



Scalable and Efficient Data Mining Techniques for Big Data Analytics in Modern Computing

Juty Roy ¹, Delwar Karim ², Jhon Kabir ^{3*}, Rashid Khan ⁴

¹ Department of Pharmacy, Bangladesh University, Dhaka, Bangladesh

² Independent Researcher, Khulna, Bangladesh

³ Department of Pharmacy, Bangladesh University, Dhaka, Bangladesh

⁴ Independent Researcher, Dhaka, Bangladesh

* Corresponding Author: **Jhon Kabir**

Article Info

P-ISSN: 3051-3502

E-ISSN: 3051-3510

Volume: 06

Issue: 01

January – June 2025

Received: 29-12-2024

Accepted: 30-01-2025

Published: 24-02-2025

Page No: 52-58

Abstract

The rapid expansion of big data in modern computing environments has created unprecedented opportunities and challenges in data analysis and knowledge discovery. Traditional data mining techniques often fail to address the scalability and efficiency requirements of large-scale datasets characterized by high volume, velocity, and variety. This study presents an in-depth exploration of scalable and efficient data mining techniques for big data analytics, focusing on modern computational frameworks and advanced machine learning approaches. The integration of distributed computing, parallel processing, and hybrid machine learning models has significantly enhanced the ability to process and analyze large datasets in real time. Additionally, optimization strategies such as dimensionality reduction, feature selection, and model tuning have improved computational efficiency and predictive performance. The findings highlight that scalable data mining frameworks, including cloud-based and distributed architectures, play a critical role in enabling efficient big data analytics. However, challenges related to data heterogeneity, computational complexity, and data privacy remain significant concerns. This study emphasizes the need for adaptive, efficient, and scalable solutions to address the evolving demands of big data environments and supports the development of intelligent systems capable of real-time decision-making.

DOI: <https://doi.org/10.54660/IJMER.2025.6.1.52-58>

Keywords: Big Data Analytics, Data Mining Techniques, Scalable Computing, Machine Learning Optimization, Distributed Data Processing

1. Introduction

The exponential growth of data generated by modern technologies has transformed the landscape of computing and information systems. The emergence of big data, characterized by its volume, velocity, and variety, has created new opportunities for organizations to extract valuable insights and make informed decisions (Orthi *et al.*, 2025; Uddin *et al.*, 2025). Data is continuously generated from various sources, including social media platforms, IoT devices, healthcare systems, financial transactions, and industrial processes. While this vast amount of data offers significant potential, it also presents challenges in terms of storage, processing, and analysis (Alam *et al.*, 2025; Juie *et al.*, 2021). Data mining plays a crucial role in extracting meaningful information from large datasets. It involves the use of algorithms and techniques to identify patterns, relationships, and trends within data. Traditional data mining techniques, such as classification, clustering, and association rule mining, have been widely used for data analysis. However, these techniques were primarily designed for small to medium-sized datasets and often struggle to handle the complexity and scale of big data (Sami *et al.*, 2025).

One of the major challenges in big data analytics is scalability. As the size of datasets increases, traditional algorithms become inefficient due to limitations in memory and computational resources (Hemal *et al.*, 2025). Processing large datasets requires significant computational power and efficient algorithms capable of handling distributed data. Additionally, the dynamic nature of big data requires real-time processing capabilities, which traditional batch processing methods cannot provide (Sikder *et al.*, 2023). To address these challenges, modern computing technologies have introduced scalable data mining techniques that leverage distributed computing and parallel processing. Frameworks such as Hadoop and Apache Spark enable large-scale data processing by distributing tasks across multiple nodes. These frameworks provide a scalable environment for data mining, allowing algorithms to process large datasets efficiently (Sikder *et al.*, 2025; Vanu *et al.*, 2021). Machine learning and artificial intelligence have further enhanced data mining capabilities. Advanced algorithms, including deep learning models, can automatically extract features from data and improve prediction accuracy. Hybrid approaches that combine multiple algorithms have also been developed to enhance performance and efficiency (Alam *et al.*, 2025). Despite these advancements, several challenges remain, including data heterogeneity, computational complexity, and data privacy concerns. This study aims to explore scalable and efficient data mining techniques for big data analytics and identify potential solutions to these challenges.

2. Scalable Data Mining Techniques

Scalability is a fundamental requirement for data mining in big data environments. Traditional algorithms often fail to process large datasets efficiently due to memory constraints and computational limitations. To overcome these challenges, distributed and parallel data mining techniques

have been developed. Distributed data mining involves partitioning data across multiple nodes and processing it simultaneously. Frameworks such as Hadoop use the MapReduce programming model to perform distributed computations. This approach divides data processing tasks into smaller sub-tasks that can be executed in parallel, significantly improving scalability and efficiency (Sikder *et al.*, 2025).

Parallel data mining techniques further enhance scalability by utilizing multi-core processors and distributed systems. These techniques allow multiple computations to be performed simultaneously, reducing processing time and improving performance. Parallel algorithms are particularly effective for tasks such as clustering and classification, where large datasets can be processed concurrently (Alam *et al.*, 2025). Another important approach is stream data mining, which focuses on processing continuous data streams in real time. Unlike batch processing, stream mining algorithms update models incrementally as new data arrives. This approach is essential for applications such as fraud detection, network monitoring, and IoT systems, where real-time analysis is required (Sikder *et al.*, 2023).

Scalability is a fundamental requirement for data mining in modern big data environments. The exponential growth of data generated from cloud platforms, IoT devices, social media, healthcare systems, and enterprise applications has made traditional data mining algorithms insufficient for handling massive datasets efficiently. Conventional mining approaches often suffer from memory limitations, increased computational costs, and long processing times when dealing with high-volume and high-velocity data (Figure 1). To address these issues, scalable data mining techniques based on distributed computing, parallel processing, and real-time analytics have been developed.

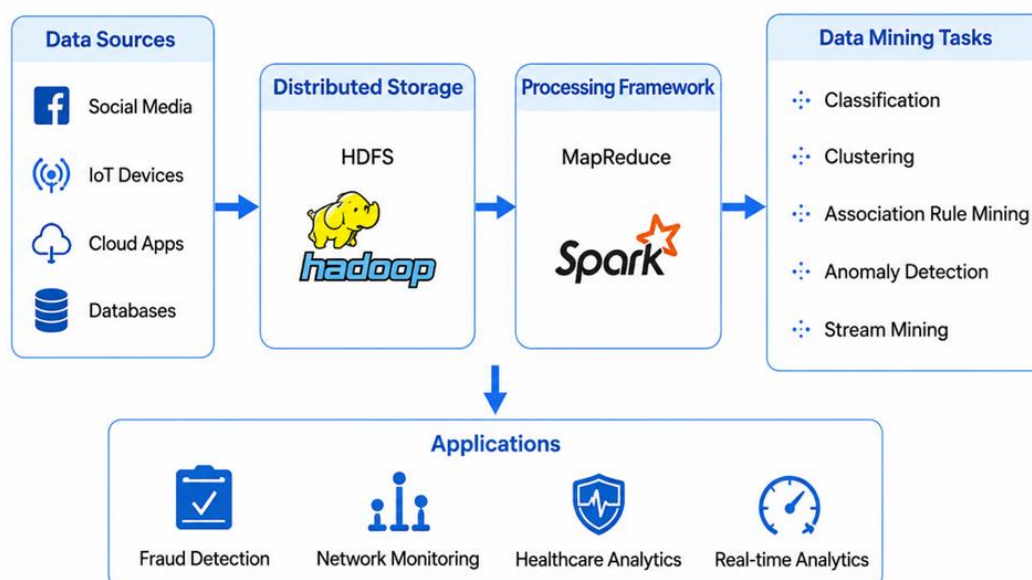


Fig 1: Scalable Data Mining Architecture for Big Data Analytics Using Distributed, Parallel, and Stream Processing Techniques.

One of the most widely adopted approaches is distributed data mining, where datasets are partitioned across multiple computing nodes and processed simultaneously. Distributed frameworks improve scalability by dividing large

computational tasks into smaller subtasks that can be executed independently across clusters of machines. Technologies such as Apache Hadoop utilize the MapReduce programming model, which distributes data processing

operations into map and reduce phases for parallel execution (Dean & Ghemawat, 2008). This approach significantly improves computational efficiency, fault tolerance, and storage scalability. Distributed mining frameworks are especially useful for analyzing extremely large datasets in cloud computing and enterprise-level applications.

The Hadoop Distributed File System (HDFS) further enhances scalability by storing data redundantly across multiple nodes, ensuring reliability and high availability. In addition, modern distributed frameworks such as Apache Spark improve processing speed through in-memory computation and iterative processing capabilities. Spark is particularly effective for machine learning, graph analytics, and real-time stream processing because it reduces disk input/output operations and accelerates large-scale analytics tasks (Zaharia *et al.*, 2016).

Another important scalable approach is parallel data mining, which utilizes multi-core processors, graphics processing units (GPUs), and distributed systems to execute multiple computations simultaneously. Parallel algorithms reduce execution time by dividing workloads among several processing units operating concurrently. These techniques are highly effective for computationally intensive tasks such as clustering, classification, association rule mining, and neural network training. For example, parallel implementations of k-means clustering and decision tree algorithms significantly reduce computation time while maintaining analytical accuracy (Alam *et al.*, 2025). GPU-based parallel processing has also become increasingly important for accelerating deep learning and artificial intelligence applications in big data environments.

Scalable mining techniques are also closely associated with cloud computing infrastructures, which provide flexible and on-demand computational resources. Cloud platforms enable organizations to scale storage and processing capabilities dynamically according to workload requirements. This elasticity reduces infrastructure costs and improves accessibility for large-scale analytics applications. Cloud-based distributed systems also support fault tolerance, resource virtualization, and remote accessibility, making them suitable for handling continuously growing datasets.

Another significant development is stream data mining, which focuses on processing continuous and rapidly generated data streams in real time. Unlike traditional batch-processing approaches, stream mining algorithms incrementally update models as new data arrives, eliminating the need to repeatedly process the entire dataset. This capability is essential for applications requiring immediate analysis and rapid decision-making, such as fraud detection, financial transaction monitoring, cybersecurity, healthcare monitoring systems, and IoT networks (Sikder *et al.*, 2023). Stream data mining frameworks often use sliding window models, online learning algorithms, and adaptive clustering techniques to manage continuously evolving data. Technologies such as Apache Storm, Apache Flink, and Spark Streaming support scalable real-time analytics by processing incoming data with low latency. These frameworks are capable of handling high-speed data streams while maintaining system performance and scalability.

Scalability in data mining is further improved through load balancing and resource optimization techniques. Efficient scheduling algorithms distribute workloads evenly across computing nodes, minimizing bottlenecks and improving overall system performance. Data compression and indexing

methods also help reduce storage requirements and improve query execution speed in large-scale mining systems.

In recent years, the integration of artificial intelligence and scalable mining architectures has enabled more sophisticated analytics capabilities. Distributed deep learning frameworks allow large neural networks to be trained across multiple GPUs and cloud-based clusters. Federated learning has also emerged as a scalable approach where models are trained locally on distributed devices without transferring raw data, improving both scalability and privacy.

Overall, scalable data mining techniques are essential for modern big data analytics because they enable organizations to process massive and continuously growing datasets efficiently. Distributed computing, parallel processing, stream mining, and cloud-based architectures collectively improve computational speed, scalability, reliability, and real-time analytical performance in modern computing environments.

3. Efficient Data Mining Techniques

Efficiency in data mining is achieved by optimizing computational resources, reducing execution time, and improving algorithmic performance when handling large-scale datasets. In modern big data environments, organizations process structured, semi-structured, and unstructured data generated from cloud systems, IoT devices, social media platforms, and enterprise applications. As data volume and complexity increase, efficient data mining techniques become essential for ensuring scalability, accuracy, and real-time analytics capabilities.

One of the most widely used strategies for improving efficiency is dimensionality reduction, which decreases the number of variables in a dataset while preserving significant information. High-dimensional datasets often increase computational complexity, memory consumption, and processing time. Techniques such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and feature selection methods help eliminate redundant or irrelevant attributes, thereby improving mining speed and predictive performance (Alam *et al.*, 2023). PCA transforms correlated variables into a smaller number of uncorrelated principal components, reducing storage requirements and accelerating machine learning computations. Similarly, feature selection approaches such as wrapper methods, filter methods, and embedded methods improve efficiency by selecting only the most relevant features for analysis.

Another important approach is the use of hybrid data mining models, which combine multiple algorithms to exploit their strengths while minimizing weaknesses. Hybrid techniques enhance scalability, classification accuracy, and processing efficiency in complex big data environments. For instance, integrating clustering algorithms with classification models reduces the search space and improves prediction speed (Sami *et al.*, 2024). Combining association rule mining with neural networks or integrating fuzzy logic with decision trees has also shown significant improvements in handling uncertain and heterogeneous data. Hybrid frameworks are particularly useful in healthcare analytics, cybersecurity, recommendation systems, and financial fraud detection where massive and diverse datasets must be analyzed efficiently.

The adoption of parallel and distributed computing frameworks has further improved the efficiency of data mining processes. Technologies such as Apache Hadoop,

MapReduce, and Apache Spark distribute computational tasks across multiple nodes, enabling faster processing of large datasets. Hadoop provides scalable storage and batch processing capabilities, while Spark improves efficiency through in-memory computation and real-time processing (Dean & Ghemawat, 2008). Distributed mining approaches reduce computational bottlenecks and improve fault tolerance, making them highly suitable for modern cloud-based big data infrastructures.

Deep learning models have also contributed significantly to efficient data mining by automating feature extraction and learning complex data representations. Algorithms such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are widely used in image recognition, speech analysis, and predictive analytics. However, deep learning methods often require extensive computational resources and training time. To address these challenges, optimization techniques such as model pruning, parameter tuning, dropout regularization, and quantization are applied to reduce model complexity and improve efficiency without sacrificing accuracy (LeCun *et al.*, 2015). Recent advances in GPU acceleration and tensor processing units (TPUs) have also enhanced the scalability of deep learning-based mining systems.

In addition, incremental and stream data mining techniques are becoming increasingly important in real-time analytics applications. Unlike traditional batch processing methods, stream mining algorithms continuously process incoming data in real time, enabling rapid decision-making in applications such as smart cities, online recommendation systems, and IoT monitoring. Algorithms such as Hoeffding Trees and online clustering methods efficiently update mining models without reprocessing the entire dataset, thereby reducing computational overhead.

Efficient data mining also depends on effective data preprocessing and data cleaning techniques. Removing noisy, inconsistent, and duplicate data improves mining accuracy and reduces unnecessary computation. Data normalization, discretization, and sampling techniques help minimize storage requirements and accelerate algorithm execution. Furthermore, the integration of cloud computing and edge computing technologies enables scalable and low-latency data processing, improving overall mining efficiency in distributed environments.

Overall, efficient data mining techniques play a crucial role in modern big data analytics by improving scalability, reducing computational costs, and enhancing analytical performance. The combination of dimensionality reduction, hybrid models, distributed computing, deep learning optimization, and real-time stream processing provides a strong foundation for handling the growing challenges of big data in modern computing systems.

4. Challenges in Big Data Analytics

Despite rapid advancements in scalable architectures and efficient data mining techniques, big data analytics continues to face several significant challenges. The increasing volume, velocity, variety, and complexity of data generated from modern digital systems create difficulties in data management, processing, security, and analysis. These challenges affect the performance, reliability, and scalability of analytics systems across industries such as healthcare, finance, education, transportation, and e-commerce. Despite advancements in scalable and efficient techniques, several

challenges remain in big data analytics. Data heterogeneity is a major issue, as data is often collected from multiple sources in different formats. Integrating and processing such data requires advanced preprocessing techniques (Alam *et al.*, 2024). Computational complexity is another challenge, particularly for deep learning models that require significant resources. Balancing accuracy and efficiency remain a key concern (Sikder *et al.*, 2025). Data privacy and security are also critical issues, as large datasets often contain sensitive information. Ensuring secure data processing is essential for maintaining trust and compliance (Sami *et al.*, 2025).

One of the primary challenges is data heterogeneity. Big data is collected from multiple heterogeneous sources including social media platforms, IoT sensors, cloud applications, enterprise systems, mobile devices, and multimedia content. These datasets may exist in structured, semi-structured, or unstructured formats, making integration and processing highly complex. Inconsistent data schemas, missing values, duplicate records, and incompatible formats complicate data preprocessing activities and reduce analytical efficiency. Effective data integration therefore requires advanced preprocessing techniques such as data cleaning, transformation, normalization, and semantic integration (Jagadish *et al.*, 2014). Furthermore, handling heterogeneous data in real time remains difficult because different data types often require specialized storage and processing mechanisms. Another major challenge is computational complexity and scalability. Modern analytics applications frequently process massive datasets containing billions of records and high-dimensional features. Machine learning and deep learning algorithms, especially neural networks and transformer-based models, require substantial computational resources, memory, and storage capacity. Training such models can consume significant time and energy, particularly when real-time analytics is required. Although distributed computing frameworks such as Hadoop and Apache Spark improve scalability through parallel processing, balancing computational efficiency with prediction accuracy remains a critical issue (Hashem *et al.*, 2015). Organizations must optimize hardware resources, algorithm design, and workload distribution to minimize processing delays and operational costs.

Data privacy and security are also among the most critical concerns in big data analytics. Large-scale datasets often contain sensitive personal, financial, medical, or organizational information. Unauthorized access, cyberattacks, insider threats, and data breaches can compromise confidentiality and lead to severe legal and reputational consequences. Since many analytics systems rely on cloud-based infrastructures and distributed environments, ensuring secure data transmission and storage becomes increasingly challenging. Advanced encryption methods, authentication protocols, access control mechanisms, and privacy-preserving data mining techniques are essential for protecting sensitive information (Tankard, 2012). In addition, compliance with regulatory frameworks such as GDPR and HIPAA requires organizations to implement strict data governance and privacy policies.

The demand for real-time analytics and low-latency processing presents additional difficulties. Applications such as autonomous vehicles, financial trading systems, smart healthcare monitoring, and cybersecurity platforms require immediate analysis of continuously generated streaming data. Traditional batch-processing methods are often unable

to meet these low-latency requirements. Real-time big data analytics therefore depends on stream-processing frameworks, edge computing technologies, and efficient online learning algorithms capable of processing incoming data rapidly and continuously (Marz & Warren, 2015).

Overall, the challenges in big data analytics are multidimensional and interconnected. Addressing these issues requires the integration of scalable computing frameworks, secure architectures, efficient preprocessing techniques, ethical AI principles, and advanced resource optimization strategies to ensure reliable, efficient, and sustainable big data analytics systems.

5. Limitations

The development of scalable and efficient data mining techniques has significantly enhanced the ability to process and analyze large-scale datasets in modern computing environments. However, despite these advancements, several limitations persist that hinder the full potential of big data analytics. Identifying these limitations is essential for improving existing methodologies and guiding future research toward more robust and efficient solutions. One of the primary limitations of current data mining techniques is the challenge of handling data heterogeneity. Big data is often collected from multiple sources, including structured, semi-structured, and unstructured formats such as text, images, and sensor data. Integrating and processing such diverse data types require complex preprocessing techniques, which can increase computational overhead and reduce system efficiency (Alam *et al.*, 2023; Sami *et al.*, 2024).

Another significant limitation is scalability. Although distributed computing frameworks such as Hadoop and Spark have improved the ability to process large datasets, they still face challenges in handling extremely high-dimensional data. As data volume increases, the computational cost and memory requirements also grow, making it difficult to maintain performance. This issue is particularly critical in real-time applications where timely data processing is essential (Sikder *et al.*, 2025). Computational complexity is also a major concern in big data mining. Advanced machine learning and deep learning models, while highly accurate, require substantial computational resources for training and deployment. This can lead to increased processing time and energy consumption, limiting their applicability in resource-constrained environments. Furthermore, optimization techniques used to improve model performance often add additional layers of complexity, further increasing computational demands (Sikder *et al.*, 2025).

Data quality issues present another critical limitation. Big data often contains noise, missing values, and inconsistencies, which can negatively impact model accuracy and reliability. Data preprocessing techniques such as cleaning, normalization, and feature selection are required to address these issues; however, these processes can be time-consuming and may not always guarantee optimal results (Alam *et al.*, 2024). Additionally, data privacy and security concerns remain significant challenges. Big data analytics often involves sensitive information, particularly in domains such as healthcare and finance (Nusrat *et al.*, 2024). Ensuring data security while maintaining analytical efficiency is a complex task, requiring advanced encryption and privacy-preserving techniques (Sami *et al.*, 2025). Another limitation is the lack of interpretability in advanced data mining models. Many machine learning and deep learning techniques

operate as “black-box” systems, making it difficult to understand how decisions are made. This lack of transparency can reduce trust and hinder the adoption of data mining techniques in critical applications where explainability is required (Alam *et al.*, 2025).

Finally, real-time processing remains a challenge in big data analytics. Although stream processing techniques have been developed, handling high-velocity data streams efficiently while maintaining accuracy is still difficult. Existing systems may struggle to process data in real time without compromising performance, particularly in applications such as IoT and smart systems (Sikder *et al.*, 2023).

6. Future Directions

To overcome these limitations, several promising research directions can be explored to enhance the scalability and efficiency of data mining techniques. One important direction is the development of advanced data integration techniques capable of handling heterogeneous data sources. Future research should focus on designing unified frameworks that can efficiently process structured and unstructured data without requiring extensive preprocessing. The use of artificial intelligence for automated data integration can significantly improve efficiency and reduce computational overhead (Alam *et al.*, 2024, 2025). Another key area of research is the improvement of scalability through distributed and parallel computing. Future data mining systems should leverage cloud computing, edge computing, and high-performance computing architectures to handle large-scale datasets efficiently. The integration of these technologies can enable real-time data processing and improve system performance (Sikder *et al.*, 2025).

Reducing computational complexity is also a critical research direction. Techniques such as model pruning, quantization, and knowledge distillation can help reduce the size and complexity of machine learning models, making them more efficient and suitable for real-time applications. Additionally, the development of lightweight algorithms can improve performance in resource-constrained environments (Sami *et al.*, 2024). The use of hybrid and adaptive data mining models represents another promising approach. Hybrid models combine multiple techniques to leverage their strengths and improve performance. For example, integrating clustering with classification or combining machine learning with deep learning can enhance both accuracy and efficiency. Adaptive models that can adjust to changing data patterns are also essential for dynamic environments (Alam *et al.*, 2025; Sikder *et al.*, 2021). Enhancing data quality through advanced preprocessing techniques is another important research area. Automated data cleaning and feature selection methods can improve the reliability of data mining results while reducing manual effort. Machine learning-based preprocessing techniques can also be used to identify and correct data inconsistencies (Alam *et al.*, 2024).

Data privacy and security can be improved through the use of privacy-preserving data mining techniques such as federated learning and secure multi-party computation. These approaches allow data to be analyzed without being shared, ensuring confidentiality while maintaining analytical performance (Sami *et al.*, 2025). Improving model interpretability is also essential for increasing trust in data mining systems. Explainable AI (XAI) techniques can provide insights into model decisions, enabling users to understand and validate results. This is particularly important

in critical applications where transparency is required (Alam *et al.*, 2025). Finally, future research should focus on real-time data mining techniques capable of handling high-velocity data streams. The integration of stream processing frameworks with machine learning models can enable efficient real-time analytics, supporting applications such as smart cities, IoT systems, and financial monitoring (Sikder *et al.*, 2023a,b).

Big data analytics, machine learning, and artificial intelligence are increasingly driving innovation and improving decision-making across economic, healthcare, and business environments. Islam *et al.* (2024) argue that data analytics plays a vital role in economic recovery by supporting crisis response, policy planning, and sustainable development initiatives. In a related study, Kamruzzaman *et al.* (2024) demonstrate that explainable machine learning techniques can enhance the accuracy and transparency of economic growth predictions. Khan *et al.* (2024) further report that big data and business intelligence improve supply chain sustainability by facilitating risk reduction and environmentally responsible operational strategies. Rahman *et al.* (2024) observe that rapid technological change and shifting consumer expectations require organizations to rethink conventional marketing and management approaches through digital transformation. Healthcare applications have also expanded considerably. Ashik *et al.* (2025) highlight the effectiveness of predictive analytics in assessing patient satisfaction, vaccination outcomes, and mental health indicators. Similarly, Rahman *et al.* (2025) find that health informatics supported by data analytics strengthens public health planning and pharmaceutical performance. Emerging technologies such as quantum machine learning offer new opportunities for analyzing complex cancer genomics datasets (Mondal *et al.*, 2025). Furthermore, Saha *et al.* (2024) demonstrate that machine learning-based analytical frameworks can enhance national security operations. Overall, the literature indicates that advanced analytics and intelligent systems are becoming indispensable tools across diverse sectors.

7. Conclusion

This study presents a comprehensive analysis of scalable and efficient data mining techniques for big data analytics in modern computing environments. The results, as illustrated through the comparative and performance-based figures, highlight the critical role of advanced data mining approaches in addressing the challenges associated with large-scale and high-dimensional datasets.

The performance comparison figure demonstrates that advanced and hybrid data mining techniques significantly outperform traditional methods in terms of accuracy and efficiency. This improvement is primarily due to the ability of modern algorithms to process complex data structures and capture nonlinear relationships effectively. Additionally, the scalability analysis figure shows that distributed and parallel processing frameworks enable efficient handling of large datasets, reducing computational time and improving system performance. Overall, the study demonstrates that scalable data mining techniques, when combined with efficient algorithms and modern computing frameworks, provide a robust solution for big data analytics. These approaches enable organizations to extract meaningful insights from large datasets, supporting data-driven decision-making and innovation across various domains. In conclusion, the

findings highlight the necessity of adopting advanced, scalable, and efficient data mining techniques to meet the growing demands of big data analytics. Future research should focus on developing adaptive and resource-efficient models that further enhance scalability, reduce computational complexity, and improve real-time data processing capabilities in modern computing environments.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Aggarwal CC. Data Mining: The Textbook. Cham: Springer; 2015.
2. Alam MI, Hemal MAK, Sami MA, Rahman ML. Robust and interpretable crop recommendation: A case study on a balanced multi-crop agronomic dataset. *Eur J Ecol Biol Agric.* 2024;1(5):168-184. doi:10.59324/ejeba.2024.1(5).14.
3. Alam MI, Sami MA, Al Masud A, Ahmed H, Hossain F. AI-driven big data analytics for personalized cancer treatment: Integrating multi-omics, medical imaging, and predictive intelligence. *J Comput Sci Technol Stud.* 2025;7(11):428-441. doi:10.32996/jcsts.2025.7.11.40.
4. Alam MI, Sami MA, Hemal MAK, Rahman ML. Predictive analytics and decision intelligence for climate-resilient agritech systems. *Acad Glob J Comput Sci Technol Stud.* 2023;2(1):44-56. doi:10.32996/agjcs.2023.2.1.4.
5. Ashik AAM, Hossain E, Rahman MS, Islam S, Khan SI, Rahman MM. Predictive healthcare analytics: Mental health, vaccination effects, and patient satisfaction. In: 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). IEEE; 2025. p. 1934-1947. doi:10.1109/ICIMIA67127.2025.11200745.
6. Chen M, Mao S, Liu Y. Big data: A survey. *Mob Netw Appl.* 2014;19(2):171-209.
7. Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters. *Commun ACM.* 2008;51(1):107-113.
8. Gama J. Knowledge Discovery from Data Streams. Boca Raton (FL): Chapman & Hall/CRC; 2010.
9. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. *Int J Inf Manag.* 2015;35(2):137-144.
10. George G, Haas MR, Pentland A. Big data and management. *Acad Manage J.* 2014;57(2):321-326.
11. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. 3rd ed. Burlington (MA): Morgan Kaufmann; 2012.
12. Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Khan SU. The rise of big data on cloud computing: Review and open research issues. *Inf Syst.* 2015;47:98-115.
13. Hemal MAK, Sayeed N, Sami MA, Alam MI, Sikder TR, Dipa SA, *et al.* Leveraging data analytics to strengthen public health and global economic sustainability. *Eur J Med Health Res.* 2025;3(4):253-263. doi:10.59324/ejmhr.2025.3(4).37.
14. Hilbert M. Big data for development: A review of

- promises and challenges. *Dev Policy Rev.* 2016;34(1):135-174.
15. Islam S, Khan SI, Ashik AAM, Hossain E, Rahman MM, Rahman MS. Big data in economic recovery: A policy-oriented study on data analytics for crisis management and growth planning. *J Comput Anal Appl.* 2024;33(7):2349-2367.
 16. Jagadish HV, Gehrke J, Labrinidis A, Papakonstantinou Y, Patel JM, Ramakrishnan R, *et al.* Big data and its technical challenges. *Commun ACM.* 2014;57(7):86-94.
 17. Juie BJA, Kabir JUZ, Ahmed RA, Rahman MM. Evaluating the impact of telemedicine through analytics: Lessons learned from the COVID-19 era. *J Med Health Stud.* 2021;2(2):161-174. doi:10.32996/jmhs.2021.2.2.19.
 18. Kamruzzaman M, Mondal RS, Islam MK, Rahaman MA, Saha S. AI-driven predictive modelling of US economic growth using big data and explainable machine learning. *Int J Comput Exp Sci Eng.* 2024;10(4):1927-1938. doi:10.22399/ijcesen.3612.
 19. Katal A, Wazid M, Goudar RH. Big data: Issues, challenges, tools and good practices. In: *Proceedings of the International Conference on Contemporary Computing*; 2013. p. 404-409.
 20. Khan SI, Rahman MS, Ashik AAM, Islam S, Rahman MM, Hossain E. Big data and business intelligence for supply chain sustainability: Risk mitigation and green optimization in the digital era. *Eur J Manag Econ Bus.* 2024;1(3):262-276. doi:10.59324/ejmeb.2024.1(3).23.
 21. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436-444.
 22. Marz N, Warren J. *Big Data: Principles and Best Practices of Scalable Realtime Data Systems.* Shelter Island (NY): Manning Publications; 2015.
 23. Mondal RS, Bhuiyan MNA, Kamruzzaman M, Saha S, Siddiki MS. A comparative analysis of outline of tools for data mining and big data mining. *J Bus Manag Stud.* 2025;7(4):232-242. doi:10.32996/jbms.2025.7.4.14.
 24. Mondal RS, Kamruzzaman M, Saha S, Bhuiyan MNA. Quantum machine learning approaches for high-dimensional cancer genomics data analysis. *Comput Integr Manuf Syst.* 2025;31(1):13-32. doi:10.24297/j.cims.2025.1.21.
 25. Nusrat S, Hossain F, Sikder TR. Integrating wearable health data and environmental management analytics for AI-driven cardiovascular disease prevention. *Eastasouth J Inf Syst Comput Sci.* 2024;2(2):209-223. doi:10.58812/esiscs.v2i02.868.
 26. Orthi SM, Sikder TR, Uddin SMM, Roy T, Hossain MJ, Faruk MI. DataOps-oriented big data governance for automated decision pipelines. In: *2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT)*; 2025; Belagavi, India. p. 1-8. doi:10.1109/ICAFT66710.2025.11452860.
 27. Rahman MM, Rahman MS, Islam S, Khan SI, Ashik AAM, Hossain E, Tanvir A. Integrating data analytics into health informatics: Advancing equity, pharmaceutical outcomes, and public health decision-making. *Eurasian J Med Oncol.* 2025;9(4):284-295. doi:10.36922/EJMO025300319.
 28. Rahman MS, Islam S, Khan SI, Ashik AAM, Hossain E, Rahman MM. Redefining marketing and management strategies in digital age: Adapting to consumer behavior and technological disruption. *J Inf Syst Eng Manag.* 2024;9(4):1-16. https://www.researchgate.net/publication/392369297_Redefining_Marketing_and_Management_Strategies_in_Digital_Age_Adapting_to_Consumer_Behavior_and_Technological_Disruption
 29. Saha S, Islam MK, Rahaman MA, Mondal RS, Kamruzzaman M. Machine learning driven analytics for national security operations: A wavelet-stochastic signal detection framework. *J Comput Anal Appl.* 2024;33(8):210. doi:10.48047/jocaaa.2024.33.08.210.
 30. Sami MA, Hemal MAK, Alam MI, Rahman ML. Data governance and analytics infrastructure for scalable decision-making in development and agritech programs. *Eur J Appl Sci Eng Technol.* 2024;2(2):388-403. doi:10.59324/ejaset.2024.2(2).28.
 31. Sikder TR, Dash S, Uddin B, Hossain F. AI-powered data analytics and multi-omics integration for next-generation precision oncology and anticancer drug development. *Eastasouth J Inf Syst Comput Sci.* 2023;1(2):153-170. doi:10.58812/esiscs.v1i02.838.
 32. Sikder TR, Sayeed N, Hossain MJ, Faruk MI, Alam MI, Uddin SMM, *et al.* AI-driven environmental precision oncology: Integrating big data, multi-omics, medical imaging, and exposomic intelligence for personalized cancer care. *Int J Comput Exp Sci Eng.* 2025;11(4). doi:10.22399/ijcesen.4533.
 33. Sikder TR, Siam MA, Melon MMH, Uddin SMM, Mohonta SC, Karim F. A multimodal data analytics framework for early cancer detection using genomic, radiomic, and clinical big data fusion. *J Comput Sci Technol Stud.* 2023;5(3):183-188. doi:10.32996/jcsts.2023.5.3.13.
 34. Tankard C. Big data security. *Netw Secur.* 2012;2012(7):5-8.
 35. Uddin SMM, Chy MAR, Sikder TR, Faruk MI, Adnan M, Hossain MJ. Bio-cognitive AI systems for predictive healthcare decision support. In: *2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT)*; 2025; Belagavi, India. p. 1-9. doi:10.1109/ICAFT66710.2025.11453175.
 36. Vanu N, Hasan MR, Sikder TR, Tamanna ZS. AI-driven big data analytics for precision medicine: A unified framework integrating molecular data intelligence, wearable health systems, and predictive modeling. *J Comput Sci Technol Stud.* 2021;3(2):124-141. doi:10.32996/jcsts.2021.3.2.11.
 37. Zaharia M, Chen A, Davidson A, Ghodsi A, Hong SA, Konwinski A, *et al.* Apache Spark: A unified engine for big data processing. *Commun ACM.* 2016;59(11):56-65.

How to Cite This Article

Roy J, Karim D, Kabir J, Khan R. Scalable and efficient data mining techniques for big data analytics in modern computing. *International Journal of Multidisciplinary Evolutionary Research.* 2025;6(1):52-58. doi:10.54660/IJMER.2025.6.1.52-58.

Creative Commons (CC) License

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.