International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I11P120

Volume 72 Issue 11, 184-191, November 2024

Original Article

Proposing a Hybrid Machine Learning Technique for Efficient Outlier Detection

Bondu Venkateswarlu[1], C. Jayaramulu[2]

[1]*Department of CSE, Dayananda Sagar University, Bangalore, India.*
[2]*Dayananda Sagar University, Bangalore, India.*

[1]*Corresponding Author : bonduvenkat-cse@dsu.edu.in*

Received: 25 June 2024 Revised: 02 October 2024 Accepted: 06 November 2024 Published: 29 November 2024

Abstract - *Finding the best results when outliers in high-dimensional data is challenging because of data imbalance and dimensionality. Several algorithms were created in an attempt to address this problem. However, they are made using either an unsupervised learning approach or a supervised learning technique, and they can discover outliers from such data. Supervised learning methods utilize training data, whereas unsupervised learning methods provide tools for finding and applying complex patterns. The central claim of this article is that it is possible to "merge two approaches to produce a hybrid and gain from both worlds." A state-of-the-art ML system that combines both under supervision and without methods for efficiently identifying outliers is provided to assess this claim. Furthermore, the Outlier Detection Multi-Model Approach (MMA-OD) is a method proposed in this paper. This method combines the advantages of both supervised and unsupervised learning models to improve performance. Its strength is increasing the size of the feature space. Many benchmark datasets are used to assess the suggested approach. The empirical results show that MMA-OD outperforms a wide range of alternative approaches.*

Keywords - *Hybrid Machine Learning, Outlier detection, Framework, ROC comparison, Execution time comparison.*

1. Introduction

Outlier identification is a crucial technique for addressing specific challenges in real-world applications. It has the potential to uncover hidden information that may not be apparent in situations with frequent item sets or recurring issues. As a result, outlier detection has emerged as a vital field of research with numerous practical applications. Its significance spans various domains, including education, healthcare, and networking. While there are several established methods for identifying outliers, the advent of machine learning has expanded the options. Various machine learning methods, including supervised ones that can locate outliers without needing labeled data, are now being used. However, it is important to note that specific outlier identification methods based on supervised learning require training. One may also find semi-supervised methods in the literature. Many hybrid models have been created to achieve the goal of identifying anomalies. The research gap found in the literature is that the existing methods lack novelty in exploiting hybrid approaches in machine learning. This paper strives to exploit a hybrid methodology for discovering outliers from data. The following contributions to this paper are proposed: an algorithm known as the Outlier Detection Multi-Model Approach (MMA-OD). The technique enhances performance by combining the benefits of supervised and unsupervised learning models. Its strength is obtaining a more significant feature space. The proposed method is evaluated using many benchmark datasets. The empirical results show that MMA-OD outperforms a wide range of alternative approaches.

The remainder of the document is organized as follows: Section 2 reviews the literature on previous works related to outlier detection. Section 3 presents the proposed methodology for outlier detection using a hybrid approach. Section 4 presents the results of our empirical study. Section 5 discusses the paper's research and gives the study's limitations. Section 6 presents the conclusions drawn and discusses the scope of future research.

2. Related Work

Several outlier detection methods are available for managing high-dimensional datasets. Thudumu et al. [1] highlighted the challenges of anomaly detection due to the large dimensionality of the data, which negatively impacts performance and accuracy. To address this issue, novel frameworks are required. Chen et al. [2] introduced a framework for hybrid anomaly detection in optical networks that integrates supervised regression with unsupervised clustering, resulting in a system with minimal false positives and high accuracy. Jeffrey et al. [3] proposed a combined anomaly detection technique that utilizes threshold-based, behavior-based, and signature-based methods specifically for Cyber-Physical Systems (CPS). Poutre et al. [4] developed an innovative hybrid anomaly detection system to identify manipulations related to trading activities in Limit Order Books (LOBs).

This system employs simulation techniques and a modified Transformer autoencoder to achieve advanced fraud detection capabilities. Lastly, Terbuch et al. [5] explored using unsupervised variational autoencoders alongside physics-based indicators in a hybrid machine-learning approach for detecting anomalies in multivariate time series within geotechnical engineering. Their findings indicated that optimizing hyperparameters and experimenting with different architectures lead to improved performance.

Sikder and Batarseh [6] examined OD techniques, grouping them into six categories and reviewing how they work and what they might be used for. Ali et al. [7] discussed formation-related abnormalities in well-logging data and suggested a mix of supervised and unsupervised techniques for ML to reconstruct density logs properly. Without requiring data labeling, Lee et al. [8], an anomaly detection technique based on hybrid DL for intelligent factories, improves output quality and production efficiency. Cao et al. [9] assessed the predictive power of gradient boosting, random forests, and extra trees for ozone concentration in membrane contactors. With the PO algorithm tuned, the models reached excellent accuracy levels. Acharya et al. [10]depended on computer networks, which are more vulnerable to attack with Digital information transport. High accuracy is attained in anomaly detection with deep learning, especially with CNN Bi-LSTM models.

Cai et al. [11] assert that for the safety and effectiveness of smart grids, it is essential to have the capability to detect power theft. This study introduces HFR-WSVDD, an ensemble learning technique that enhances feature representation and classification. Matinkia et al. [12] found that compared to conventional methods, hybrid approaches such as MLP-SSD offer improvements in both efficiency and accuracy. The estimation of rock permeability, which is crucial for hydrocarbon production, can be achieved through various techniques. Gutierrez et al. [13] presented the STFA-HDLID algorithm, which combines STFA, DBN, and SSO to deliver highly accurate intrusion detection and classification in IoD environments. Velasquez et al. [14] enhanced performance metrics by implementing Industry 4.0 systems that utilize a mixed machine learning ensemble to detect real-time anomalies. Ibrahim et al. [15] improved previous methods by introducing a hybrid deep learning methodology for anomaly identification in univariate time series, specifically using a 1D CNN-BiLSTM model.

Sakhnenko et al. [16] explored PQCs and measurement methodologies; combining quantum and classical methods improves anomaly identification in the HAE model. Zhou et al. [17] introduced HAD-MDGAT, a hybrid approach that considerably outperforms baselines for multivariate time series anomaly detection by combining GAT and MDA. Rosenberger et al. [18] addressed several anomalies in edge computing; the EEM-KDE method improves anomaly detection for industrial applications. Umer et al. [19] approach, concentrating on intrusion and anomaly detection methods, helps Industrial Control Systems withstand cyberattacks. Ilyas et al. [20] offered a hybrid technique for identifying unusual crowds in videos that combines handmade and deep features.Mustaqeem and Saqib [21] presented a hybrid method for predicting software defects that combines PCA and SVM, yielding better accuracy than previous approaches—Sudar et al. [22] state that network security becomes more critical when data use rises. Deep learning and machine learning are two aspects of big data analytics that help in ID. Liu et al. [23] suggested a hybrid ID model that effectively classifies data by combining RF, k-means, and deep learning.

Kurt et al. [24] presented scalable, efficient, and assumption-free Nonparametric Methods for detecting anomalies in high-dimensional data streams in real-time. Rabbani et al. [25] presented a PSO-PNN technique that exhibits encouraging outcomes for identifying and detecting harmful activities in cloud settings. Alirezaei et al. [26] investigated the genetic and lifestyle factors associated with diabetes. They utilized data mining methods to facilitate the early detection of the disease. The study employed meta-heuristic and clustering techniques for data preprocessing and feature selection.

In another study, Pu et al. [27] developed the SSC-OCSVM, an approach to unsupervised anomaly detection that combines Sub-Space Clustering with One-Class Support Vector Machine (SVM). This method outperforms existing techniques on the NSL-KDD dataset, emphasizing efficient feature selection and parallelization as goals for future research. Toshniwal et al. [28] utilized Anomaly Detection Techniques (ADT), exploring various methods such as clustering and classification.

They emphasized that ADT should be rapid, flexible, and unsupervised. Zhou et al. [29] proposed a VLSTM-based anomaly detection model designed explicitly for Industrial Big Data (IBD), which enhances accuracy by effectively addressing class imbalance and high dimensionality issues. Future research aims to improve this model further. Lastly, Ma et al. [30] introduced AD-H1CD, a method for detecting abnormalities in large-scale Local Area Networks (LANs) based on a hybrid neural network. This approach demonstrates superior performance by utilizing multiple types of characteristics.

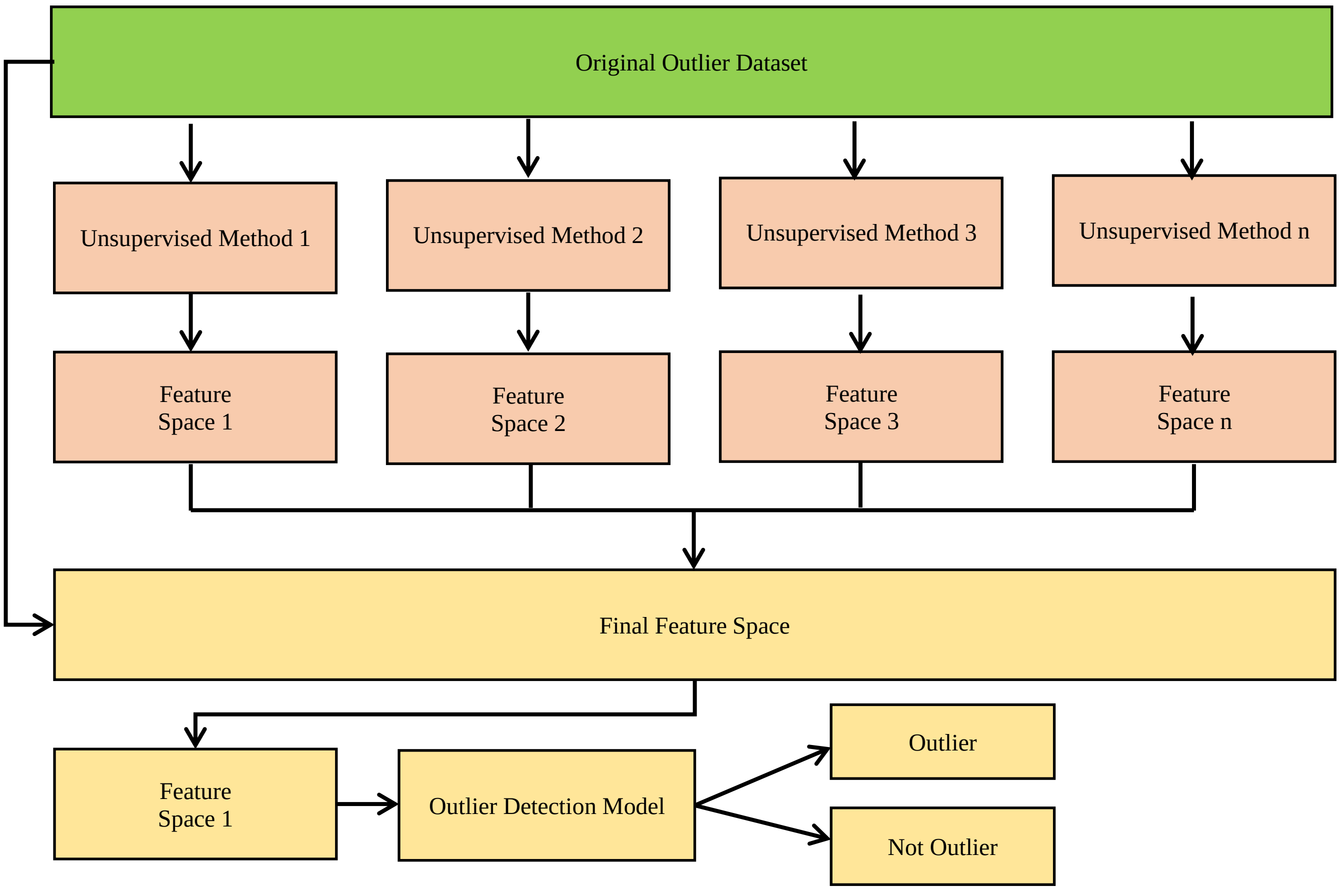

Fig. 1 Proposed framework for outlier detection

3. Methodology

A framework named Outlier Detector with Machine Learning Basis (ML-OD) is proposed and implemented. The following subsections provide additional details on the suggested method, including the underpinning algorithm and framework.

3.1. The Framework

The proposed framework is based on ML methods that utilize supervised and unsupervised techniques, as illustrated in Figure 1. Various unsupervised baseline techniques provide the feature space required for outlier detection. Each method used in the empirical study can generate a unique feature space. Multiple feature areas are created using this method, which are then combined to produce a merged feature space. The final feature space is formed by amalgamating the features obtained from the source data and the combined feature space. The feature space is constructed using unsupervised learning techniques because it is well-established that they provide a more accurate reflection of the data. After recovering the feature space, supervised classifier training may finish the outlier discovery process. This is a detailed explanation of the proposed structure. Let X be an array of annotated observations. There are several points and elements. Equation 1 may represent the connection between the extracted feature space and the tagged observations.

$$L = \{(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^{n\times d}\} \quad (1)$$

The algorithm scores outliers using the formula. $\phi(.)$, which results in a vector-like $\phi_i(x) \in \mathbb{R}^{n\times d}$on X. The degree of outlierness is shown. Outlier findings are shown, with a 1 signifying an outlier and a 0 denoting none. An unsupervised learning strategy using features is equivalent to an outlier scoring function.

A multitude of scoring functions are generated since there are several comparable approaches. There exist k scoring functions that function on X R, with the combined feature space represented as A matrix of these scoring functions as shown in Equation 2.

$$\phi(\mathrm{x}) = [\phi_1(\mathrm{x})^{\mathrm{T}}, \dots, \phi_{\mathrm{k}}(\mathrm{x})^{\mathrm{T}}] \in \mathbb{R}^{n\times k} \quad (2)$$

Differentiating the data across a range of scoring functions produces unique representations. Diversity is guaranteed since closely related scoring functions are not good predictors. Following the computation of $\phi_i(x)$, feature space may be calculated using Equation 3.

$$FeatureSpace_{new} = [x, \phi(\mathrm{x})] \in \mathbb{R}^{n\times l} \quad (3)$$

Table 1. Use of notations

Notation	Description
$\varphi(\phi_i)$	Accuracy function
$\rho(\phi_i, \phi_j)$	Denotes Pearson correlation function
X	Denotes Input
S	Outlier score functions
$\mathbb{R}$	Represents actual values
$\phi = [\phi_1, \dots . \phi_k]$	Outlier function matrix
$\phi(.)$	Represents an outlier scoring function.
N	Every point in the dataset
L	Annotated observations
D	characteristics of the dataset
ACC_i	ROC of $\phi_i(X)^T$
$\{(x_1, y_1), \dots, (d_n, y_n)\}$	Collection of labelled observations

Apart from offering many unsupervised ways to generate outlier results, the proposed algorithm also provides A system for determining the optimal choice method that balances diversity and accuracy. This is reflected in the expression in Equation 4.

$$\varphi(\phi_i) = \frac{ACC_I}{\sum_{j=1}^{\#(S)} |\rho(\phi_i, \phi_j)|} \quad (4)$$

The Pearson correlation function is employed to determine the connection. This function meets the requirement.$\phi_j \in \{S\}, ACC_I \geq 0$, and is represented as $\rho(\phi_i, \phi_j)$.

$$\sum_{j=1}^{\#(S)} |\rho(\phi_i, \phi_j)| \quad (5)$$

Equation 5 describes how the correlation function is connected to the aggregation process.

3.2. Design of Algorithm

The proposed algorithm, the Multi-Model Method for Outlier Identification (MMA-OD), uses a hybrid strategy that blends unsupervised and supervised ML methods. This approach recognizes that while unsupervised learning effectively represents feature space, supervised learning is more accurate in predicting class labels [28]. The algorithm operates based on these principles. Algorithm 1 returns R, the outliers found in D. It has several outlier functions to calculate the outlier score. The optimal score for each outlier function score and the accuracy function are found in an iterative process.

A supervised model is trained once the feature space has been computed, and the output is used to predict outliers. The suggested strategy is better than the current ones since it uses a deliberate selection process and various supervised learning approaches.

Algorithm 1: Multi-model technique for outlier detection

Algorithm: Multi-Model Method for Outlier Detection (MMA-OD)
Input: High dimensional dataset D
Output: Outlier detection results R

1. F←OutlierFunctions()
2. S←Null //outlier scores
3. V←BestScoreDynamics()
4. Fr each function f in F
5. s←GetOutlierScore()
6. Add function and s to S
7. End For
8. For each entry s found in S
9. Compute function for accuracy as in Eq. 4
10. b←BestScoreComputation()
11. Add b to V
12. Remove s in S
13. End For
14. Use D to add V details to S along with feature space
15. model←TrainTheMachneLearning()
16. R←ModelTesting(model)
17. Return R

4. Experimental Results

Large-scale experimental datasets are extracted from [21]. After 30 iterations, the mean score obtained from each trial is evaluated. Error rates and Receiver Operating Characteristic Curve (ROC) are the metrics used for assessment. The MMA-OD algorithm is compared with the latest methods in the field. This approach defines the feature space in an unsupervised environment using baseline outlier functions. The fundamental procedures referenced are from [22], [23], [24], along with LoOP, LOF, k-Median, KNN, and average-KNN. Additionally, one-class SVM methods are drawn from [23]. The proposed approach is compared with those identified in [25] and [26]. All the datasets used in the empirical study are widely used in outlier detection research. As shown in Table 2, The suggested approach, MMA-OD, outperforms current techniques in terms of ROC.

Table 2. ROC comparison

DATASET	ROC Performance			
	XGB	**PSO-PNN [25]**	**K-Means Clustering [26]**	**Proposed**
Speech	0.7593	0.8515	0.8534	0.8591
Satellite	0.9656	0.9156	0.9096	0.9666
Mnist	0.9963	0.989	0.988	0.9999
Mammography	0.9515	0.941	0.9415	0.9431
Letter	0.9399	0.9653	0.9685	0.9729
Cardio	0.9966	0.9953	0.9879	0.9976
Arrhythmia	0.8698	0.8537	0.8545	0.8816

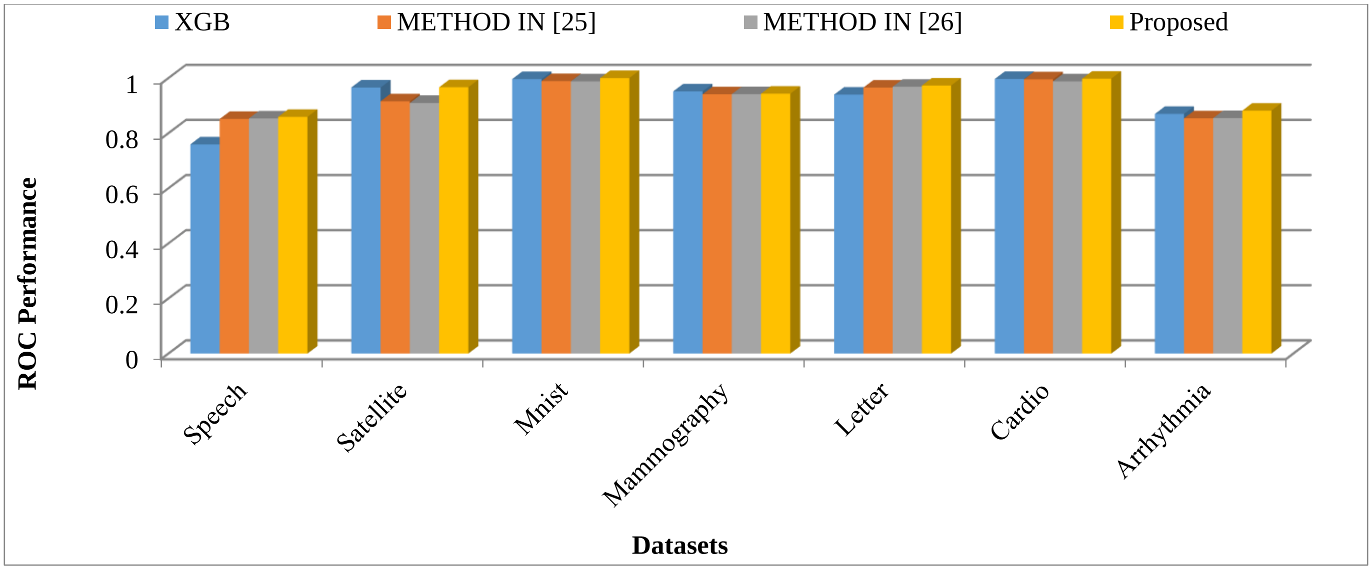

Fig. 2 ROC comparison

Figure 2 displays the results of the proposed method, MMA-OD, regarding ROC. The observations are compared to the latest methods. Higher ROC scores indicate better performance. According to the results, the recommended ROC value of the approach is higher than the existing techniques for all datasets presented. The proper method selection made the performance boost possible, which struck the ideal balance between accuracy and diversity. The recommended MMA-OD technique's maximum The Mnist dataset has a ROC of 0.9999.

As indicated in Table 3, the precision performance of the suggested MMA-OD method is contrasted with that of the current approaches.

Table 3. Precision@N comparison

DATASET	PRECISION@N			
	XGB	PSO-PNN [25]	K-Means Clustering [26]	Proposed
Speech	0.1696	0.1569	0.2082	0.2561
Satellite	0.8508	0.7566	0.7508	0.8568
Mnist	0.9195	0.8409	0.8379	0.9901
Mammography	0.6877	0.6013	0.5754	0.6677
Letter	0.6181	0.6653	0.6874	0.732
Cardio	0.9302	0.8925	0.8508	0.9377
Arrhythmia	0.5932	0.553	0.5449	0.6002

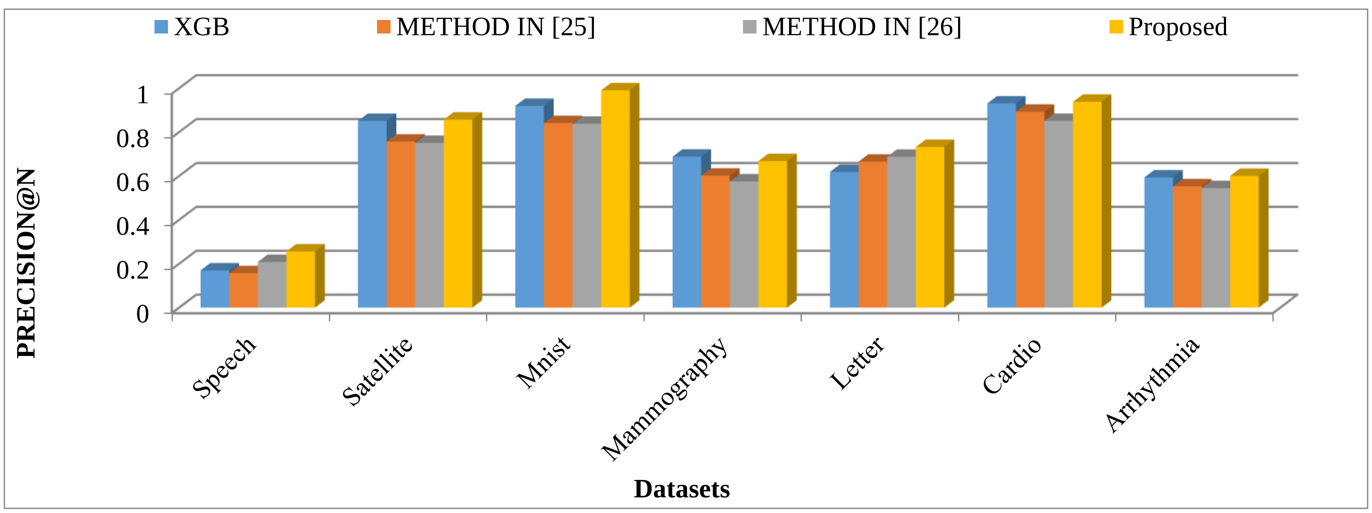

Fig. 3 Performance comparison of outlier detection

Table 4. Execution time comparison

DATASET	EXECUTION TIME (Sec)			
	XGB	PSO-PNN [25]	K-Means Clustering [26]	Proposed
Speech	8.1853	5.9067	2.2505	10.177
Satellite	0.5018	144.79	28.252	4.157
Mnist	1.6107	19.359	26.428	4.9213
Mammography	0.2968	38.182	1.9436	5.4165
Letter	0.152	1.2231	0.5719	1.0286
Cardio	0.1456	3.1053	0.6215	1.1174
Arrhythmia	0.3963	1.5521	0.6281	0.6181

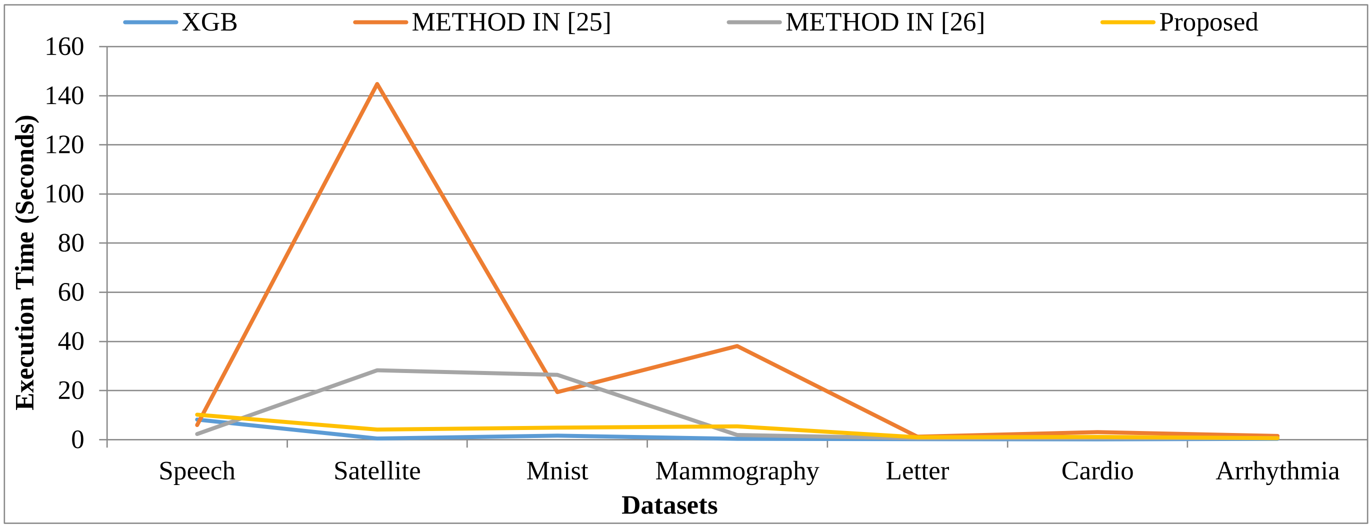

Fig. 4 Execution time comparison

The precision@N outcomes of the suggested MMA-OD approach are shown in Figure 3. The most recent methods are contrasted with the observations. Higher precision levels correspond to better performance. The findings indicate that the precision@N value of the suggested strategy is more significant compared to the current approaches for all datasets presented. The proper method selection made the performance boost possible, which struck the ideal balance between accuracy and diversity. The results of the recommended MMA-OD approach for the Mnist dataset show the highest precision@N of 0.9901.

As indicated in Table 4, the suggested technique MMA-OD is evaluated in terms of execution time against current procedures. Figure 4 displays the execution time results of the proposed MMA-OD technique. The latest methods are compared with the observations. In shorter execution times, better performance is evident. The results demonstrate that the recommended strategy takes longer than existing methods for all datasets presented. Performance has increased because the correct approach was chosen to strike the right balance between accuracy and diversity. With an execution time of 0.6181 seconds, the recommended method The quickest speed for the Arrhythmia dataset is MMA-OD. Many empirical results demonstrate that the proposed strategy outperforms existing approaches in outlier detection.

5. Discussion

Outlier detection using learning-based approaches, specifically with artificial intelligence, has become increasingly important due to the vast amounts of data that require intelligent handling. The application of learning-based methods, whether supervised or unsupervised, has become quite common in recent years. However, supervised learning can need to help perform well because of the challenges of providing sufficient labeled data. Similarly, unsupervised learning is limited due to the absence of labeled data. This underscores the need to utilize both supervised and unsupervised approaches, which can enhance performance in outlier detection. In this paper, a hybrid methodology that leverages both approaches is proposed. This includes methods for extracting features and generating the feature space needed for the supervised learning process. Essentially, the unsupervised methods generate ground truth data that the supervised method can use, enabling each to compensate for the other's shortcomings and improve overall outlier detection performance. However, as discussed in Section 5.1, our proposed methodology does have certain limitations.

5.1. Limitations

Although the proposed system presented in this paper employs an efficient hybrid approach, it has certain limitations. First, the datasets used for outlier detection have a relatively finite number of samples. Due to the lack of diversity in the data, the results of the experiments cannot be easily generalized. Additionally, a significant shortcoming of the proposed system is its reliance on ground truth generated by unsupervised learning methods rather than utilizing a standard and proven ground truth.

6. Conclusion and Future Work

Our research introduced a new framework that uses supervised and unsupervised ML methods for outlier detection. Among the suggested tactics is the MMA-OD. This method leverages supervised and unsupervised learning models for higher performance and an expanded feature space. We have evaluated this approach using numerous benchmark datasets, and the empirical results demonstrate that MMA-OD outperforms other commonly used methods.

Our central claim is that it is feasible to integrate these two approaches to create a hybrid model that incorporates both strengths. Our experimental data have confirmed this assertion. Our ultimate goal is to develop an ensemble learning technique to enhance outlier detection performance in the future. Another important direction for the future scope of the research is to create a methodology that uses autoencoders to detect outliers efficiently.

References

[1] Srikanth Thudumu et al., "A Comprehensive Survey of Anomaly Detection Techniques for High Dimensional Big Data," *Journal of Big Data*, vol. 7, no. 1, pp. 1-30, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Xiaoliang Chen et al., "Self-Taught Anomaly Detection with Hybrid Unsupervised/Supervised Machine Learning in Optical Networks," *Journal of Lightwave Technology*, vol. 37, no. 7, pp. 1742-1749, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[3] Nicholas Jeffrey, Qing Tan, and José R. Villar, "A Hybrid Methodology for Anomaly Detection in Cyber–Physical Systems," *Neurocomputing*, vol. 568, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] CédricPoutré, Didier Chételat, and Manuel Morales, "Deep Unsupervised Anomaly Detection in High-Frequency Markets," *The Journal of Finance and Data Science*, vol. 10, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Anika Terbuch et al., "Detecting Anomalous Multivariate Time-Series via Hybrid Machine Learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Nazmul Kabir Sikder, and Feras A. Batarseh, "Outlier Detection using AI: A Survey," *AI Assurance*, pp. 231-291, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Muhammad Ali et al., "A Novel Machine Learning Approach for Detecting Outliers, Rebuilding Well Logs, and Enhancing Reservoir Characterization," *Natural Resources Research*, vol. 32, pp. 1047-1066, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kyung Sung Lee, Seong Beom Kim, and Hee-Woong Kim, "Enhanced Anomaly Detection in Manufacturing Processes Through Hybrid Deep Learning Techniques," *IEEE Access*, vol. 11, pp. 93368-93380, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Yusong Cao, Chunqin Hu, and Limin Ma, "Separation of Organic Compound from Water Using Membrane Process: Hybrid Machine Learning-Based Modeling and Validation," *Case Studies in Thermal Engineering*, vol. 51, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Toya Acharya, Annamalai Annamalai, and Mohamed F Chouikha, "Efficacy of CNN-Bidirectional LSTM Hybrid Model for Network-Based Anomaly Detection," *In Proceedings 2023 IEEE 13th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Penang, Malaysia, pp. 348-353, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Qingyuan Cai, Peng Li, and Ruchuan Wang, "Electricity Theft Detection Based on Hybrid Random Forest and Weighted Support Vector Data Description," *International Journal of Electrical Power & Energy Systems*, vol. 153, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Morteza Matinkia et al., "Prediction of Permeability from Well Logs Using a New Hybrid Machine Learning Algorithm.," *Petroleum*, vol. 9, no. 1, pp. 108-123, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Jos´e Escorcia-Gutierrez et al., "Sea Turtle Foraging Algorithm with Hybrid Deep Learning-Based Intrusion Detection for the Internet of Drones Environment," *Computers and Electrical Engineering*, vol. 108, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] David Velásquez et al., "A Hybrid Machine-Learning Ensemble for Anomaly Detection in Real-Time Industry 4.0 Systems," *IEEE Access*, vol. 10, pp. 72024-72036, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Moamen Ibrahim, Khaled M. Badran, and Ahmed Esmat Hussien, "Artificial Intelligence-Based Approach for Univariate Time-Series Anomaly Detection Using Hybrid CNN-BiLSTM Model," *In Proceedings 2022 13th International Conference on Electrical Engineering (ICEENG)*, Cairo, Egypt, pp. 129-133, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Alona Sakhnenko et al., "Hybrid Classical-Quantum Autoencoder for Anomaly Detection," *Quantum Machine Intelligence*, vol. 4, no. 27, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Liwen Zhou, Qingkui Zeng, and Bo Li, "Hybrid Anomaly Detection via Multihead Dynamic Graph Attention Networks for Multivariate Time Series," *IEEE Access*, vol. 10, pp. 40967-40978, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Julia Rosenberger et al., "Extended Kernel Density Estimation for Anomaly Detection in Streaming Data," *Procedia CIRP*, vol. 112, pp. 156-161, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] Muhammad Azmi Umer et al., "Machine Learning for Intrusion Detection in Industrial Control Systems: Applications, Challenges, and Recommendations," *International Journal of Critical Infrastructure Protection*, vol. 38, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Zirgham Ilyas et al., "A hybrid Deep Network Based Approach for Crowd Anomaly Detection," *Multimedia Tools and Applications*, vol. 80, pp. 24053-24067, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] Mohd. Mustaqeem and Mohd. Saqib, "Principal Component Based Support Vector Machine (PC-SVM): A Hybrid Technique for Software Defect Detection," *Cluster Computing*, vol. 24, pp. 2581-2595, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[22] K. Muthamil Sudar et al., "Analysis of Intruder Detection in Big Data Analytics," *In Proceedings 2021 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, pp. 1-5, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[23] Chao Liu, Zhaojun Gu, and Jialiang Wang, "A Hybrid Intrusion Detection System Based on Scalable K-Means+ Random Forest and Deep Learning," *IEEE Access*, vol. 9, pp. 75729-75740, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Mehmet Necip Kurt, Yasin Yılmaz, and Xiaodong Wang, "Real-Time Nonparametric Anomaly Detection in High-Dimensional Settings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2463-2479, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[25] Mahdi Rabbani et al., "A Hybrid Machine Learning Approach for Malicious Behaviour Detection and Recognition in Cloud Computing," *Journal of Network and Computer Applications*, vol. 151, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[26] Mahsa Alirezaei, Seyed Taghi Akhavan Niaki, and Seyed Armin Akhavan Niaki, "A bi-objective Hybrid Optimization Algorithm to Reduce Noise and Data Dimension in Diabetes Diagnosis Using Support Vector Machines," *Expert Systems with Applications*, vol. 127, pp. 47-57, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[27] Guo Pu et al., "A Hybrid Unsupervised Clustering-Based Anomaly Detection Method," *Tsinghua Science and Technology*, vol. 26, no. 2, pp. 146-153, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[28] Akanksha Toshniwal, Kavi Maheshand, and R. Jayashree, "Overview of Anomaly Detection Techniques in Machine Learning," *In Proceedings 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, India, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[29] Xiaokang Zhou et al., "Variational LSTM Enhanced Anomaly Detection for Industrial Big Data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3469-3477, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[30] Chencheng Ma, Xuehui Du, and Lifeng Cao, "Analysis of Multi-Types of Flow Features Based on Hybrid Neural Network for Improving Network Anomaly Detection," *IEEE Access*, vol. 7, pp. 148363-148380, 2019. [CrossRef] [Google Scholar] [Publisher Link]