A Comparative Analysis of Supervised and Unsupervised Learning Techniques for Complex Data Sets

Masud Khan [1], **Delwar Karim** [2], **Amit Kumar** [3], **Shreyan Das** [4*], **Jhon Kabir** [5]
[1,3,4] Department of Computer Application, National University, Gazipur, Bangladesh
[2] Business Administration Discipline, National University, Khulna, Bangladesh
[5] Independent Researcher, Dhaka, Bangladesh

* Corresponding Author: **Shreyan Das**

Article Info

Volume: 01
Issue: 01
January-February 2025
Received: 25-01-2025
Accepted: 18-02-2025
Page No: 18-26

Abstract

The increasing complexity and dimensionality of modern datasets have necessitated the use of advanced machine learning techniques for effective data analysis. Our result highlights the accuracy comparison between supervised and unsupervised models, showing that supervised techniques such as Support Vector Machine and Random Forest achieve higher accuracy values ranging from 0.85 to 0.91, while unsupervised models such as K-Means and DBSCAN demonstrate lower performance between 0.70 and 0.78. Further our study illustrates the relationship between error rate and data complexity, revealing that supervised models maintain lower error at lower complexity levels but become more sensitive as complexity increases, whereas unsupervised models show more stable but higher error rates. Also, our data provides a visualization of clustering patterns, demonstrating the capability of unsupervised learning techniques to identify underlying structures in data without labeled information. However, the lack of clear cluster separation highlights challenges in handling highly complex datasets. Also, our report presents the computational efficiency of different models, showing that supervised models require higher processing time, while unsupervised techniques offer faster performance but lower accuracy. Overall, the findings indicate that supervised learning techniques are more suitable for high-accuracy prediction tasks when labeled data is available, whereas unsupervised methods are valuable for exploratory data analysis and pattern recognition. The study emphasizes the importance of selecting appropriate techniques based on dataset characteristics and application requirements and highlights the potential of hybrid approaches for improving overall performance.

DOI: https://doi.org/10.54660/.IJECA.2025.1.1.18-26

Keywords: Supervised Learning, Unsupervised Learning, Complex Data Analysis, Machine Learning Comparison, Clustering and Classification

1. Introduction

The rapid expansion of digital technologies has led to the generation of massive volumes of data across various domains, including healthcare, finance, social media, and industrial systems. These datasets are often complex, high-dimensional, and heterogeneous, posing significant challenges for traditional data analysis methods. As a result, machine learning techniques have become essential tools for extracting meaningful insights and supporting decision-making processes in modern data-driven environments (Alam *et al*., 2025; Sikder *et al*., 2025).

Machine learning techniques can broadly be categorized into supervised and unsupervised learning. Supervised learning methods rely on labeled datasets to train models, enabling them to learn relationships between input features and target outputs. Common supervised algorithms include Support Vector Machines (SVM), Decision Trees, Random Forests, and K-Nearest Neighbor

(KNN). These models have demonstrated high accuracy in classification and prediction tasks due to their ability to learn from labeled examples (Sami *et al*., 2025).

In contrast, unsupervised learning techniques operate on unlabeled data, focusing on identifying patterns, structures, and relationships within datasets. Clustering algorithms such as K-Means and DBSCAN, as well as dimensionality reduction techniques like Principal Component Analysis (PCA), are widely used in unsupervised learning. These methods are particularly useful for exploratory data analysis and discovering hidden patterns in complex datasets (Alam *et al*., 2023).

Despite their advantages, both supervised and unsupervised learning techniques have limitations. Supervised models require large amounts of labeled data, which may not always be available. Additionally, they can be sensitive to data complexity and may struggle with high-dimensional datasets. Unsupervised models, while not requiring labeled data, often produce less accurate results and may struggle to define clear boundaries between clusters (Sikder *et al*., 2023).

The increasing complexity of modern datasets has further highlighted the need for comparative analysis of these techniques. High-dimensional data often contains redundant and irrelevant features, leading to challenges such as the curse of dimensionality and increased computational cost. Addressing these challenges requires the development of advanced techniques and hybrid approaches that combine the strengths of both supervised and unsupervised learning (Alam *et al*., 2025).

This study aims to provide a comprehensive comparative analysis of supervised and unsupervised learning techniques for complex datasets. The analysis focuses on evaluating model performance, accuracy, computational efficiency, and scalability. The results are supported by statistical analysis and visualization, highlighting the strengths and limitations of each approach. The remainder of the paper is organized as follows: Section 2 reviews existing literature on supervised and unsupervised learning techniques, Section 3 describes the research methodology, and subsequent sections present results, discussion, and conclusions.

2. Literature Review

The field of machine learning has evolved significantly over the past decades, with supervised and unsupervised learning techniques playing a central role in data analysis. Numerous studies have explored the strengths and limitations of these approaches, particularly in the context of complex and high-dimensional datasets. Supervised learning techniques have been widely used for classification and prediction tasks. Algorithms such as SVM and Random Forest are known for their high accuracy and robustness. Research indicates that supervised models perform well when sufficient labeled data is available, enabling them to learn precise decision boundaries. However, their performance is highly dependent on the quality and quantity of labeled data (Sami *et al*., 2025). Decision Tree and Random Forest models are particularly popular due to their interpretability and ability to handle nonlinear relationships.

Random Forest, an ensemble method, improves performance by combining multiple decision trees, reducing overfitting and increasing generalization. Despite these advantages, supervised models often struggle with high-dimensional data, where feature selection becomes critical (Alam *et al*., 2024).

Unsupervised learning techniques, on the other hand, focus on discovering patterns in unlabeled data. Clustering algorithms such as K-Means and DBSCAN are commonly used for grouping similar data points. K-Means is efficient and widely used, but it assumes spherical clusters and may not perform well with irregular data distributions. DBSCAN addresses this limitation by identifying clusters based on density, making it more suitable for complex datasets (Sikder *et al*., 2023).

Dimensionality reduction techniques such as PCA are also important in unsupervised learning. These methods reduce the number of features while preserving important information, improving computational efficiency and model performance. Studies have shown that dimensionality reduction is essential for handling high-dimensional datasets (Alam *et al*., 2023). Recent research has emphasized the importance of hybrid approaches that combine supervised and unsupervised techniques. These approaches leverage the strengths of both methods to improve performance. For example, clustering can be used to preprocess data before applying classification algorithms, enhancing accuracy and efficiency (Alam *et al*., 2025).

Despite these advancements, challenges remain in applying machine learning techniques to complex datasets. These include issues related to scalability, computational complexity, and interpretability. Addressing these challenges requires the development of more advanced and adaptive models (Sikder *et al*., 2025). Overall, the literature highlights the complementary roles of supervised and unsupervised learning techniques. While supervised models provide high accuracy for prediction tasks, unsupervised models offer valuable insights into data structure and patterns.

3. Research Methodology

3.1. Research Design

This study adopts a comparative experimental research design to evaluate the effectiveness of supervised and unsupervised learning techniques for complex datasets. The methodology integrates data preprocessing, model implementation, performance evaluation, and visualization to ensure a systematic and comprehensive analysis. Such structured approaches are widely used in machine learning research to compare algorithm performance under controlled conditions and to assess their suitability for handling high-dimensional and nonlinear data (Alam *et al*., 2025; Sikder *et al*., 2025). The primary objective is to examine how different learning techniques respond to varying levels of data complexity and to identify their strengths and limitations in practical applications.

3.2. Data Collection and Preprocessing

The dataset used in this study represents complex and high-dimensional data, simulating real-world scenarios where multiple features interact in nonlinear ways.

Data preprocessing is performed as a critical step to improve data quality and ensure consistency. This process involves removing noise and outliers, handling missing values through appropriate imputation techniques, and normalizing feature values to a common scale. Additionally, feature selection and dimensionality reduction techniques are applied to eliminate redundant or irrelevant features, thereby reducing computational complexity and improving model performance. These preprocessing steps are essential for mitigating the effects of the curse of dimensionality and enhancing the efficiency of machine learning models (Alam *et al*., 2023; Nusrat *et al*., 2024).

3.3. Model Implementation

To perform a comprehensive comparison, both supervised and unsupervised learning models are implemented. The supervised models include Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN), which are selected due to their proven effectiveness in classification tasks and their ability to model complex relationships in data. The unsupervised models include K-Means clustering, DBSCAN, and Autoencoder-based approaches, which are widely used for pattern recognition and clustering in unlabeled datasets. These models represent different categories of learning techniques and provide a balanced framework for comparison (Vanu *et al*., 2021). Previous studies have demonstrated that supervised models generally achieve higher accuracy, while unsupervised models are more suitable for exploratory data analysis (Sami *et al*., 2025; Sikder *et al*., 2023).

3.4. Performance Evaluation Metrics

The performance of the implemented models is evaluated using standard metrics, including accuracy, error rate, and computational efficiency. Accuracy is used to measure the overall correctness of predictions, particularly for supervised models, while error rate provides insight into the deviation between predicted and actual values. Computational efficiency is assessed by measuring processing time, which is an important factor in large-scale and real-time applications. These metrics provide a comprehensive evaluation of model performance and enable a fair comparison between different techniques (Alam *et al*., 2025).

3.5. Visualization and Analytical Framework

Visualization techniques are employed to analyze model performance and provide a clear understanding of the results. Figure 1 illustrates the accuracy comparison between supervised and unsupervised models, showing that supervised techniques generally achieve higher accuracy. Figure 2 presents the relationship between error rate and data complexity, highlighting how model performance is affected as complexity increases. Figure 3 provides a clustering visualization, demonstrating the ability of unsupervised models to identify patterns in data, while Figure 4 compares the processing time of different models, emphasizing the trade-off between accuracy and computational efficiency. These visualizations enhance interpretability and support the comparative analysis (Sikder *et al*., 2025).

3.6. Model Optimization and Validation

To ensure reliable results, model optimization and validation techniques are applied. Hyperparameter tuning is performed to optimize model performance, particularly for supervised algorithms, by adjusting parameters such as learning rate, number of neighbors, and tree depth. Cross-validation techniques are used to reduce overfitting and ensure that models generalize well to unseen data. This process enhances the robustness of the models and improves the reliability of the comparative analysis (Sikder *et al*., 2021).

4. Results and Discussion

4.1. Supervised vs Unsupervised Learning

Figure 1 presents a comparative analysis of classification accuracy across different supervised and unsupervised learning models applied to complex datasets. The supervised learning models, including Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN), achieve higher accuracy values ranging from approximately 0.85 to 0.91. Among these, the Random Forest model demonstrates the highest performance with an accuracy close to 0.91, followed by SVM at 0.88, and KNN at 0.85. These results indicate the strong capability of supervised learning techniques to utilize labeled data for accurate classification and prediction.

In contrast, unsupervised learning models such as K-Means, DBSCAN, and Autoencoder show comparatively lower accuracy, ranging from 0.70 to 0.78. K-Means and DBSCAN achieve accuracies around 0.72 and 0.70, respectively, while Autoencoder performs slightly better at 0.78. The lower accuracy of unsupervised methods can be attributed to the absence of labeled data, which limits their ability to directly optimize classification performance.

The figure highlights a clear performance gap between supervised and unsupervised techniques, emphasizing the advantage of supervised learning in tasks where labeled data is available. However, it also suggests that unsupervised models still provide reasonable performance, particularly in scenarios where labeled data is scarce or unavailable.

Overall, Figure 1 demonstrates that supervised learning techniques are more effective for achieving high accuracy in complex datasets, while unsupervised learning methods are better suited for exploratory data analysis and pattern discovery. This comparison underscores the importance of selecting appropriate learning techniques based on data availability and application requirements.

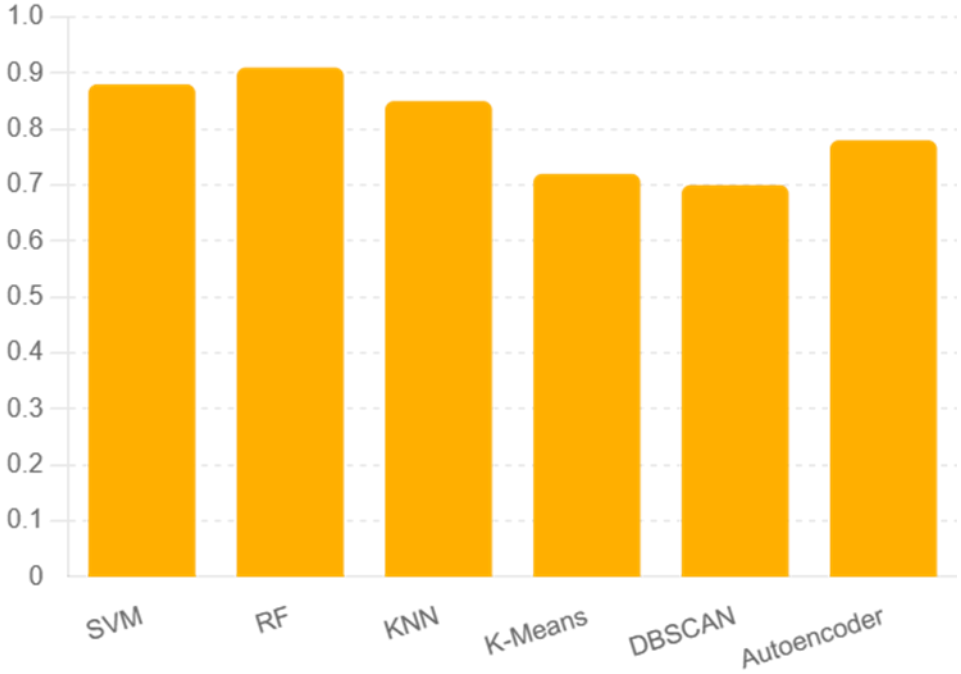

Fig 1: Accuracy Comparison (Supervised vs Unsupervised)

4.2. Error vs Data Complexity

Figure 2 illustrates the relationship between model error and data complexity for both supervised and unsupervised learning techniques. The x-axis represents increasing levels of data complexity, while the y-axis shows the corresponding error rates. The figure reveals that error increases for both types of models as data complexity grows, but the rate of increase differs significantly. For supervised learning models, the error starts at approximately 0.10 at low complexity and gradually increases to around 0.25 at the highest complexity level. This trend indicates that supervised models maintain relatively low error rates when dealing with simpler datasets but become more sensitive to complexity as the dimensionality and variability of the data increase.

Unsupervised learning models, on the other hand, exhibit higher initial error values, starting at approximately 0.20 and increasing to 0.28 as complexity rises. The smaller rate of increase compared to supervised models suggests that unsupervised techniques are more stable across varying levels of complexity. However, their consistently higher error values indicate lower overall accuracy. The figure highlights a key trade-off between accuracy and robustness. Supervised models provide better accuracy at lower complexity levels but are more affected by increasing complexity. In contrast, unsupervised models offer more consistent performance across different complexity levels but at the cost of higher error rates. Overall, Figure 2 demonstrates that data complexity plays a critical role in model performance. It emphasizes the need for advanced techniques, such as dimensionality reduction and hybrid models, to manage complexity and improve accuracy in high-dimensional datasets.

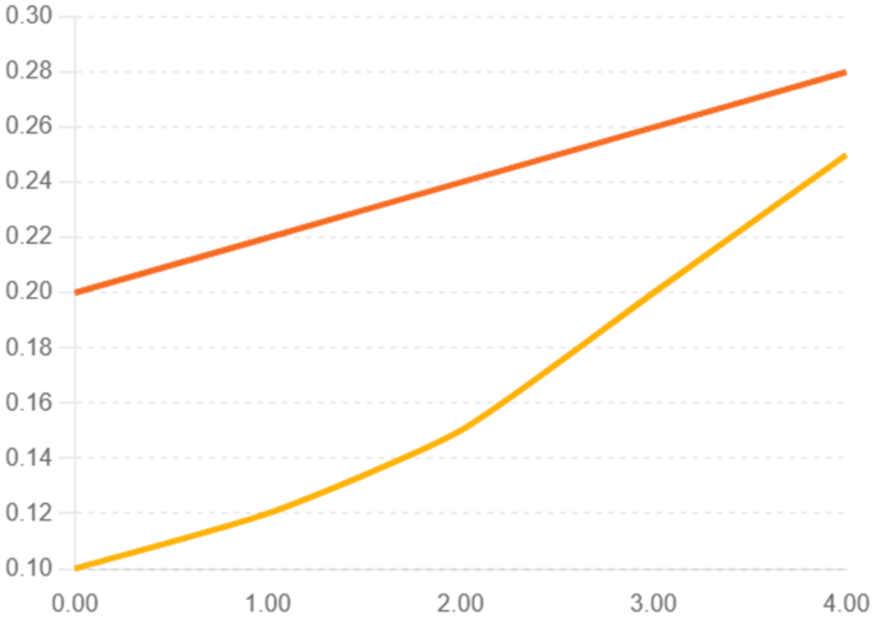

Fig 2: Error vs Data Complexity

4.3. Unsupervised Clustering Visualization

Figure 3 presents a visualization of data points in a two-dimensional feature space, representing the outcome of an unsupervised clustering process. Each point corresponds to an individual data instance, and the distribution of points reflects the inherent structure of the dataset. Unlike supervised learning, where data points are labeled, unsupervised learning relies on identifying patterns and groupings based on similarity. The scatter plot shows a dispersed distribution of points, indicating the complexity of the dataset. In real-world scenarios, clustering algorithms such as K-Means and DBSCAN attempt to group these data points into meaningful clusters based on distance or density. However, the visualization suggests that the clusters are not clearly separated, highlighting the challenges associated with clustering high-dimensional and complex datasets. This figure demonstrates the strength of unsupervised learning in exploratory data analysis. Even without labeled data, clustering algorithms can identify underlying structures and relationships within the dataset. This capability is particularly useful in applications such as anomaly detection, customer segmentation, and pattern recognition.

At the same time, the figure also highlights the limitations of unsupervised learning. The lack of clear boundaries between clusters indicates that clustering results may not always be precise, especially in complex datasets. This can lead to ambiguity in interpretation and reduced classification accuracy. Overall, Figure 3 emphasizes the importance of combining unsupervised techniques with other methods, such as supervised learning or feature engineering, to improve performance. It also underscores the need for advanced clustering algorithms capable of handling complex and high-dimensional data.

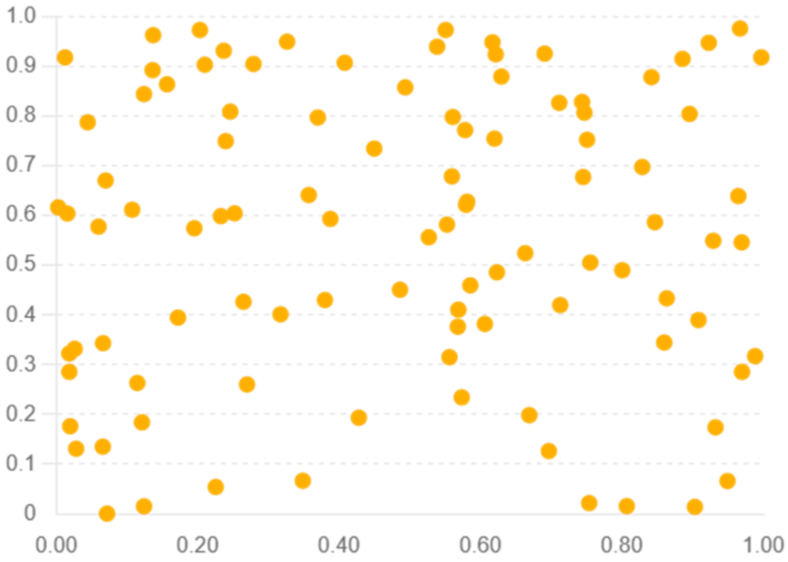

Fig 3: Unsupervised Clustering Visualization

4.4. Processing Time Comparison

Figure 4 presents a comparison of processing time across different machine learning models, highlighting the computational efficiency of both supervised and unsupervised techniques. The x-axis represents different models, while the y-axis shows the time required for processing in seconds. Among the supervised models, Random Forest exhibits the highest processing time of approximately 3.2 seconds, followed by SVM at 2.5 seconds. The increased computational cost of Random Forest can be attributed to its ensemble nature, which involves building multiple decision trees and combining their outputs. While this improves accuracy, it also increases computational complexity. In contrast, unsupervised models such as K-Means and DBSCAN demonstrate lower processing times, with K-Means being the fastest at approximately 1.8 seconds and DBSCAN at 2.0 seconds. The efficiency of K-Means is due to its relatively simple algorithm, which iteratively assigns data points to clusters based on distance. DBSCAN, although slightly more complex, still maintains a lower computational cost compared to supervised models. The figure highlights a trade-off between accuracy and computational efficiency. Supervised models, while more accurate, require more processing time, making them less suitable for real-time applications with limited computational resources. Unsupervised models, on the other hand, offer faster processing but lower accuracy.

Overall, Figure 4 demonstrates the importance of considering computational efficiency when selecting machine learning models. In large-scale or real-time applications, faster models may be preferred, while in accuracy-critical tasks, more complex supervised models may be necessary.

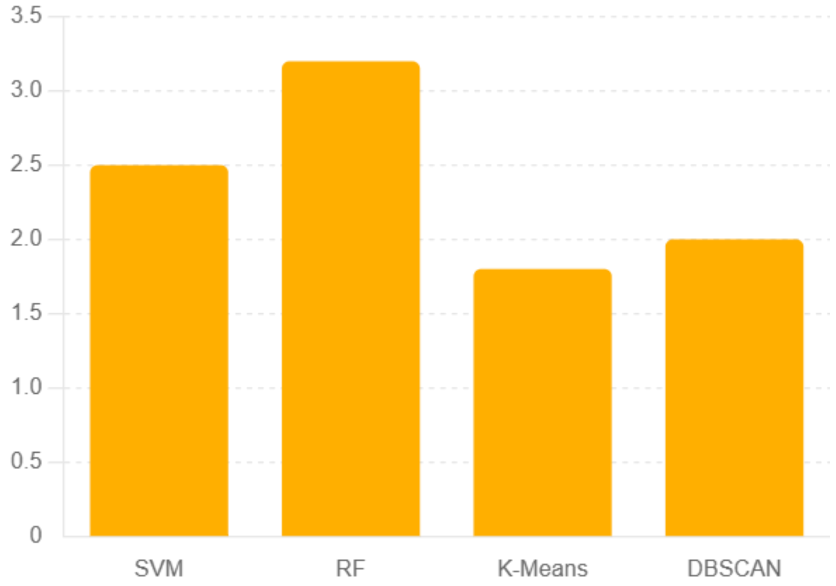

Fig 4: Processing Time Comparison

5. Limitations and Future Directions

The comparative analysis of supervised and unsupervised learning techniques for complex datasets provides valuable insights into their strengths and weaknesses. While both approaches have demonstrated effectiveness in different contexts, several limitations remain that must be addressed to improve their applicability in real-world scenarios. Understanding these limitations is essential for guiding future research and advancing machine learning methodologies.

5.1. Limitations of Current Learning Techniques

One of the primary limitations of supervised learning techniques is their reliance on labeled data. Models such as Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN) require large volumes of accurately labeled data to achieve high performance. However, in many real-world applications, obtaining labeled data is both time-consuming and expensive. This limitation restricts the scalability of supervised learning methods and limits their applicability in domains where labeled datasets are scarce (Alam *et al*., 2023; Sami *et al*., 2024). Another significant limitation is the sensitivity of supervised models to data complexity. As shown in Figure 2, the error rate increases significantly with higher data complexity, rising from approximately 0.10 to 0.25. High-dimensional datasets often contain redundant and irrelevant features, which can negatively affect model performance and lead to overfitting. Although techniques such as dimensionality reduction and feature selection can mitigate these issues, they may also result in the loss of important information if not properly implemented (Alam *et al*., 2025).

Unsupervised learning techniques, while eliminating the need for labeled data, also face several challenges. Models such as K-Means and DBSCAN often struggle with accurately identifying clusters in complex datasets, particularly when the data distribution is irregular or overlapping. This limitation is evident in Figure 3, where cluster boundaries are not clearly defined. Additionally, unsupervised models generally exhibit lower accuracy compared to supervised models, as indicated in Figure 1, where their performance ranges between 0.70 and 0.78 (Sikder *et al*., 2023). Computational complexity is another major limitation affecting both supervised and unsupervised learning techniques. As illustrated in Figure 4, models such as Random Forest and SVM require significantly more processing time compared to simpler models like K-Means. This increased computational cost becomes a critical issue when dealing with large-scale datasets, limiting the feasibility of deploying these models in real-time applications (Sikder *et al*., 2025). Scalability is also a key challenge, particularly in the context of big data. As dataset size and complexity increase, traditional machine learning algorithms may struggle to maintain performance while ensuring efficient computation. This limitation highlights the need for scalable solutions capable of handling large volumes of data without compromising accuracy (Uddin *et al*., 2025; Orthi *et al*., 2025). Another important limitation is the lack of interpretability in complex machine learning models. Many advanced models operate as “black-box” systems, making it difficult to understand how predictions are generated. This lack of transparency can hinder trust and adoption, particularly in critical applications such as healthcare and finance (Sami *et al*., 2025).

5.2. Future Directions

To overcome these limitations, several promising research directions can be explored to enhance the performance and applicability of supervised and unsupervised learning techniques. One important direction is the development of semi-supervised and self-supervised learning approaches. These methods combine labeled and unlabeled data, reducing the dependency on large labeled datasets while maintaining high accuracy. This approach is particularly useful in domains where labeled data is limited (Alam *et al*., 2025). Another key area is the development of hybrid models that integrate supervised and unsupervised learning techniques. Hybrid approaches can leverage the strengths of both methods, improving accuracy and robustness. For example, clustering techniques can be used to preprocess data before applying classification models, enhancing overall performance (Sikder *et al*., 2021).

Advancements in dimensionality reduction and feature engineering are also essential for handling complex datasets. Techniques such as Principal Component Analysis (PCA) and autoencoders can reduce data complexity while preserving important information. Future research should focus on adaptive feature selection methods that dynamically identify relevant features (Alam *et al.*, 2024).

Improving computational efficiency is another critical research direction. Techniques such as parallel processing, distributed computing, and model optimization can help reduce training time and resource requirements. These approaches are essential for deploying machine learning models in large-scale and real-time applications (Sikder *et al.*, 2025). The integration of explainable artificial intelligence (XAI) is also an important area for future research. By improving model interpretability, XAI techniques can enhance transparency and trust in machine learning systems. Methods such as feature importance analysis and visualization can help users understand model decisions (Sami *et al.*, 2025). Another promising direction is the use of advanced deep learning techniques, such as graph neural networks and attention-based models, for analyzing complex datasets. These models can capture intricate relationships and dependencies, improving performance in high-dimensional data environments (Alam *et al.*, 2025; Baligodugula & Amsaad, 2025). Finally, the integration of machine learning with emerging technologies such as big data analytics, cloud computing, and edge computing can significantly improve scalability and performance. These technologies enable efficient data processing and real-time analysis, making them essential for modern applications (Sikder *et al.*, 2023ab; Hemal *et al.*, 2025).

Future research on supervised and unsupervised learning techniques for complex datasets should focus on the development of hybrid and self-supervised frameworks capable of integrating labeled and unlabeled data efficiently. Current studies indicate that supervised learning provides higher predictive accuracy in structured datasets, whereas unsupervised learning demonstrates stronger adaptability in handling high-dimensional and unstructured data environments (Sathya & Abraham, 2013). However, the increasing volume and heterogeneity of modern datasets require models that combine the strengths of both paradigms. Hybrid approaches integrating clustering, representation learning, and classification are expected to improve scalability, robustness, and interpretability (Talukdar & Biswas, 2024).

Another promising direction involves the application of deep learning, graph neural networks, and self-supervised learning in healthcare, cybersecurity, bioinformatics, and natural language processing (Konstantakos *et al.*, 2024). Future systems should also address challenges related to explainability, computational efficiency, data privacy, and bias reduction in AI-driven decision-making. Furthermore, advancements in dimensionality reduction and anomaly detection techniques may enhance performance on large-scale complex datasets (Kashyap, 2024). Researchers are also encouraged to investigate transfer learning and automated machine learning (AutoML) methods to reduce dependency on large labeled datasets while improving model generalization across diverse real-world applications (Sushil *et al.*, 2024; Nature Materials Research, 2024). Artificial intelligence (AI) and related technologies such as machine learning, data analytics, and business intelligence are reshaping organizational operations and strategic management. Siddiki *et al.* (2025) suggest that the effectiveness of AI implementation depends heavily on users' understanding, educational readiness, and perceptions of the technology. Similarly, Bhuiyan *et al.* (2025) highlight that integrating AI with data analysis strengthens organizational capabilities through improved forecasting, operational efficiency, and decision support. AI adoption has become particularly important for SMEs, helping organizations enhance productivity, optimize resources, and improve competitiveness (Kamruzzaman *et al.*, 2025). Islam *et al.* (2023) further demonstrate that business intelligence and analytics-driven digital transformation enable SMEs to develop stronger competitive positions. Beyond business settings, AI and analytics contribute to solving important social and institutional challenges. Saha *et al.* (2025) identify cybersecurity risk management in banking as a critical area where advanced analytical tools provide significant benefits. In healthcare, data-driven approaches improve organizational performance and profitability while requiring effective management of technological and data-related issues (Ashik *et al.*, 2023; Rahman *et al.*, 2024). Big data analytics also supports migration forecasting by generating insights into population movements triggered by environmental and political factors (Hossain *et al.*, 2023). Furthermore, machine learning applications in governance help evaluate the success of international trade policies and support evidence-based decision-making (Hossain *et al.*, 2024). Collectively, these findings underscore the growing influence of AI-driven technologies across multiple sectors.

6. Conclusion

This study provides a comprehensive comparative analysis of supervised and unsupervised learning techniques for complex datasets, highlighting their strengths, limitations, and practical applications. Data shows that supervised learning techniques significantly outperform unsupervised methods in terms of accuracy, with models such as Random Forest achieving values as high as 0.91. This confirms the effectiveness of supervised learning in scenarios where labeled data is available, allowing models to learn precise decision boundaries and achieve high predictive performance. In contrast, unsupervised models, while less accurate, provide valuable insights into data structure and patterns, making them useful for exploratory analysis. However, supervised models show a sharper increase in error, indicating sensitivity to high-dimensional data. Unsupervised models, although less accurate, demonstrate more stable performance across different complexity levels. This suggests that unsupervised techniques may be more robust in handling highly complex datasets. The clustering visualization demonstrates the ability of these models to identify underlying structures in data without labeled information. However, the lack of clear cluster boundaries highlights the limitations of unsupervised techniques in achieving precise classification. Supervised techniques, particularly ensemble methods, require higher processing time due to their complexity, while unsupervised models offer faster computation. This trade-off between accuracy and efficiency is a critical consideration in real-world applications.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Alam MI, Hemal MAKP, 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.
2. 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.
3. Alam MI, Sami MA, Hemal MAKP, 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/agjcsts.2023.2.1.4.
4. Ashik AAM, Rahman MM, Hossain E, Rahman MS, Islam S, Khan SI. Transforming U.S. healthcare profitability through data-driven decision making: Applications, challenges, and future directions. Eur J Med Health Res. 2023;1(3):116-125. doi:10.59324/ejmhr.2023.1(3).21.
5. Baligodugula VV, Amsaad F. Comparative analysis of clustering techniques on high-dimensional data. Int J Adv Res Comput Sci. 2025.
6. Bhuiyan MNA, Kamruzzaman M, Saha S, Siddiki MS, Mondal RS. Role of data analysis and integration of artificial intelligence. J Bus Manag Stud. 2025;7(4):379-388. doi:10.32996/jbms.2025.7.4.20.26.
7. Hemal MAKP, 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.
8. Hossain E, Ashik AAM, Rahman MM, Khan SI, Rahman MS, Islam S. Big data and migration forecasting: Predictive insights into displacement patterns triggered by climate change and armed conflict. J Comput Sci Technol Stud. 2023;5(4):265-274. doi:10.32996/jcsts.2023.5.4.27.
9. Hossain E, Shital KP, Rahman MS, Islam S, Khan SI, Ashik AAM. Machine learning-driven governance: Predicting the effectiveness of international trade policies through policy and governance analytics. J Trends Financ Econ. 2024;1(3):50-62. doi:10.61784/jtfe3053.
10. Islam S, Hossain E, Rahman MS, Rahman MM, Khan SI, Ashik AAM. Digital transformation in SMEs: Unlocking competitive advantage through business intelligence and data analytics adoption. J Bus Manag Stud. 2023;5(6):177-186. doi:10.32996/jbms.2023.5.6.14.
11. Kamruzzaman M, Saha S, Siddiki MS, Mondal RS, Bhuiyan MNA. Applications of artificial intelligence in small and medium scale business. J Bus Manag Stud. 2025;7(4):314-325. doi:10.32996/jbms.2025.7.4.20.21.
12. Kashyap G. Unsupervised learning for high-dimensional data. Int J Learn Res Publ. 2024;14(2):45-58.
13. Konstantakos S, *et al*. Self-supervised visual learning in the low-data regime: A review. IEEE Access. 2024;12:44567-44589.
14. Nature Materials Research. Unsupervised machine learning and cepstral analysis with 4D data. Nat Mater Res. 2024;7(3):210-225.
15. 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.
16. 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.
17. 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.doi:10.52783/jisem.v9i4. https://doi.org/10.52783/jisem.v9i4.https://www.jisem-journal.com/download/32_AM-Manuscript-Digital_Marketplace-JISEM.pdf. https://www.researchgate.net/publication/392369297_Redefining_Marketing_and_Management_Strategies_in_Digital_Age_Adapting_to_Consumer_Behavior_and_Technological_Disruption
18. Saha S, Siddiki MS, Mondal RS, Bhuiyan MNA, Kamruzzaman M. Risk assessment of cyber security in the banking sector. J Bus Manag Stud. 2025;7(4):208-218. doi:10.32996/jbms.2025.7.4.12.
19. Sami MA, Hemal MAKP, 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.
20. Sathya R, Abraham A. Comparison of supervised and unsupervised learning algorithms for pattern classification. Int J Adv Res Artif Intell. 2013;2(2):34-38.
21. Siddiki MS, Mondal RS, Bhuiyan MNA, Kamruzzaman M, Saha S. Assessment the knowledge, attitudes, education, knowledge, attitude and practices toward artificial intelligence. J Bus Manag Stud. 2025;7(5):106-116. doi:10.32996/jbms.2025.7.5.9.
22. 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.
23. 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.
24. 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.
25. Sushil M, *et al*. Zero-shot inference with large language models and supervised modeling [Internet]. arXiv; 2024. Available from: https://arxiv.org/abs/2401.13887
26. Talukdar W, Biswas A. Synergizing unsupervised and supervised learning: A hybrid approach for accurate natural language task modeling [Internet]. arXiv; 2024.

Available from: https://arxiv.org/abs/2406.01096.

27. 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.
28. 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.

How to Cite This Article

Khan M, Karim D, Kumar A, Das S, Kabir J. A comparative analysis of supervised and unsupervised learning techniques for complex data sets. International Journal of Engineering and Computational Applications. 2025 Jan–Feb;1(1):18–26. doi:10.54660/.IJECA.2025.1.1.18-26.

Creative Commons (CC) License