data processing. Current Bioinformatics, 13(4), pp.352-359.

[24] Mohammadi, M., Al-Fuqaha, A., Sorour, S. and Guizani, M., 2018. Deep learning for IoT big data and streaming analytics: A survey. IEEE Communications Surveys & Tutorials, 20(4), pp.2923-2960.

[25] Elaraby, N.M., Elmogy, M. and Barakat, S., 2016. Deep Learning: Effective tool for big data analytics. International Journal of Computer Science Engineering (IJCSE), 9.

[26] Khoshaba, F., Kareem, S., Awla, H. and Mohammed, C., 2022, June. Machine learning algorithms in Bigdata analysis and its applications: A Review. In 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (pp. 1-8). IEEE.

[27] Yousif, M.K., Dallalbashi, Z.E. and Kareem, S.W., 2023. Information security for big data using the NTRUEncrypt method. Measurement: Sensors, 27, p.100738.

[28] Kareem, S.W., 2022. A nature-inspired metaheuristic optimization algorithm based on crocodiles hunting search (CHS). International Journal of Swarm Intelligence Research (IJSIR), 13(1), pp.1-23.

[29] Khan, M., Jan, B. and Farman, H., 2019. Deep learning: convergence to big data analytics (p. 31). Singapore: Springer.

[30] Muhamad, H.A., Kareem, S.W. and Mohammed, A.S., 2022. A deep learning method for detecting leukemia in real images. NeuroQuantology, 20(7), p.2358.

[31] Li, Y., Huang, C., Ding, L., Li, Z., Pan, Y. and Gao, X., 2019. Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. Methods, 166, pp.4-21.