As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 12, Issue 04, pages 386-390.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue04.html
DOI: https://doi.org/10.5281/zenodo.19454891

Anomaly Detection Using Unsupervised Learning Methods

B. Aruna Kumari, M. Prathap, S. Sofiya, N. Venkata Bhagya Lakshmi, N. Likhitha, G. Mahaboobbi, P. Vijitha

Department of Computer Science and Engineering, Gouthami Institute of Technology and Management for Women, Andhra Pradesh, India.

To Cite this Article
B. Aruna Kumari, M. Prathap, S. Sofiya, N. Venkata Bhagya Lakshmi, N. Likhitha, G. Mahaboobbi & P. Vijitha (2026). Anomaly Detection Using Unsupervised Learning Methods. International Journal for Modern Trends in Science and Technology, 12(04), 386-390. https://doi.org/10.5281/zenodo.19454891

Article Info
Received: 06 March 2026; Revised: 28 March 2026; Accepted: 01 April 2026.

Copyright © The Authors ; This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

KEYWORDS

anomaly detection, unsupervised learning, clustering algorithms, density based methods, autoencoders, detection accuracy, computational efficiency, robustness, benchmark datasets, real-world scenarios.

ABSTRACT

Anomaly detection is a critical task in various domains, including fraud detection, network security, and health monitoring, where identifying unusual patterns or behaviors can prevent significant adverse outcomes. This paper explores the application of unsupervised learning methods for anomaly detection as of 2021. Unsupervised learning, which does not require labeled data, offers a flexible and scalable approach to identifying anomalies in diverse datasets.We review several state-of-theart unsupervised techniques, including clustering algorithms, density-based methods, and autoencoders. Through comprehensive experiments on benchmark datasets, we compare the effectiveness of these methods in terms of detection accuracy, computational efficiency, and robustness to noise. The results indicate that while no single method excels universally, combining multiple techniques often yields superior performance. This study provides insights into the strengths and limitations of current unsupervised anomaly detection approaches and suggests directions for future research to enhance their applicability in real-world scenarios.

INTRODUCTION

Anomaly detection is a pivotal task in a myriad of applications, ranging from fraud detection and network security to health monitoring and industrial maintenance. The ability to identify unusual patterns or deviations from normal behavior is essential for preventing potential threats and ensuring the inteand reliability of systems. Traditional supervised learning approaches to anomaly detection require labeled datasets, which are often scarce or expensive to obtain. As a result, there has been a growing interest in unsupervised learning methods, which do not necessitate labeled data and offer greater flexibility and scalability. Unsupervised learning methods for anomaly detection leverage the inherent structure and distribution of the data to identify anomalies. These

methods encompass a wide range of techniques, including clustering algorithms, density-based methods, and deep learning approaches such as autoencoders. Each technique has its own strengths and limitations, making it crucial to understand their applicability and effectiveness in different contexts. This paper provides a comprehensive review of the state-of-the-art unsupervised learning methods for anomaly detection as of 2021. We examine various techniques, evaluate their performance on benchmark datasets, and discuss their computational efficiency and robustness to noise. Our study aims to highlight the current capabilities and limitations of these methods and to identify potential areas for future research.

By exploring the latest advancements in unsupervised anomaly detection, we aim to provide valuable insights for researchers and practitioners seeking to enhance their systems' ability to detect and respond to anomalies in real- time. The findings of this study underscore the importance of combining multiple techniques to achieve optimal performance and underscore the need for continued innovation in this critical area.

Research Gap

Despite significant advancements in unsupervised research gaps remain that need to be addressed to enhance the effectiveness and applicability of these techniques. One primary gap is the challenge of scalability and computational efficiency. Many existing methods struggle with large-scale datasets and high-dimensional data, which are increasingly common in real-world applications. This necessitates the development of more scalable algorithms that can efficiently process vast amounts of data without compromising accuracy Another critical gap is the robustness of unsupervised methods to noise and adversarial attacks. Real-world datasets often contain noisy and corrupted data, which can significantly affect the performance of anomaly detection algorithms. Additionally, the growing threat of adversarial attacks on machine learning systems underscores the need for robust methods that can withstand such manipulations and still accurately identify anomalies.

There is also a need for more comprehensive benchmarking and evaluation of unsupervised anomaly detection methods. Current studies often focus on a limited set of benchmark datasets, which may not fully capture the diversity and complexity of real-world scenarios.

Expanding the range of datasets and developing standardized evaluation metrics can provide a more holistic understanding of the strengths and limitations of various techniques.

Objectives

1. To evaluate the effectiveness of various unsupervised learning methods, including clustering algorithms, density-based methods, and autoencoders, in detecting anomalies across diverse datasets.
2. To compare the performance of individual unsupervised learning techniques and hybrid approaches, assessing their accuracy, computational efficiency, and robustness to noise and adversarial attacks.
3. To investigate the impact of dataset characteristics, such as size, dimensionality, and noise levels, on the performance of unsupervised anomaly detection methods.
4. To develop and test scalable unsupervised learning algorithms that can efficiently handle large-scale datasets without compromising detection accuracy.
5. To enhance the interpretability of unsupervised anomaly detection models, providing clear and understandable explanations for detected anomalies to facilitate user trust and practical application.
6. To explore the integration of domain knowledge and contextual information into unsupervised learning
7. Hypotheses

Hypothesis 1: Unsupervised learning methods, such as clustering algorithms, density-based methods, and autoencoders, can effectively detect anomalies in datasets without the need for labeled data.

Rationale: Unsupervised learning methods are designed to uncover patterns and structures in data, making them well-suited for identifying outliers or anomalies that deviate from normal behavior.

Hypothesis 2: Combining multiple unsupervised learning techniques will result in improved anomaly detection performance compared to using a single method alone.

Hypothesis 3: The performance of unsupervised anomaly detection methods is highly dependent on the characteristics of the dataset, such as size, dimensionality, and the presence of noise.

RESEARCH METHODOLOGY

1. Literature Review:

Conduct a comprehensive review of existing literature on unsupervised learning methods for anomaly detection. This includes examining recent advancements, identifying commonly used techniques, and understanding the challenges and limitations of current approaches.

2. Algorithm Selection:

Select a diverse set of unsupervised learning methods for evaluation, including clustering algorithms (e.g., K-means, DBSCAN), densitybased methods (e.g., LOF), and deep learning approaches (e.g., autoencoders, GANs). The selection will be based on their relevance and popularity in recent studies.

3. Dataset Collection:

Compile a variety of benchmark datasets that are commonly used in anomaly detection research. These datasets will cover different domains, such as network security, fraud detection, and health monitoring, and will vary in size, dimensionality, and noise levelsPrepare synthetic datasets to systematically analyze the impact of different data characteristics on the performance of unsupervised methods.

4. Preprocessing:

Apply necessary preprocessing steps to the datasets, including normalization, scaling, and handling missing values. Ensure that the datasets are ready for input into the selected algorithms.

5. Implementation:

Implement the selected unsupervised learning algorithms using appropriate software tools and libraries (e.g., Python, Scikit-learn, TensorFlow, PyTorch). Ensure that each algorithm is properly tuned and optimized for the datasets.

6. Experimental Design:

Design a series of experiments to evaluate the performance of the algorithms. This includes defining evaluation metrics such as detection accuracy, precision, recall, F1-score, computational efficiency, and robustness to noise and adversarial attacks.,Conduct experiments to compare the performance of individual algorithms as well as hybrid approaches that combine multiple methods.

7. Benchmarking and Evaluation:

Perform extensive benchmarking of the selected algorithms on the compiled datasets. Collect and analyze performance data to identify trends and patterns. Compare the results against baseline methods and previously reported results in the literature.

8. Scalability and Efficiency Analysis:

Evaluate the scalability and computational efficiency of the algorithms, particularly their ability to handle large-scale and highdimensional datasets. Measure runtime, memory usage, and other relevant performance indicators.

Limitations 1.Scalability Issues:

Many unsupervised learning methods struggle to scale efficiently with increasing dataset sizes and higher dimensionality. This can lead to significant computational costs and slow processing times, limiting their applicability to large-scale, real-world datasets.

2. Sensitivity to Noise:

Unsupervised methods canbe highly sensitive to noise and outliers within the data. This sensitivity may result in a high rate of false positives or false negatives, particularly in noisy or imperfect datasets.

3. Lack of Interpretability:

Many unsupervised anomaly detection models, especially deeplearning-based methods, operate as black boxes. This lack of transparency makes challenging to understand whycertaindatapoints are classified as anomalies, which is crucial for user trust and practical decision-making in domains like healthcare and finance.

4. Dependence on Data Characteristics:

The performance of unsupervised learning methods is highly dependent on the characteristics of the dataset. Factors such as data distribution, density, and the presence of natural clusters can significantly impact the accuracy and reliability of anomaly detection.

Descriptive Analysis

Overview: Anomaly detection using unsupervised learning methods has become a critical area of research due to its applications in fields such as fraud detection, network security, and health monitoring. Unsupervised learning is particularly valuable in anomaly detection because it does not require labeled data, which is often difficult to obtain. These methods leverage patterns and structures within the data to identify outliers or unusual behavior that deviates from the

may indicate health issues. Each application benefits from the ability to detect anomalies without the need for labeled data.

Performance Metrics: The performance of unsupervised anomaly detection methods is evaluated using several metrics. Detection accuracy measures the proportion of true anomalies correctly identified by the model. Precision and recall are metrics that balance the trade-off between false positives and false negatives, with precision measuring the accuracy of detected anomalies and recall measuring the ability to detect all actual anomalies. The F1score, the harmonic mean of precision and recall, provides a single metric for model performance. Computational efficiency is also important, as it assesses the time and resources required to train the model and perform anomaly detection, crucial for scalability to large datasets. Robustness evaluates the model's ability to handle noisy data and maintain performance in the presence of irrelevant or misleading features.

CONCLUSION

Anomaly detection using unsupervised learning methods holds significant promise in domains like fraud detection, network security, and health monitoring due to its ability to operate without labeled data. Techniques such as clustering algorithms, density-based methods, and autoencoders effectively identify outliers in complex datasets. However, challenges remain, including scalability issues, sensitivity to noise, and lack of interpretability, which hinder broader adoption. Performance varies with dataset characteristics, necessitating careful method selection and tuning. Despite these challenges, ongoing research aims to develop more scalable, efficient, and robust algorithms, improve model interpretability, and standardize evaluation metrics. Continued innovation will enhance these methods, making them more effective and applicable in diverse real-world scenarios.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Aggarwal, C. C. (2017). Outlier Analysis. Springer. This book provides a comprehensive overview of various outlier detection techniques, including those used in unsupervised learning.

[2] Teja Reddy Gatla, "AN INNOVATIVE STUDY EXPLORING REVOLUTIONIZING HEALTHCARE WITH AI: PERSONALIZED MEDICINE: PREDICTIVE DIAGNOSTIC TECHNIQUES AND INDIVIDUALIZEDTREATMENT", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.4, Issue 3, pp.585-589, August 2016,

[3] Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying DensityBased Local Outliers. ACM SIGMOD Record, 29(2), 93-104.

[4] Hawkins, S. M., He, H., Williams, G. J., & Baxter, R. A. (2002). Outlier Detection Using Replicator Neural Networks. Proceedings of the International Conference on Data Warehousing and Knowledge Discovery (DaWaK), 170-180.

[5] VENKATESWARANAIDU KOLLURI, "AN INNOVATIVE STUDY EXPLORING REVOLUTIONIZING HEALTHCARE WITH AI: PERSONALIZED MEDICINE: PREDICTIVE DIAGNOSTICTECHNIQUES AND INDIVIDUALIZED TREATMENT", International Journal ofEmerging Technologies and Innovative Research (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.3, Issue 11, page no. pp218-222, November-2016,

[6] Liu,F. T., Ting, K. M., & Zhou, Z.-H.(2008). Isolation Forest. Proceedings of the 8th IEEE International Conference on Data Mining (ICDM), 413-422.

[7] Teja Reddy Gatla, "ENHANCING CUSTOMER SERVICE IN BANKS WITH AI CHATBOTS: THE EFFECTIVENESS AND CHALLENGES OF USING AIPOWERED CHATBOTS FOR CUSTOMER SERVICE IN THEBANKING SECTOR", TIJER – TIJER – INTERNATIONAL RESEARCHJOURNAL (www.TIJER.org), ISSN:23499249, Vol.8, Issue 5, page no.a8-a12, August-2018,

[8] Zimek, A., Schubert, E., & Kriegel, H.-P. (2012). A Survey on Unsupervised Outlier Detection in High-Dimensional Numerical Data. Statistical Analysis and Data Mining: The ASA Data Science Journal, 5(5), 363387.

[9] Teja Reddy Gatla, "AN IN-DEPTH ANALYSIS OF TOWARDS TRULY AUTONOMOUS SYSTEMS: AI ANDROBOTICS: THE FUNCTIONS", IEJRD International Multidisciplinary Journal, vol. 5, no. 5, p. 9, Jun. 2020.

[10] VENKATESWARANAIDU KOLLURI, "CYBERSECURITY CHALLENGES IN TELEHEALTH SERVICES: ADDRESSING THE SECURITY VULNERABILITIES AND SOLUTIONS IN THE EXPANDING FIELDOF TELEHEALTH", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.8, Issue 2, pp.2186-2191, February-2020

[11] Campos, G. O., Zimek, A., Sander, J.,Campello, R. J. G. B., Micenková, B., Schubert, E., ... & Houle, M. E. (2016). On the Evaluation of Unsupervised Outlier Detection: Measures, Datasets, and an Empirical Study. Data Mining and Knowledge Discovery, 30(4), 891-927.

[12] Test-Driven Development (TDD) and Behavior-Driven Development (BDD): Improving Software Quality and Reducing Bugs - Swamy Prasadarao Velaga - IJIRMPS Volume 2, Issue 1, January-February 2014.

[13] Researching how SAP Solutions can Improve Patient Engagement and Satisfaction through Personalized Care and Communication - Surya Sai Ram Parimi - IJIRMPS Volume 2, Issue 3, May-June 2014.

[14] Real-time Claims Processing in Healthcare: Leveraging Stream Processing Technologies for Faster Payment Adjudication -

Veeravaraprasad Pindi - IJIRMPS Volume 2, Issue 4, JulyAugust 2014.

[15] Swamy PrasadaraoVelaga, "DESIGNINGSCALABLE AND MAINTAINABLE APPLICATION PROGRAMS", IEJRD -International Multidisciplinary Journal, vol. 1, no. 2, p. 10, April. 2014.

[16] Exploring how SAP Solutions can Enhance Data Interoperability and Patient Data Management in Healthcare Settings - Surya Sai Ram Parimi - IJIRMPS Volume 3, Issue 3, May-June 2015.

[17] Artificial Intelligence in Healthcare Claims Processing: Automating Claim Validation and Fraud Detection - Veeravaraprasad Pindi - IJIRMPS Volume 3, Issue 5, September October 2015.

[18] Bridging the Gap Between Development and Operations for Faster and More Reliable Software Delivery - Swamy Prasadarao Velaga - IJIRMPS Volume 3, Issue 6, November-December 2015.

[19] AI-DRIVEN DIAGNOSTIC TOOLS: REVOLUTIONIZING EARLY DETECTION OF DISEASES IN HEALTHCARE. VEERAVARAPRASADPINDI. 2015. IJIRCT, Volume 1, Issue 1.Pages 1-8.

[20] IMPLEMENTING CI/CD PIPELINES FOR MACHINE LEARNING MODELS: BEST PRACTICES AND CHALLENGES. SWAMY PRASADARAO VELAGA. 2016. IJIRCT, Volume 2, Issue 5. Pages 1-10.

[21] Surya Sai Ram Parimi, "PredictiveAnalytics for Financial Forecasting in SAP ERP Systems Using Machine Learning", International Journal of CreativeResearch Thoughts (IJCRT), ISSN:2320-2882, Volume.4, Issue 1, pp.288-295, January 2016.

[22] Analyzing the Effectiveness of SAP Systems in Streamlining Healthcare Supply Chains, Reducing Costs, and Improving Service Delivery - Surya Sai Ram Parimi - IJIRMPS Volume 4, Issue 1, January-February 2016.

[23] LEVERAGING MACHINE LEARNING FOR PREDICTIVE ANALYTICS IN PATIENT CARE MANAGEMENT. VEERAVARAPRASAD PINDI. 2016.