A Big Data, Bigger Impact: A Comprehensive Review of Machine Learning Advancements

1st Didar Dlshad Hamad Ameen
Computer Science Department
Soran University
Faculty of Science
Soran, Kurdistan-Iraq
didar.dlshad@soran.edu.iq

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

3rd Sarhad Baez Hasan
Computer Science Department
Soran University
Faculty of Science
Soran, Kurdistan-Iraq
sarhad.hasan@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/ICECOS63900.2024.10791082

***Abstract*— The exponential growth of data driven by the internet has necessitated effective extraction of insights, with big data and machine learning standing as pivotal tools. This paper aims to provide insights into the evolving landscape of machine learning and big data analytics through a comprehensive literature review. It seeks to underscore the key advancements and intersections between these domains and their transformative applications across industries. Specifically, the study will delve into the foundational theories that underpin machine learning algorithms, data collection and preprocessing techniques involved in big data analysis, and the prominent challenges associated with extracting insights from large and complex datasets. It will further explore major studies that exemplify the integration of machine learning methodologies and big data technologies for addressing real-world problems. The study will also examine visualization methods that facilitate comprehending insights gleaned from big data. Finally, it will discuss future research directions and highlight best practices for leveraging machine learning and big data analytics in decision-making processes.**

Keywords*— *Big data, Machine learning, Data analyze, unstructured data, structured data

I. INTRODUCTION

The exponential growth of data, fueled by the rapid development of the internet and connected devices, has created unprecedented challenges and opportunities for organizations. Effective extraction of valuable insights from these vast datasets is crucial for making informed decisions and maintaining a competitive edge. However, the sheer volume, variety, and velocity of big data make it difficult to analyze using traditional methods. This has led to the integration of machine learning techniques, which have shown significant promise in processing and analyzing large-scale data to uncover hidden patterns and insights.

In recent years, big data has become an increasingly important field of study, with applications in various industries such as healthcare, finance, and marketing. Machine Learning (ML), a subset of artificial intelligence, has emerged as a powerful tool for analyzing large and complex datasets. One of the key advancements in this field has been the development of machine learning algorithms, which can analyze large and complex datasets and extract valuable insights. These algorithms have the potential to make a significant impact on a variety of industries, Big data is the result of technological advancements in all facets of human existence, A vast quantity of data is generated by human beings, machines, and Internet of Things devices [2]. Big data is the main focus of every meaningful effort. Effective decision-making and proactive actions in the sector also heavily depend on the knowledge and insights obtained from this sort of field [3].

Despite the advancements in machine learning and big data analytics, organizations still face significant challenges in extracting meaningful insights from large and complex datasets. These challenges include handling the vast amount of unstructured data, ensuring data quality, and addressing the computational power required for real-time analysis. Additionally, the lack of expertise in applying advanced machine learning techniques to big data further complicates the process, often resulting in suboptimal decision-making.

This study aims to provide a comprehensive review of the advancements in machine learning and their application in big data analytics. Specifically, it seeks to explore the foundational theories that underpin machine learning algorithms, the data collection and preprocessing techniques involved in big data analysis, and the prominent challenges associated with extracting insights from large and complex datasets. By examining major studies that exemplify the integration of machine learning methodologies and big data technologies, this paper will highlight the transformative potential of these technologies across various industries. Additionally, the study will discuss visualization methods that aid in understanding insights derived from big data and propose future research directions to further enhance the field.

There is a lot of interest in machine learning because of the rise of big data [4]. Although their position has been questioned, the promise of massive data to apply sophisticated machine learning algorithms to novel types of markets and human behaviour is greater than ever [5, 6]. The source claims that although huge data offers an amazing degree of informative depth, standard machine learning is hindered by the enormous number of variables. All are growing larger and more complex, requiring a great deal of research and ML breakthroughs [7]. Machine learning will soon replace a number of occupations as a result of artificial learning approaches because learning algorithms are extremely reliable and can perform continuous learning, which reduces the need for human involvement [8].

II. BACKGROUND THEORY

A. Overview of Machine Learning

Theories from several scientific disciplines, such as artificial intelligence, mathematics, knowledge theories, and analytics, are included into the topic of machine learning. Experts in machine learning state that creating quick and effective learning algorithms that can handle prediction data is the primary goal [9]. A number of successful applications of machine learning techniques have already surfaced in domains such as voice recognition, computer vision, cognitive science, healthcare, and the Internet of Things [10].

ML is one of the many subfields within artificial intelligence that does numerical analysis and adapts according to training outcomes.

For estimate and prediction, machine learning may rely on real data recordings. These datasets must be utilized as test datasets for the computer algorithms that make decisions and the recognition system in order for the systems to start learning. One way to assess an algorithm's efficacy is to see if it works in the real world [11]. The resulting models in machine learning would be more precise and accurate the more and better training databases it had access to [12]. Machine learning uses a variety of methods, including Unsupervised (Clustering, Association Analysis, and Hidden Markov Model) and Supervised (Regression, Decision Tree, Random Forest, Classification) approaches. Supervised, unsupervised, and reinforcement learning are the three primary categories into which these algorithms fall. Supervised learning uses labeled data to make predictions, whereas unsupervised learning uses unlabeled data to make inferences about patterns. Reward-based learning, or reinforcement learning, gains knowledge through feedback systems or trial and error [13]. As illustrated in Fig. 1, the classification of ML techniques provides a clear understanding of the various approaches used in big data analytics. Fig. 2 compares the application of big data across various sectors from 2010 to 2017.

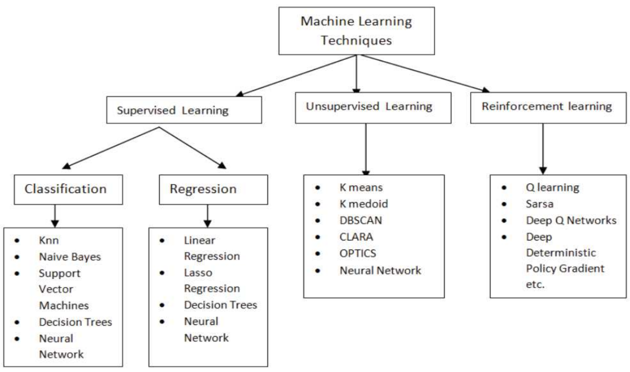

Fig. 1. Classification of machine learning techniques [22]

Typically, machine learning procedures include testing and training stages. While testing verifies the model's functionality, training involves developing models with labeled data. To categorize unknown data, machine learning involves building a model, utilizing it to identify characteristics and classes from training data, and lowering dimensionality as needed Machine learning applications include, for example, financial forecasting, text classification (including spam filtering and webpage categorization), handwriting recognition, and medical diagnosis support [7].

B. Big Data Analysis

Big data has emerged as a revolutionary technology in recent years, with the ability to process and analyze vast amounts of data in real-time. This has led to its application in various industries, including healthcare, transportation, banking, and insurance [14].

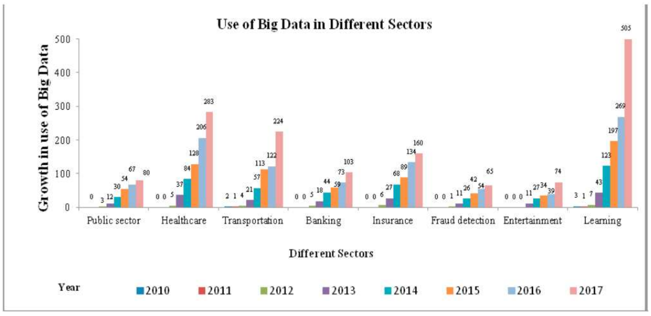

Fig. 2. Comparison of big data applications in various sectors from 2010 to 2017 [18].

Big data analytics refers to methods that analyze and produce large amounts of data in support of strategic planning. Finding the relationship between a dataset's hitherto unknown qualities by a variety of field techniques including machine learning, database systems, statistics, and mathematics is the aim of big data analysis, which specifically makes use of data mining [15](Fig.3).

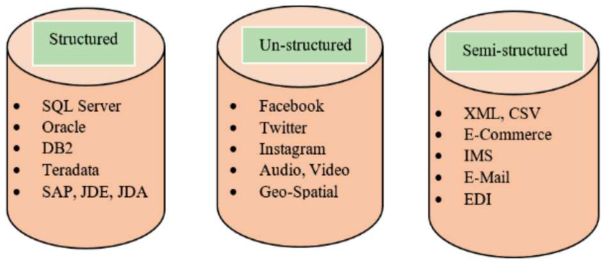

Fig. 3. Categorization of big data [16]

Aspects that are illustrated in Fig. 4, and explained below can be used to describe big data.

- Volume: Over time, the amount of data has grown tremendously. Big data systems get input data from many sources such as social media, banking transactions, sensors, and tags.
- Velocity: Over the past few years, data has been created at an alarming rate. Therefore, in order to handle the data that is coming in at such a rapid rate, specialist tools are needed.
- Variety: The generated data is not limited to any one type. Data can be found in text, JSON, XML, social media, and other formats.
- Variability: This describes facts whose dictionary definitions alter over time.
- Veracity: This term refers to the degree of reliability of the data.
- Visualization: Without the use of visual aids like word clouds, dashboards, graphs, and the like, one cannot comprehend the insights contained in data.
- Value: This is the completed item. One wants to confirm the Big Data Characteristics after resolving all the Vs, which need a significant investment of time and resources [16].

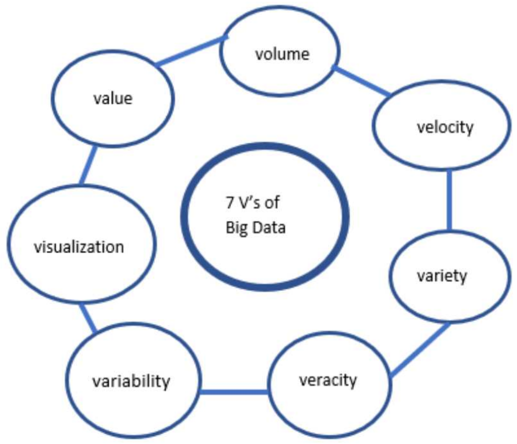

Fig. 4. Big data characteristics [17]

C. Big Data Challenges in ML

Big Data analytics presents the following challenges:

- Privacy should be viewed as a set of norms that govern ethical information flows but do not grant the capacity to keep data hidden.
- Confidential information can still be shared.
- Transparency is required for big data mining.
- Privacy may be at risk from big data.

Big data analytics presents several significant issues, including the need for specialized computing power, unstructured data and data provenance, missing or incomplete data, data quality, data security, and a shortage of expertise. As of right now, real-time analysis is not feasible with Hadoop's standard version. Purchase and usage of new tools are required. Hadoop will eventually be able to handle data in real-time, and there are now a number of solutions available to do the task. Using real-time insights necessitates a new method of working inside your organization. For example, obtaining insights every second would demand a different strategy and manner of working if your organization typically only receives insights once a week, which is extremely frequent in many organizations. Action is now necessary in real-time, rather than on a weekly basis, in response to insights. The culture will be impacted by this. Making your company an information-centric organization should be the goal [16].

III. Literature Review

This literature review was conducted by systematically collecting and analyzing relevant academic papers, conference proceedings, and technical reports published in the field of big data and ML. The sources were identified through searches in major electronic databases such as IEEE Xplore, SpringerLink, and Google Scholar, using keywords like "machine learning," "big data analytics," "data preprocessing," and "real-time data analysis." The selection criteria focused on studies that discussed the integration of ML techniques with big data, addressed the challenges of data analysis, and provided insights into the practical applications of these technologies. A total of 45 papers were reviewed, encompassing both foundational theories and recent advancements in the field.

[17] This study discusses the crucial role of big data in healthcare, emphasizing its continuous support in handling the growing data volumes in the industry. Through an empirical study, the authors explore the widespread use of big data and machine learning in healthcare, noting a lack of attention to privacy and security in existing implementations for disease diagnosis. In response, the paper proposes an innovative and secure healthcare information system design. This solution incorporates optimal storage, data security layers, and various techniques like masking encryption, activity monitoring, granular access control, dynamic data encryption, and endpoint validation. The resulting hybrid healthcare model is positioned as an effective and secure solution for disease diagnostic big data systems.

[18] The study explored the role of sensors, big data, and machine learning technologies in modern animal farming in Sensing and Bio-Sensing Research. In the study, shows how these advanced technologies can help animal farmers address several challenges related to production costs, efficiency, animal health and welfare. Also It's been described how sensors collected vast amounts of data from farms on factors like animal behaviors, movements, feeding intake and air quality. This big data was then analyzed using machine learning algorithms to gain insights and make predictions.

[19] This study examines the promises and challenges of using big data and machine learning approaches in personality measurement. The paper provide an overview of the types of digital data sources that are emerging as a result of modern technologies like smartphones, social media, wearable devices etc. They discuss how this digital trace data has already been used in initial research efforts to predict scores on traditional personality models like the Big Five. Several studies have shown it is possible to infer personality traits from digital behaviors and content on platforms like Facebook, Twitter, Instagram and more. However, the study note that more research is still needed to understand if big data captures unique personality-relevant aspects not captured by traditional surveys. Overall, this article aims to provide a framework and guiding principles for how personality psychologists can approach working with big data sources, while addressing important considerations around reliability, validity, fairness and ethical use of such data.

[20] This study aimed to propose an online client algorithm based on machine learning for IoT unstructured big data analysis. It first introduced the background of big data, machine learning, Internet of Things, and big data analysis. It then proposed the machine learning-based online client algorithm for unstructured big data analysis from IoT. The algorithm used machine learning techniques such as the K-nearest neighbor algorithm to perform data mining on online data entered by customers. Experiments were conducted on a big data platform testing the performance of the original IoT sensor data under different node sizes. The results showed that as the number of nodes increased, the time required for data processing also increased. With 2000 nodes, the analysis time was 200,000 seconds, while with 240,000 nodes, the time increased significantly to 7,400,000 seconds. This highlights how the proposed algorithm can analyze unstructured big data from IoT at scale using machine learning techniques.

[21] This study aims to discuss various machine learning techniques that can be used for big data analytics and address

the challenges faced in applying traditional machine learning algorithms on big data. A comprehensive survey of the literature was conducted to identify machine learning algorithms that have been successfully used for big data analytics. Various research papers that discussed deep learning techniques such as convolutional neural networks, deep belief networks, stacked auto-encoders etc. and their implementation on big data problems were reviewed. The findings from the literature reveal that deep learning models are useful for big data analytics due to their ability to learn feature hierarchies and handle large, complex datasets. Techniques such as distributed deep learning architectures, GPU processing and incremental/online learning were found to help address the issues of volume, velocity and variety associated with big data. The relevant studies highlighted in the literature survey provide valuable insights into the application of machine learning, particularly deep learning techniques, to big data analytics and the promising results achieved so far.

[7] The study conducted a systematic literature to analyze and compare various machine learning techniques for predictive analysis on complex datasets. 45 relevant research papers were identified, revealing the prevalence of supervised algorithms such as Random Forest and Support Vector Machine, along with growing interest in unsupervised techniques like clustering. Common evaluation metrics included accuracy, precision, recall, and AUC. The review emphasized research gaps in algorithm diversity, dataset heterogeneity, dynamic modeling, interpretability, and benchmarking. Overall, it provided insights into the landscape of machine learning applications on large real-world datasets.

[16] This paper aimed to perform a comparative analysis of the prominent big data processing frameworks namely Apache Hadoop, Apache Spark and Apache Storm. A descriptive research methodology was employed that involved analyzing the features, components and working mechanisms of these frameworks based on previous research works and literature. Relevant studies were identified through searches in major electronic databases using defined keywords related to big data, Hadoop, Spark and Storm. The key features of these frameworks including scalability, processing model, fault tolerance, iterative computation ability and language support were compared based on the findings from over 15 research papers. The results indicate that Spark has advantages over Hadoop and Storm in terms of faster processing time, supporting both batch and stream processing, ease of iterative computation and programming languages. However, Hadoop is still preferable for scenarios requiring high persistence storage and batch processing of huge datasets, while Storm is suitable for applications involving low latency stream processing. The study highlights the relative strengths of these frameworks to help researchers and practitioners select the most appropriate one based on their unique application requirements.

[4] This paper reviewed Machine Learning and Big Data Processing techniques proposed in recent research. The aim was to assess the role of ML-based algorithms and methods for Big Data Processing and Analytics when dealing with large, complex datasets. A variety of ML techniques were discussed for tasks like classification, clustering, recommendation systems and more. The paper also analyzed challenges in pre-processing large, heterogeneous datasets and proposed solutions like distributed feature selection and representation learning. A methodology of categorizing ML techniques into first, second and third generation based on data and model parallelism was discussed. The results indicate promising trends in combining ML and distributed computing frameworks to efficiently extract patterns and insights from Big data. However, more work is still needed to address issues around veracity, privacy and context-awareness as Big Data size and sources continue to grow exponentially.

[22] The study aimed to evaluate the performance of ensemble machine learning and scalable parallel discriminant analysis (EML-SPDA) for hyperspectral data analytics in agriculture. A review of existing literature on the use of machine learning techniques for hyperspectral data was conducted. Several representative works that applied support vector machines, random forests, artificial neural networks and other methods to classify crops, detect diseases and stresses, and perform other agriculture tasks were summarized. However, no prior work had explored the combined use of ensemble machine learning methods and scalable parallel discriminant analysis that takes into account both the spatial and spectral properties of hyperspectral data in an agriculture context. To address this gap, the EML-SPDA approach was developed and tested on hyperspectral data from agriculture. Preliminary results found that EML-SPDA achieved accurate classification and showed promising potential for hyperspectral data analytics in agriculture. Further experimental validation on more datasets is still needed to fully evaluate the performance and advantages of the proposed method.

[23] This study aimed to develop an accurate and efficient machine learning model for predicting disease risk using big data from healthcare communities. A convolutional neural network (CNN) based multimodal disease risk prediction (CNN-MDRP) algorithm was proposed to handle both structured and unstructured data for risk assessment. Real-life hospital data collected from 2013-2015 in central China was used, consisting of structured data such as patient demographics and examination results, as well as unstructured medical text. A latent factor model was applied to reconstruct missing values from the structured data. Cerebral infarction was identified as a major chronic disease affecting the region. The CNN-MDRP algorithm extracted features automatically from the different data types and fused them to generate predictions. Experimental results found that the CNN-MDRP approach achieved 94.8% accuracy and converged faster than alternatives, demonstrating its effectiveness in leveraging big data for precision medicine through multimodal predictive modeling. The proposed framework has potential for early detection of disease outbreaks and optimized healthcare resource allocation.

[24] This study aims to examine the applications of machine learning techniques for text analysis of big data documents. A wide range of machine learning and natural language processing methods were explored in existing scholarly literature and publications. Supervised learning algorithms such as Naive Bayes, decision trees, and support vector machines have been commonly used for tasks like document classification and sentiment analysis. Unsupervised techniques including topic modeling and clustering have also been applied to analyze large corpora and identify important themes or categories in unstructured

text data (cite studies). This study utilizes supervised machine learning models for predictive text analysis. A set of representative machine learning classifiers including Logistic Regression, Random Forest and Deep Neural Networks are trained on a manually annotated corpus of documents. The performance of these models is evaluated using standard metrics like accuracy, precision, recall to identify the most suitable technique for the task. The results provide insights into the effectiveness of different machine learning approaches for text analytics applications involving big data documents

TABLE I. A SUMMARY OF EXEMPLARY WORKS FOR COMPARISON ACROSS EACH STUDY

Authors	Methods	Results
[17]	Empirical study on big data and ML in healthcare, proposing optimal storage, data security layers, and various techniques.	Presents a hybrid healthcare model as an effective and secure solution for disease diagnostic big data systems.
[18]	Use sensors to collect data on animal behaviors, movements, feeding intake, and air quality. Analyze data using ML algorithms to gain insights and make predictions.	Demonstrates how advanced technologies can address challenges in animal farming, improving production costs, efficiency, and animal health.
[19]	Review types of digital data sources emerging from modern technologies. Discuss potential of digital trace data in predicting personality traits.	Indicates potential of big data in personality measurement, highlights need for further research on reliability, validity, and ethical use of data.
[20]	Introduce machine learning-based online client algorithm for unstructured big data analysis from IoT. Conduct experiments on a big data platform testing performance under different node sizes.	Shows scalability of proposed algorithm for analyzing unstructured big data from IoT using machine learning techniques.
[21]	Survey literature on machine learning algorithms used for big data analytics, particularly deep learning techniques. Review studies on distributed deep learning architectures, GPU processing, and incremental learning.	Highlights usefulness of deep learning models for big data analytics, addressing challenges of volume, velocity, and variety associated with big data.
[7]	Systematic literature review on machine learning techniques for predictive analysis. Identify prevalence of supervised and unsupervised algorithms.	Provides insights into prevalent machine learning techniques for predictive analysis on large real-world datasets.
[16]	Analyze features, components, and working mechanisms of Hadoop, Spark, and Storm. Compare key features based on findings from previous research.	Highlights relative strengths of Hadoop, Spark, and Storm for different application scenarios, aiding researchers and practitioners in framework selection.
[4]	Discuss ML techniques for classification, clustering, recommendation systems, etc. Analyze challenges and propose solutions like distributed feature selection and representation learning.	Indicates promising trends in combining ML and distributed computing frameworks for efficient extraction of patterns and insights from big data.
[22]	Review existing literature on ML techniques for hyperspectral data in agriculture. Develop and test EML-SPDA approach on hyperspectral data.	Preliminary results suggest accurate classification and promising potential for hyperspectral data analytics in agriculture with EML-SPDA approach.
[23]	Propose CNN-MDRP algorithm for multimodal disease risk prediction. Apply latent factor model to handle missing values. Test on real-life hospital data.	Achieves high accuracy and fast convergence, demonstrating effectiveness in leveraging big data for precision medicine through multimodal predictive modeling.
[24]	decision tree learning, Naive Bayes classification, and SVM classification; include k-means clustering, hierarchical clustering, and SOM clustering; deep learning methods apply multiple information processing stages in hierarchical architectures to learn feature representations from big data.	Provides insights into effectiveness of different ML approaches for text analytics applications involving big data documents.

IV. ANALYSIS AND DISCUSSION

The table I, summarizes the aims, objectives, methods, and results of various studies related to big data and machine learning. A total of 11 studies have been considered, and they cover a wide range of applications. In terms of methods, a variety of techniques have been used, such as empirical studies, machine learning algorithms, sensors, and deep learning models. The results of these studies show that machine learning can be an effective tool for analyzing big data and extracting valuable insights. However, the studies also highlight some challenges associated with big data and machine learning. For instance, the need for specialized computing power, the lack of expertise in this field, and privacy concerns are some of the challenges that need to be addressed. The study by Abdal wahid et al. (2019) discusses an enhancing approach using hybrid pailler and RSA for information security in big data, while the study by Hossain et al. (2019) reviews the application of big data and machine learning in smart grids, highlighting the associated security concerns.

Overall, the studies suggest that machine learning can transform industries and improve decision-making processes by providing valuable insights from big data. However, further research and development are needed to address the challenges associated with big data analysis. The studies also emphasize the importance of collaboration between different fields, such as computer science, statistics, and domain-specific knowledge, to develop effective machine learning solutions for big data.

V. CONCLUSION

In conclusion, this literature review has highlighted the significant role that ML plays in big data analytics, showcasing its potential to transform industries by providing deeper insights and improving decision-making processes. However, the challenges associated with big data, such as data quality, computational requirements, and the need for specialized expertise, remain significant barriers. Future research should focus on developing more efficient ML algorithms that can handle the scale and complexity of big data while minimizing computational costs. Additionally, there is a need for better data preprocessing techniques that can automate the cleaning and structuring of unstructured data. Research should also explore the integration of ML with other emerging technologies, such as edge computing and block-chain, to enhance data security and privacy. Lastly, more interdisciplinary studies that combine expertise from computer science, statistics, and domain-specific knowledge will be essential in advancing the field and addressing the remaining challenges. By addressing these areas, the field of ML and big data analytics can continue to evolve, offering even greater benefits across a wide range of industries.

REFERENCES

[1] Ozgur, C., Kleckner, M. and Li, Y., 2015. Selection of statistical software for solving big data problems: a guide for businesses, students, and universities. Sage Open, 5(2), p.2158244015584379.

[2] Ajin, V.W. and Kumar, L.D., 2016, May. Big data and clustering algorithms. In 2016 international conference on research advances in integrated navigation systems (RAINS) (pp. 1-5). IEEE.

[3] Abdalwahid, S.M.J., Yousif, R.Z. and Kareem, S.W., 2019. Enhancing approach using hybrid pailler and RSA for information security in bigdata. Applied Computer Science, 15(4), pp.63-74.

[4] Bhatnagar, R., 2018. Machine learning and big data processing: a technological perspective and review. In The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018) (pp. 468-478). Springer International Publishing.

[5] Hossain, E., Khan, I., Un-Noor, F., Sikander, S.S. and Sunny, M.S.H., 2019. Application of big data and machine learning in smart grid, and associated security concerns: A review. Ieee Access, 7, pp.13960-13988.

[6] 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.

[7] Tyagi, A.K., 2019, February. Machine learning with big data. In Machine Learning with Big Data (March 20, 2019). Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India.

[8] Son, L.H., Tripathy, H.K., Acharya, B.R., Kumar, R. and Chatterjee, J.M., 2019. Machine learning on big data: A developmental approach on societal applications. Big Data Processing Using Spark in Cloud, pp.143-165.

[9] Athmaja, S., Hanumanthappa, M. and Kavitha, V., 2017, March. A survey of machine learning algorithms for big data analytics. In 2017 International conference on innovations in information, embedded and communication systems (ICIIECS) (pp. 1-4). IEEE.

[10] Zhou, L., Pan, S., Wang, J. and Vasilakos, A.V., 2017. Machine learning on big data: Opportunities and challenges. Neurocomputing, 237, pp.350-361.

[11] Ray, S., 2019, February. A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 35-39). IEEE.

[12] Teri, S.S. and Musliman, I.A., 2019. Machine learning in big lidar data: A review. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, pp.641-644.

[13] Kareem, S. and Okur, M.C., 2018. Bayesian network structure learning using hybrid bee optimization and greedy search. Çukurova University.

[14] Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. META group research note, 6(70), p.1.

[15] KAREEM, S.W., 2020. Secure Cloud Approach Based on Okamoto-Uchiyama Cryptosystem. Journal of Applied Computer Science & Mathematics, 14(29).

[16] 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.

[17] Kaur, P., Sharma, M. and Mittal, M., 2018. Big data and machine learning based secure healthcare framework. Procedia computer science, 132, pp.1049-1059.

[18] Neethirajan, S., 2020. The role of sensors, big data and machine learning in modern animal farming. Sensing and Bio-Sensing Research, 29, p.100367.

[19] Alexander III, L., Mulfinger, E. and Oswald, F.L., 2020. Using big data and machine learning in personality measurement: Opportunities and challenges. European Journal of Personality, 34(5), pp.632-648.

[20] Hou, R., Kong, Y., Cai, B. and Liu, H., 2020. Unstructured big data analysis algorithm and simulation of Internet of Things based on machine learning. Neural Computing and Applications, 32, pp.5399-5407.

[21] Mittal, S. and Sangwan, O.P., 2019, January. Big data analytics using machine learning techniques. In 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 203-207). IEEE.

[22] Ang, K.L.M. and Seng, J.K.P., 2021. Big data and machine learning with hyperspectral information in agriculture. IEEE Access, 9, pp.36699-36718.

[23] Chen, M., Hao, Y., Hwang, K., Wang, L. and Wang, L., 2017. Disease prediction by machine learning over big data from healthcare communities. Ieee Access, 5, pp.8869-8879.

[24] Wang, L. and Alexander, C.A., 2016. Machine learning in big data. International Journal of Mathematical, Engineering and Management Sciences, 1(2), pp.52-61.