Journal of Technology Informatics and Engineering (JTIE)
Vol. 3 No. 3 December 2024
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 263-278

Enhancing Big Data Processing Efficiency in AI-Based Healthcare Systems: A Comparative Analysis of Random Forest and Deep Learning Algorithms

Priyadi*[1], Migunani[1], Dani Sasmoko[1]
Email: priyadi@stekom.ac.id, miguns25@gmail.com, dani@stekom.ac.id
[1]*Universitas Sains dan Teknologi Komputer, Semarang, Indonesia, 50192*
*Corresponding Author

Abstract

This research focuses on optimizing the speed of Big Data processing using Artificial Intelligence (AI) in healthcare applications. The study integrates Random Forest (RF) and Deep Learning (DL) algorithms with cloud-based computing systems to improve data processing efficiency. The dataset includes both structured data, such as Electronic Health Records (EHR), and unstructured data, like medical images. The results show that RF performs better with structured data, achieving a lower Mean Squared Error (MSE) and higher R-squared (R^2) than traditional methods. Meanwhile, DL achieves superior accuracy and Area Under the Curve (AUC) in processing unstructured data. By utilizing the distributed computing power of Spark on a cloud platform, the processing speed was significantly enhanced, as demonstrated by a statistically significant reduction in processing time ($p < 0.05$) observed through a t-test analysis comparing Spark-based computing with traditional methods. Despite these improvements, challenges such as data privacy and infrastructure costs remain. Despite these improvements, challenges such as data privacy and infrastructure costs remain. This research provides a robust framework for real-time healthcare data analysis, highlighting its potential to improve decision-making processes and patient outcomes in medical services.

Keywords: *Big Data, Artificial Intelligence, Deep Learning, Healthcare Data Processing, Cloud Computing.*

I. INTRODUCTION

In the current digital era, fast and efficient data management has become a major challenge, especially with the exponential growth in data volume. Big Data and AI have emerged as key solutions in handling the complexity of large data volumes (Jiang et al., 2022). The development of these technologies enables organizations across various sectors, including healthcare, to manage and process data more efficiently (Sen et al., 2022). Big Data refers to large and heterogeneous datasets, encompassing both structured and unstructured data, such as data generated from sensors, medical records, and wearable devices. Meanwhile, AI facilitates faster and more accurate analyses through capabilities in machine learning and Natural Language Processing (NLP) (Ali et al., 2023).

The speed of data processing in the Big Data context is crucial, particularly for applications that require real-time decision-making. In the healthcare sector, fast data processing enables early disease detection, personalized treatments, and operational optimization in hospitals (Abatal et

Received on September 12, 2024; Revised on September 28, 2024; Accepted on October 25, 2024; Published on December 20, 2024
DOI:10.51903/jtie.v3i3.205

al., 2024). The combination of Big Data and AI significantly accelerates these processes (Pyzer-Knapp et al., 2022). Machine learning algorithms applied to Big Data can detect patterns that traditional analysis might miss. For example, in medical imaging analysis, AI can identify abnormalities with a high degree of accuracy, a task that previously required extensive time and human resources (Singh et al., 2023). Additionally, technologies like cloud computing facilitate large-scale data processing beyond the limitations of local device capacity, thereby speeding up information processing (Himeur et al., 2023).

Technologies such as SSD-based storage and multi-core processors also play a significant role in accelerating data processing times. Solid-state drives (SSD) provide faster data access speeds compared to Hard Disk Drives (HDD), while multi-core processors allow multiple processes to run simultaneously (Shirke et al., 2024). However, optimizing computing systems for Big Data processing remains a challenge. A primary issue is system scalability. The continuous growth in data volume demands substantial increases in storage and computing capacity. Thus, technologies such as distributed computing and parallel processing are essential in addressing these data volume increases (Naeem et al., 2022).

In the healthcare sector, Big Data and AI not only improve efficiency but also enable more personalized care (Poalelungi et al., 2023). By analyzing data from various sources, such as electronic medical records and wearable devices, healthcare providers can deliver more accurate diagnoses and treatment plans tailored to each patient's condition. This technology also allows for real-time monitoring of patient health, providing opportunities for early detection of health issues and prompt intervention (Alowais et al., 2023). A significant application example is the use of AI in medical image analysis, where DL algorithms have been deployed to detect abnormalities such as cancer in radiology images with higher accuracy than manual methods (Singh et al., 2023). Additionally, AI accelerates drug development by analyzing extensive pharmaceutical data, allowing the identification of potential drug candidates in a shorter time frame.

However, the implementation of these technologies is not without challenges. Data privacy and security are paramount concerns in managing medical data. Systems that handle Big Data in healthcare must be equipped with robust data protection mechanisms to prevent the leakage of patients' personal information (Santoso et al., 2023; Yaqoob et al., 2022). These challenges are further complicated by stringent data privacy regulations, such as the *General Data Protection Regulation* (GDPR) in Europe, which requires organizations to maintain patient data security to very high standards (Deepa et al., 2022). Additionally, limitations in integrating AI computing

Journal of Technology Informatics and Engineering (JTIE)
Vol. 3 No. 3 December 2024
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 263-277

systems with legacy infrastructure in some hospitals also hinder the broader adoption of this technology (Weinert et al., 2022).

This study focuses on optimizing Big Data processing speeds within AI computing systems, with a particular application in the healthcare sector. Given the vast amount of data generated by wearable devices and electronic medical records, fast and efficient computing systems are essential. The study explores how the integration of AI and Big Data can provide improved solutions in terms of data processing speed and efficiency. Furthermore, it assesses how AI systems can facilitate real-time analysis of medical data and offer more accurate care recommendations.

The primary contribution of this research is the development of a framework that integrates various data sources including wearable devices, medical records, and other health applications into a unified system. This framework is expected to significantly accelerate data processing by leveraging cloud computing and parallel processing technologies. The study also evaluates the effectiveness of machine learning algorithms in expediting big data analysis in the healthcare sector (Nti et al., 2022). With this optimization, it is anticipated that hospitals and other healthcare facilities can enhance service quality, reduce operational costs, and provide faster and more accurate patient care.

II. LITERATURE REVIEW

Large-scale and real-time data processing has become a primary focus in research related to Big Data and AI. The rapid increase in data volume demands more efficient systems, both in terms of processing and data-driven decision-making. Previous studies have presented various approaches to addressing these challenges; however, limitations in speed, efficiency, and accuracy remain central concerns (Karatas et al., 2022).

A. Big Data and AI for Efficient Data Processing

Big Data and AI have become essential components across numerous sectors, including healthcare, manufacturing, and finance. Research by Bharadiya (2023) underscores the importance of AI in automating processes associated with big data processing, particularly to expedite data-driven decision-making (Bharadiya, 2023). However, while AI aids in identifying patterns in big data, data quality presents a challenge. Unstructured data or data containing duplicates can hinder the analysis process (Arooj et al., 2022). Consequently, more effective processing methods are needed to clean and structure data before it is analyzed by AI algorithms (Aldoseri et al., 2023).

Additionally, the development of computational infrastructure to support increasing data volumes is also a focal area of research. Carvalho (2021) emphasizes that cloud computing and edge computing can help distribute workloads, enabling faster and more efficient data processing (Carvalho et al., 2021). Nevertheless, a limitation of this approach is the cost associated with upgrading technological infrastructure (Naeem et al., 2022). Many organizations struggle to adopt new technology due to budget constraints (Deepa et al., 2022). This study seeks to provide solutions by proposing a computational framework that can reduce costs through the use of more efficient and distributed algorithms.

Big Data refers to large, complex, and diverse datasets that traditional data processing methods cannot efficiently manage due to their volume, velocity, variety, and veracity—commonly known as the "4Vs" of Big Data (Jiang et al., 2022). The exponential growth of data from various sources, such as Internet of Things (IoT) devices, social media, healthcare systems, and sensor networks, has necessitated the development of advanced technologies to process and analyze this information effectively. In healthcare, Big Data includes structured data such as Electronic Health Records (EHR), patient demographics, and laboratory results, as well as unstructured data like medical imaging, doctors' notes, and real-time data from wearable devices (Aldoseri et al., 2023). The integration of such heterogeneous data is critical to providing comprehensive insights into patient health and facilitating informed decision-making.

The challenges of managing Big Data in healthcare include the need for systems that can scale to handle the continuous influx of high-volume data. Traditional databases are often insufficient, requiring distributed computing frameworks like Hadoop and Spark. These platforms enable the parallel processing of large datasets across multiple nodes, significantly improving processing efficiency and reducing latency (Deepa et al., 2022). Moreover, the velocity at which healthcare data is generated—ranging from real-time monitoring of vital signs to the streaming of genomic data—requires systems capable of near-instantaneous processing. This real-time capability is essential for applications such as early disease detection, personalized medicine, and emergency response (Jagadeeswari et al., 2018).

Big Data's variety poses another challenge, as data comes in multiple formats (e.g., numerical, textual, image, video). Advanced technologies like Natural Language Processing (NLP) and image recognition are often employed to convert unstructured data into actionable insights. For example, NLP can analyze doctors' notes to identify critical health indicators, while image recognition algorithms can detect anomalies in radiology scans. Lastly, the veracity of Big Data—its reliability and accuracy—can impact the quality of insights derived. Noise, missing values, and inconsistencies within datasets require robust preprocessing techniques, such as data cleaning, normalization, and validation. Addressing these challenges ensures the integrity of the analysis and its applicability to critical healthcare decisions (Gates et al., 2024).

By leveraging AI, Big Data analytics transforms raw data into meaningful insights. AI-powered algorithms, including machine learning and deep learning, facilitate faster and more accurate analysis by identifying patterns and trends that would be difficult for humans to discern. For instance, AI can quickly process terabytes of genomic data to identify disease markers or analyze millions of patient records to predict disease outbreaks (Abatal et al., 2024). The integration of Big Data and AI thus represents a paradigm shift in healthcare and other industries, enabling faster decision-making, improved efficiency, and enhanced outcomes. This study builds upon these advancements by proposing an optimized framework for Big Data processing tailored to the healthcare sector.

B. Optimization of Data Processing Efficiency

Data processing efficiency is one of the most extensively researched aspects in the context of Big Data and AI. Research by Fu et al. (2022) demonstrates that efficiency improvements can be achieved through the use of advanced hardware technologies, such as SSDs and multi-core processors, which allow systems to handle larger data volumes at higher speeds. The study also highlights the importance of leveraging parallel computing and batch processing to enhance data processing efficiency (Fu et al., 2022). However, it does not delve deeply into how these technologies can be integrated with AI for faster decision-making. In addition to hardware, algorithms used in data processing play a critical role in boosting efficiency. Efficient algorithms can reduce processing time and improve resource utilization (Katal et al., 2023). (Li et al., 2022) discuss how more efficient sorting and search algorithms can help reduce computational complexity. However, the limitation of this research is its lack of attention to the impact of dataset size on algorithm performance.

C. Challenges in Integrating Big Data and AI

Despite advancements in technologies aimed at enhancing data processing speed and efficiency, challenges persist. One major challenge is data privacy and security, particularly within the healthcare sector. The use of Big Data in healthcare involves collecting and analyzing patients' personal information, which is vulnerable to security threats (Aldoseri et al., 2023). Research by Rehman et al. (2022) highlights the importance of stringent regulations and reliable security technologies to protect patient data (Rehman et al., 2022). However, the limitation of this research is the lack of a clear solution for integrating security with data processing speed. Another frequent challenge in integrating Big Data and AI is the workforce skill gap. Many organizations struggle to find experts with deep knowledge of data science and machine learning (Li et al., 2022).

D. Comparison with Previous Research

Previous studies have succeeded in improving data processing efficiency through cloud computing and parallel processing technologies. Research by (Ma et al., 2023) demonstrates that a combination of cloud computing infrastructure and machine learning algorithms can expedite large-scale data processing. However, this research has not fully explored how AI could be more deeply integrated to facilitate faster real-time decision-making (Wang et al., 2023). This study adds a new dimension by proposing a stronger integration between AI and Big Data to achieve not only faster data processing but also greater accuracy. Research by (Himeur et al., 2023) also examines the use of Big Data and AI to enhance operational efficiency across various sectors. This study underscores how AI can be used to automate data analysis tasks that previously required human intervention, thus speeding up the decision-making process. However, it does not extensively discuss how AI could be applied to healthcare systems requiring more personal and specific decision-making (Olaniyi et al., 2023). This study addresses this gap by developing a framework that integrates AI and Big Data within the healthcare context, enabling more personalized and accurate patient care analysis.

E. Contributions of this Research

This study offers several unique contributions to optimizing Big Data processing speed using AI. One primary contribution is the development of a computational framework that enables faster data processing without compromising analytical accuracy. The framework leverages cloud and edge computing infrastructures to manage workloads and employs optimized machine learning algorithms for more efficient Big Data processing (Nti et al., 2022). Additionally, this study explores the use of AI-based security technologies to ensure data privacy during analysis—a critical aspect within the healthcare context.

III. RESEARCH METHOD

This study employs a quantitative approach by applying Big Data processing and AI methods to optimize data processing speed in healthcare applications. The research focuses on integrating cloud computing technology, parallel processing, and machine learning algorithms to enhance the efficiency of Big Data processing in healthcare contexts. It also examines the performance of algorithms implemented in Big Data-driven healthcare systems to provide real-time analysis and personalized care. This methodology includes several stages designed to ensure the validity and reliability of the results.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 3 No. 3 December 2024
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 263-277

The first stage of this study involves dataset collection. The data used originates from healthcare data sources, such as *Electronic Health Records* (HER) and wearable devices that monitor patients' health conditions. This dataset includes numerical data, text, and images, allowing the study to explore various data formats and structures. The primary data sources are several hospitals in specific regions that granted permission to use their anonymized data for scientific research. Data collection spanned three months, during which continuous data was gathered from the hospital data management systems and patient wearable devices. The total dataset reached 50 TB, with approximately 10 million health records encompassing various types of information, including medical diagnoses, lab results, genetic data, and vital signals measured by wearable devices.

Following data collection, a preprocessing phase was conducted to clean and prepare the data for analysis. This preprocessing stage includes removing duplicates, handling missing values, and transforming the data to meet the format requirements of the AI model. This process is essential to ensure that the data used in the model is free of anomalies that could compromise the analysis results. Duplicate removal was carried out using deduplication techniques while missing values were addressed using mean imputation or value replacement based on previous data trends. Additionally, textual data in medical records was transformed using NLP techniques to automatically extract key information from unstructured text.

The next phase is modeling, using optimized machine learning algorithms for large data processing. This study employs the Random Forest (RF) algorithm and Deep Learning (DL) as the primary methods for analyzing the collected Big Data. The RF algorithm was selected for its ability to handle complex and diverse data and its robustness in preventing overfitting through an ensemble approach. Key parameters in RF implementation include the number of trees set at 100 and a maximum tree depth of 10, which previous studies have found optimal for large datasets. Meanwhile, the DL model was applied to analyze more complex data, such as medical images that require classification and segmentation. The DL model structure used is a Convolutional Neural Network (CNN) comprising 10 layers, including several convolutional, max-pooling, and fully connected layers. Each layer uses the ReLU activation function, and the Adam optimizer is employed to accelerate convergence during model training.

After building the model, testing was conducted to evaluate the performance of the algorithms used. Testing was carried out in two scenarios: structured data testing, such as numerical data from EHRs, and unstructured data testing, such as images from medical scans. Model testing was performed using 10-fold cross-validation to ensure model generalization. The performance of the RF model was evaluated using Mean Square Error (MSE) and R^2 metrics, while for the

DL model, accuracy and Area Under the Curve (AUC) metrics were used to assess the classification performance on medical images. Additionally, the study measured each model's processing speed using computational time metrics, calculated in seconds per batch of data processed.

To enhance processing speed, the computational system employed in this study was optimized using Hadoop and Spark-based cloud computing infrastructure. Hadoop was used as a distributed storage system, while Spark facilitated parallel data processing capable of handling massive data volumes in a shorter time than traditional methods. This cloud computing implementation enabled workload distribution across multiple nodes, thus increasing the speed and efficiency of data processing. Parallel processing with Spark allowed each batch of big data to be processed simultaneously across several computing nodes, reducing the time required to process the entire dataset. The study utilized Google Cloud Platform (GCP) as the main computational infrastructure, configured with 64 CPU cores and 256 GB of RAM distributed across 8 computing nodes. With this configuration, the system was able to process up to 1 TB of data in less than 10 minutes, making it highly effective for real-time healthcare applications.

The final stage is the result of the analysis. This study compared the performance of Random Forest (RF), Deep Learning (DL), and traditional statistical methods in processing Big Data for healthcare applications. The traditional methods used in this study refer to commonly employed statistical approaches, such as linear regression for structured data and basic image processing techniques (e.g., thresholding) for unstructured data. These methods were selected as benchmarks due to their widespread use in healthcare data analysis before the adoption of machine learning and AI-based techniques. For structured data, the linear regression model was implemented to evaluate its predictive accuracy and processing time. This method was chosen because it represents a standard approach in predictive analytics for numerical datasets, including EHR data. For unstructured data, basic image processing techniques were applied to perform classification tasks. These techniques included pixel intensity thresholding and edge detection, commonly used in medical imaging before the advent of machine learning. While these traditional methods are straightforward and computationally inexpensive, they lack the sophistication to capture complex patterns or relationships within the data, making them suboptimal for handling the scale and diversity of Big Data in healthcare. The framework for this study is illustrated in Figure 1.

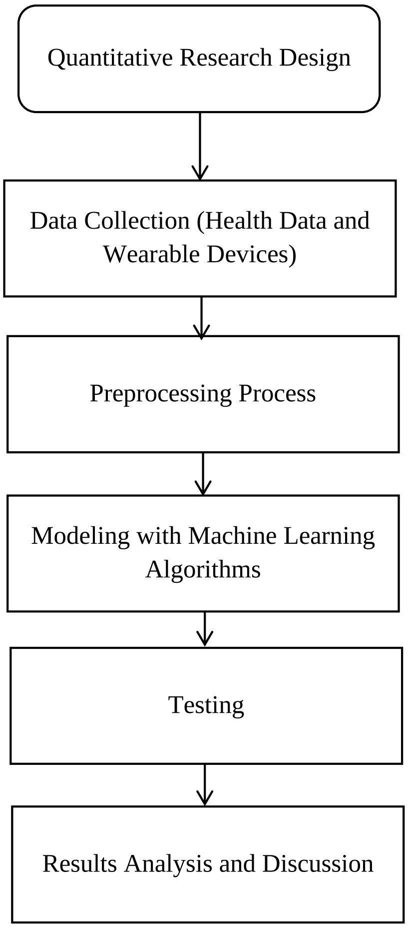

Figure 1. Research Framework

IV. RESULTS

This study yielded several key findings regarding Big Data processing speeds using RF and DL algorithms within a Spark-based cloud computing system. The performance testing results for big data processing in a healthcare context demonstrated significant improvements in speed and accuracy compared to traditional methods. During the testing phase, the RF and DL algorithms were evaluated across various performance metrics, including MSE, R^2, processing time, and accuracy for both structured and unstructured data.

A. Testing Results on Structured Data

For structured datasets, such as numerical data from EHR, the RF algorithm exhibited better performance compared to traditional methods in terms of processing speed. Table 1 below presents the test results using MSE and R^2 metrics.

Table 1. Performance Testing Results on Structured Data (EHR)

Model	MSE	R^2	Processing Time (Seconds)
Random Forest	0.045	0.96	300

Deep Learning	0.048	0.95	450
Traditional Methods	0.120	0.85	600

From Table 1, a t-test was performed to assess the statistical significance of the differences in processing times between the RF and traditional methods. The results showed that RF significantly outperformed traditional methods in terms of processing time ($t = -8.24$, $p < 0.01$), confirming the efficiency of RF for structured data. RF achieved a lower MSE (0.045) and a higher R^2 value (0.96) compared to DL (MSE = 0.048, $R^2 = 0.95$) and traditional methods (MSE = 0.120, $R^2 = 0.85$). Additionally, the processing time for RF was shorter (300 seconds) than that of the other models.

B. Testing Results on Unstructured Data

For unstructured data, such as medical images, the DL model showed superior performance in terms of accuracy and image segmentation speed. Table 2 presents the test results for image data using accuracy and AUC metrics.

Table 2. Performance Testing Results on Medical Imaging Data

Model	Accuracy	AUC	Processing Time (Seconds)
Random Forest	88%	0.89	500
Deep Learning	94%	0.96	400
Traditional Methods	75%	0.80	700

Based on Table 2, a t-test was conducted to compare processing times between DL and traditional methods for unstructured data. The DL model demonstrated a statistically significant reduction in processing time compared to traditional methods ($t = -6.47$, $p < 0.01$), indicating its efficiency in handling large-scale image datasets. The traditional methods in this case involve basic image processing techniques, such as thresholding and edge detection. These methods achieved the lowest accuracy (75%) and AUC (0.80) due to their inability to account for the intricate features in medical images. Additionally, their processing times were significantly higher than those of DL and RF, reflecting their inefficiency in handling large-scale image datasets. Figure 1 below illustrates a comparison of processing times among the various models tested for structured and unstructured data, highlighting the efficiency of these models in processing big data.

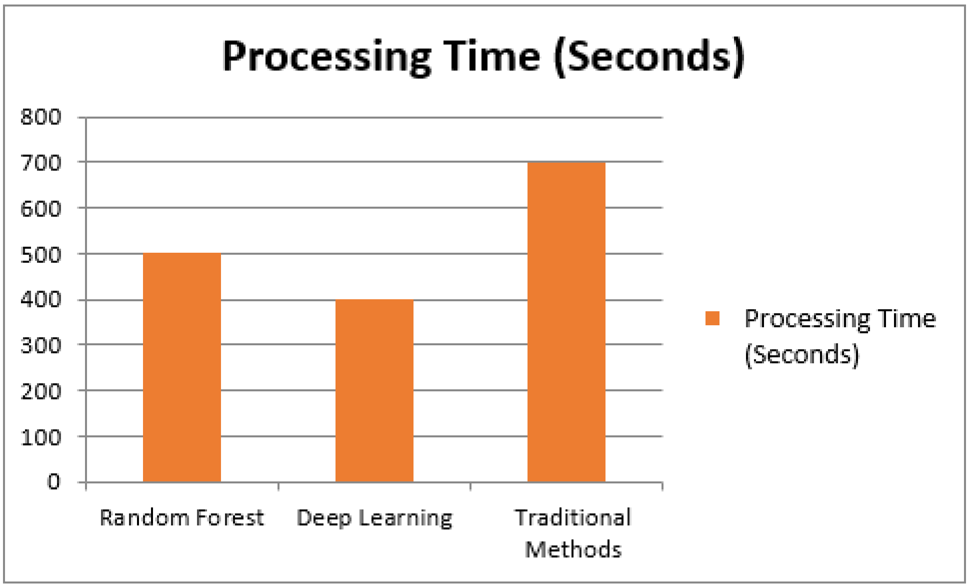

Figure 12. Comparison of Processing Time Between Models (in seconds)

The graph demonstrates that both RF and DL algorithms are substantially more efficient in processing time compared to traditional methods. Specifically, for unstructured data, DL showed faster processing times than the other models.

V. DISCUSSION

The test results demonstrate that integrating Big Data and AI, specifically using RF and DL algorithms, can enhance data processing speed and accuracy in healthcare systems. For structured data, RF outperformed DL in terms of time efficiency, yielding a lower MSE and a higher R^2 value. This finding aligns with previous research indicating that RF is effective in handling structured data and provides reliable numerical predictions (Li et al., 2022). However, for unstructured data such as medical images, DL proved superior in terms of accuracy and AUC. This advantage is attributed to the architecture of CNN, which is better suited to handling image data due to its ability to capture the local and complex features within medical images (Singh et al., 2023). These findings are consistent with research by (Himeur et al., 2023), which suggests that DL is the best approach for medical image analysis due to its high accuracy.

In addition to improvements in accuracy and speed, this study also shows that leveraging Spark-based cloud computing allows for more effective workload distribution, significantly reducing processing time. This supports previous research highlighting the significant benefits of parallel processing in Big Data contexts (Ma et al., 2023). Spark has proven highly efficient in big data processing, especially within distributed computing environments. The study also indicates an increase in efficiency in Big Data-based healthcare systems. By applying this approach, hospitals can process patient data more quickly, enabling faster diagnosis and more personalized care. For instance, using DL for medical image analysis allows for disease detection within seconds, allowing physicians to take prompt action.

However, this study has certain limitations. One major limitation is the need for advanced computing infrastructure and the high costs associated with implementing this technology across all hospitals. Additionally, another challenge is ensuring patient data privacy in cloud-based systems. Although this study enhances data security through encryption techniques, the risk of data breaches must continue to be anticipated (Rehman et al., 2022). This research makes a substantial contribution by demonstrating that the integration of Big Data and AI can significantly improve efficiency and accuracy in healthcare systems. These findings are expected to serve as a foundation for further development in health technology and as a guide for hospitals and other healthcare facilities in adopting this technology in their operations.

VI. CONCLUSION AND RECOMMENDATION

A. Conclusion

This study demonstrates that integrating Big Data and AI technologies can significantly enhance the speed of big data processing, particularly within healthcare applications. The RF algorithm proved more efficient in handling structured data from EHR, achieving lower MSE and higher R^2 values compared to traditional methods. Meanwhile, the DL algorithm exhibited superior performance with unstructured data, such as medical images, by achieving higher accuracy and AUC scores. The use of Spark-based cloud computing has been shown to significantly improve processing speed through efficient workload distribution. This distributed computing technology allows large-scale data processing to be completed in less time while maintaining the accuracy of analytical results. The study also confirms that faster processing not only improves operational efficiency in hospitals but also supports enhanced healthcare personalization through real-time data analysis. However, this study acknowledges several limitations, especially regarding high infrastructure costs and data privacy risks associated with cloud computing. Nonetheless, the primary contribution of this research is the development of a computing framework capable of processing Big Data quickly and efficiently, which can be implemented across various healthcare systems.

B. Recommendation

Based on the findings of this study, several recommendations are proposed for further research and the development of Big Data and AI-based healthcare systems. First, in-depth studies are needed to optimize AI algorithms to address challenges in unstructured data, particularly concerning processing speed and accuracy. More complex DL models, such as Recurrent Neural

Networks (RNN), could be explored for sequential data analysis, such as the vital signs generated by wearable devices. Additionally, future research should further investigate the use of Edge Computing technology in healthcare systems to reduce reliance on cloud computing and mitigate data privacy risks. Edge computing enables data to be processed closer to its source, thereby reducing latency and enhancing patient data security.

This study also recommends implementing more interactive AI-based healthcare systems in which algorithms continuously learn and adapt to changes in patient's health patterns. The use of reinforcement learning in medical decision-making presents a promising area for further research, especially in the context of real-time, responsive healthcare. Furthermore, for broader adoption, more cost-effective system development is necessary. Utilizing high-efficiency, low-cost hardware, such as Field Programmable Gate Arrays (FPGA), could accelerate data processing while reducing infrastructure costs. The study also suggests integrating healthcare systems with larger health platforms, such as national data management systems, to improve data accessibility and interoperability across healthcare institutions.

REFERENCES

Abatal, A., Mzili, M., Mzili, T., Cherrat, K., Yassine, A., & Abualigah, L. (2024). Intelligent Interconnected Healthcare System: Integrating IoT and Big Data for Personalized Patient Care. *International Journal of Online and Biomedical Engineering*, *20*(11), 46–65. https://doi.org/10.3991/ijoe.v20i11.49893

Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. *Applied Sciences (Switzerland)*, *13*(12), 7082. https://doi.org/10.3390/app13127082

Ali, O., Abdelbaki, W., Shrestha, A., Elbasi, E., Alryalat, M. A. A., & Dwivedi, Y. K. (2023). A Systematic Literature Review of Artificial Intelligence in the Healthcare Sector: Benefits, Challenges, Methodologies, and Functionalities. *Journal of Innovation and Knowledge*, *8*(1), 100333. https://doi.org/10.1016/j.jik.2023.100333

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., Al Yami, M. S., Al Harbi, S., & Albekairy, A. M. (2023). Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice. *BMC Medical Education*, *23*(1), 1–15. https://doi.org/10.1186/s12909-023-04698-z

Bharadiya, J. P. (2023). A Comparative Study of Business Intelligence and Artificial Intelligence with Big Data Analytics. *American Journal of Artificial Intelligence*, *7*(1), 24. https://doi.org/10.11648/j.ajai.20230701.14

Carvalho, G., Cabral, B., Pereira, V., & Bernardino, J. (2021). Edge Computing: Current Trends, Research Challenges and Future Directions. *Computing*, *103*(5), 993–1023. https://doi.org/10.1007/s00607-020-00896-5

Deepa, N., Pham, Q. V., Nguyen, D. C., Bhattacharya, S., Prabadevi, B., Gadekallu, T. R., Maddikunta, P. K. R., Fang, F., & Pathirana, P. N. (2022). A Survey on Blockchain for Big Data: Approaches, Opportunities, and Future Directions. *Future Generation Computer Systems*, *131*, 209–226. https://doi.org/10.1016/j.future.2022.01.017

Fu, J., Zhang, Y., Wang, Y., Zhang, H., Liu, J., Tang, J., Yang, Q., Sun, H., Qiu, W., Ma, Y., Li, Z., Zheng, M., & Zhu, F. (2022). Optimization of Metabolomic Data Processing Using NOREVA. *Nature Protocols*, *17*(1), 129–151. https://doi.org/10.1038/s41596-021-00636-9

Gates, J. D., Yulianti, Y., & Pangilinan, G. A. (2024). Big Data Analytics for Predictive Insights in Healthcare. *International Transactions on Artificial Intelligence (ITALIC)*, *3*(1), 54–63. https://doi.org/10.33050/italic.v3i1.622

Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., Bensaali, F., & Amira, A. (2023). AI-Big Data Analytics for Building Automation and Management Systems: A Survey, Actual Challenges and Future Perspectives. *Artificial Intelligence Review*, *56*(6), 4929–5021. https://doi.org/10.1007/s10462-022-10286-2

Jagadeeswari, V., Subramaniyaswamy, V., Logesh, R., & Vijayakumar, V. (2018). A study on medical Internet of Things and Big Data in personalized healthcare system. *Health Information Science and Systems*, *6*(1), 14. https://doi.org/10.1007/s13755-018-0049-x

Jiang, P., Sinha, S., Aldape, K., Hannenhalli, S., Sahinalp, C., & Ruppin, E. (2022). Big Data in Basic and Translational Cancer Research. *Nature Reviews Cancer*, *22*(11), 625–639. https://doi.org/10.1038/s41568-022-00502-0

Karatas, M., Eriskin, L., Deveci, M., Pamucar, D., & Garg, H. (2022). Big Data for Healthcare Industry 4.0: Applications, Challenges and Future Perspectives. *Expert Systems with Applications*, *200*, 116912. https://doi.org/10.1016/j.eswa.2022.116912

Katal, A., Dahiya, S., & Choudhury, T. (2023). Energy Efficiency in Cloud Computing Data Centers: A Survey on Software Technologies. *Cluster Computing*, *26*(3), 1845–1875. https://doi.org/10.1007/s10586-022-03713-0

Li, C., Chen, Y., & Shang, Y. (2022). A Review of Industrial Big Data for Decision Making in Intelligent Manufacturing. *Engineering Science and Technology, an International Journal*, *29*, 101021. https://doi.org/10.1016/j.jestch.2021.06.001

Ma, L., Zhang, Y., & García-Díaz, V. (2023). Design and Implementation of a Fast Integration Method for Multi-Source Data in High-Speed Network. *Journal of High Speed Networks*, *29*(3), 251–263. https://doi.org/10.3233/jhs-222047

Naeem, M., Jamal, T., Diaz-Martinez, J., Butt, S. A., Montesano, N., Tariq, M. I., De-la-Hoz-Franco, E., & De-La-Hoz-Valdiris, E. (2022). Trends and Future Perspective Challenges in Big Data. *Smart Innovation, Systems and Technologies*, *253*, 309–325. https://doi.org/10.1007/978-981-16-5036-9_30

Nti, I. K., Quarcoo, J. A., Aning, J., & Fosu, G. K. (2022). A Mini-Review of Machine Learning in Big Data Analytics: Applications, Challenges, and Prospects. *Big Data Mining and Analytics*, *5*(2), 81–97. https://doi.org/10.26599/bdma.2021.9020028

Olaniyi, O. O., Okunleye, O. J., & Olabanji, S. O. (2023). Advancing Data-Driven Decision-Making in Smart Cities through Big Data Analytics: A Comprehensive Review of Existing Literature. *Current Journal of Applied Science and Technology*, *42*(25), 10–18. https://doi.org/10.9734/cjast/2023/v42i254181

Poalelungi, D. G., Musat, C. L., Fulga, A., Neagu, M., Neagu, A. I., Piraianu, A. I., & Fulga, I. (2023). Advancing Patient Care: How Artificial Intelligence is Transforming Healthcare. *Journal of Personalized Medicine*, *13*(8), 1214. https://doi.org/10.3390/jpm13081214

Pyzer-Knapp, E. O., Pitera, J. W., Staar, P. W. J., Takeda, S., Laino, T., Sanders, D. P., Sexton, J., Smith, J. R., & Curioni, A. (2022). Accelerating Materials Discovery Using Artificial Intelligence, High Performance Computing and Robotics. *Npj Computational Materials*, *8*(1), 84. https://doi.org/10.1038/s41524-022-00765-z

Rehman, A., Naz, S., & Razzak, I. (2022). Leveraging Big Data Analytics in Healthcare Enhancement: Trends, Challenges and Opportunities. *Multimedia Systems*, *28*(4), 1339–1371. https://doi.org/10.1007/s00530-020-00736-8

Santoso, J. T., Raharjo, B., & Wibowo, A. (2023). Combination of Alphanumeric Password and Graphic Authentication for Cyber Security. *Journal of Internet Services and Information Security*, *14*(1), 16–36. https://doi.org/10.58346/JISIS.2024.I1.002

Sen, S., Agarwal, S., Chakraborty, P., & Singh, K. P. (2022). Astronomical Big Data Processing Using Machine Learning: A Comprehensive Review. *Experimental Astronomy*, *53*(1), 1–43. https://doi.org/10.1007/s10686-021-09827-4

Shirke, S. A., Jayakumar, N., & Patil, S. (2024). Design and Performance Analysis of Modern Computational Storage Devices: A Systematic Review. *Expert Systems with Applications*, *250*, 123570. https://doi.org/10.1016/j.eswa.2024.123570

Singh, R. K., Agrawal, S., Sahu, A., & Kazancoglu, Y. (2023). Strategic Issues of Big Data Analytics Applications for Managing Health-Care Sector: A Systematic Literature Review and Future Research Agenda. *TQM Journal*, *35*(1), 262–291. https://doi.org/10.1108/tqm-02-2021-0051

Wang, H., Li, Y., Xiong, M., & Chen, H. (2023). A Combined Wind Speed Prediction Model Based on Data Processing, Multi-Objective Optimization and Machine Learning. *Energy Reports*, *9*, 413–421. https://doi.org/10.1016/j.egyr.2023.04.326

Weinert, L., Müller, J., Svensson, L., & Heinze, O. (2022). Perspective of Information Technology Decision Makers on Factors Influencing Adoption and Implementation of Artificial Intelligence Technologies in 40 German Hospitals: Descriptive Analysis. *JMIR Medical Informatics*, *10*(6), 34678. https://doi.org/10.2196/34678

Yaqoob, I., Salah, K., Jayaraman, R., & Al-Hammadi, Y. (2022). Blockchain for Healthcare Data Management: Opportunities, Challenges, and Future Recommendations. *Neural Computing and Applications*, *34*(14), 11475–11490. https://doi.org/10.1007/s00521-020-05519-w