UDC 004.5
DOI https://doi.org/10.26661/2786-6254-2025-1-10

METHODS AND ALGORITHMS FOR ANOMALY DETECTION

Skakovskyi V. O.
Master's Student at the Department of Software Engineering
Zhytomyr Polytechnic State University
Chudnivska str., 103, Zhytomyr, Ukraine
orcid.org/0009-0001-8947-4207
vlad.sk.2003@gmail.com

Savitskyi R. S.
Senior Lecturer at the Department of Software Engineering
Zhytomyr Polytechnic State University
Chudnivska str., 103, Zhytomyr, Ukraine
orcid.org/0000-0001-9804-3604
roman.savitskyi@gmail.com

Fant M. O.
Associate Professor at the Department of Software Engineering
Zhytomyr Polytechnic State University
Chudnivska str., 103, Zhytomyr, Ukraine
orcid.org/0000-0002-4994-8009
fantkolja@gmail.com

Key words: *anomaly detection, outlier detection, machine learning, deep learning.*

The article examines several approaches to finding anomalies in complicated datasets, an essential task in domains including industrial monitoring, cybersecurity, banking, and healthcare. The paper evaluates the theoretical underpinnings, benefits, and drawbacks of anomaly detection strategies, classifying them into statistical approaches, clustering techniques, deep learning models, and hybrid/ensemble methods.

Although they provide theoretically solid methods for identifying outliers, statistical techniques such as Z-score analysis, Gaussian Mixture Models (GMM), and Kernel Density Estimation (KDE) have trouble handling high-dimensional and non-linear data. Whilst K-means, DBSCAN, and hierarchical clustering are popular clustering approaches for unsupervised anomaly detection, their efficacy depends on noise sensitivity and parameter choice.

Additionally, the study looks at deep learning and machine learning methods, including autoencoders, Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks. Particularly for time-series and picture data, these models have revolutionized anomaly detection, yet their interpretability and computational efficiency present difficulties.

Furthermore, hybrid and ensemble approaches combine many techniques to increase accuracy and resilience. In challenging anomaly detection tasks, methods such as stacking ensembles, boosting, and isolation forests show improved performance. Interpretability and computational requirements are still major issues.

Many current research issues, such as explainability, model scalability, and adaptability to changing contexts, are highlighted in the article. Future research should concentrate on federated learning, explainable AI (XAI), and semi-supervised learning to ensure more transparent and effective anomaly detection models. This study lays the groundwork for future developments in the area.

МЕТОДИ ТА АЛГОРИТМИ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ

Скаковський В. О.
магістр кафедри інженерії програмного забезпечення
Державний університет «Житомирська політехніка»
вул. Чуднівська, 103, Житомир, Україна
orcid.org/0009-0001-8947-4207
vlad.sk.2003@gmail.com

Савіцький Р. С.
старший викладач кафедри інженерії програмного забезпечення
Державний університет «Житомирська політехніка»
вул. Чуднівська, 103, Житомир, Україна
orcid.org/0000-0001-9804-3604
roman.savitskyi@gmail.com

Фант М. О.
доцент кафедри програмного забезпечення
Державний університет «Житомирська політехніка»
вул. Чуднівська, 103, Житомир, Україна
orcid.org/0000-0002-4994-8009
fantkolja@gmail.com

Ключові слова: *виявлення аномалій, виявлення викидів, машинне навчання, глибоке навчання.*

У статті розглядаються кілька підходів до пошуку аномалій у складних наборах даних, що є важливим завданням у сферах промислового моніторингу, кібербезпеки, банківської справи q охорони здоров’я. У статті оцінюються теоретичні основи, переваги та недоліки стратегій виявлення аномалій, класифіковано їх на статистичні підходи, методи кластеризації, моделі глибокого навчання та гібридні / ансамблеві методи. Статистичні методи, як-от аналіз Z-показника, моделі суміші Гауса (GMM) і оцінка щільності ядра (KDE), мають проблеми з обробкою багатовимірних і нелінійних даних. Хоча K-середні, DBSCAN та ієрархічна кластеризація є популярними методами кластеризації для неконтрольованого виявлення аномалій, ефективність залежить від чутливості до шуму та параметрів.

У дослідженні розглядаються методи глибокого й машинного навчання, зокрема автокодери, генеративні змагальні (GAN), згорткові нейронні мережі (CNN) і мережі довгострокової короткочасної пам’яті (LSTM). Для часових рядів і графічних даних моделі зробили революцію у виявленні аномалій.

Гібридний і ансамблевий підходи поєднують багато методів для підвищення точності та стійкості. У складних завданнях виявлення аномалій такі методи, як утворення ансамблів, посилення й ізоляція лісів, демонструють покращену продуктивність. Інтерпретація та вимоги до обчислень усе ще залишаються головними проблемами.

У статті висвітлено багато актуальних дослідницьких питань, як-от пояснюваність, масштабованість моделі й адаптованість до мінливих контекстів. Майбутні дослідження повинні зосередитися на федеративному навчанні, пояснюваному штучному інтелекті (XAI) і напівконтрольованому навчанні. Це дослідження закладає основу для майбутніх розробок у цій сфері.

Problem statement and its connection to significant scientific and practical tasks. Anomaly detection is critical for ensuring system stability, reliability, and security, especially in financial systems, industrial IoT, and healthcare applications. Outliers, which can be considered indicative of any disturbances that can include faults, cyber threats, or fraud, are very harmful if not recognized early. For example, possible anomalies detectable by such trained systems include unusual vibrations or leaps in temperature in manufacturing, abnormal patterns of logins or transfers of massive amounts of data, and fractures of chemical formulas in the case of pharmaceuticals. They can save a significant amount of resources for businesses, save the privacy of users, prevent breaches with leaks of sensitive data, and ensure the quality of products by detecting compromised ones in case of identifying issues in time.

Traditional methods, like rule-based systems and statistical models, struggle with complex, dynamic, and high-dimensional data; instead, they are suitable for fast, shallow initial checks, but their abilities are limited for in-depth analysis of complex data sets, including time series. By bringing more sophisticated methods that can identify intricate patterns and handle massive amounts of data, machine learning has greatly improved outlier detection. Nonetheless, there are still certain issues with interpretability, scalability, and environmental adaptation. They emphasize the necessity of continual study and advancement to improve anomaly detection methods.

Addressing these challenges is essential for improving system resilience and preventing substantial disruptions. There is a critical practical and scientific task with direct applications across various industries in improving anomaly detection methods' features such as accuracy, transparency, and adaptability. Tackling difficulties is essential for researchers and practitioners, as it addresses critical needs in today’s data-driven world.

Analysis of recent studies and publications. Significant progress in the discipline has been fueled by the growing complexity of contemporary datasets and the urgent need for real-time anomaly identification. Recent surveys, like Pang et al. (2021), highlight the use of AI techniques in tackling contemporary issues by classifying anomaly detection methods into statistical, clustering-based, and machine-learning approaches [1].

Statistical methods, including Gaussian Mixture Models (GMM), are widely used for anomaly detection, but their effectiveness is limited in high-dimensional or non-linear data scenarios. Wu et al. (2024) introduced a Trend and Variance Adaptive Bayesian Changepoint Analysis combined with Local Outlier Scoring, effectively addressing real-world data complexities [2].

Clustering-based approaches, including density-based spatial clustering of applications with noise (DBSCAN) and k-means, remain widely applied in anomaly detection. However, modern methods like Density-Based Networks introduced by Zheng et al. (2021) address the sensitivity to parameter selection and scalability issues associated with traditional clustering techniques [3].

Machine learning, particularly deep learning, has revolutionized anomaly detection. Yang et al. (2022) highlighted the advantages of transformer-based architectures in anomaly detection by proposing a Transformer-based GAN, demonstrating its effectiveness in capturing complex time-series patterns. [4]. Chabchoub et al. (2022) enhanced the Isolation Forest algorithm with Majority Voting Isolation Forest (MVIForest), improving detection accuracy and reducing execution time [5].

Hybrid and ensemble approaches are increasingly being adopted to leverage the strengths of multiple techniques. Explainable AI (XAI) has become essential in critical domains, addressing the challenge of model interpretability. By integrating SHAP, Noorchenarboo et al. (2025) research aims to improve the interpretability of detection models, facilitating better decision-making in smart grid operations [6].

Despite these advancements, challenges persist in model adaptability, interpretability, and generalization across domains, and research is actively addressing these issues by developing interpretable models, adaptive algorithms for streaming data, and techniques for transfer learning. While significant progress has been made, continuing innovation is necessary to meet the demands of modern systems and dynamic datasets.

Formulating the goals and objectives. The main goal of this article is to provide a theoretical analysis of algorithms and machine learning models for anomaly detection. Conceptual strengths, weaknesses, and suitability for various applications should be considered. By examining the theoretical underpinnings of these techniques, this study aims to bridge existing gaps in understanding and offer insights into how these models can be optimized for specific use cases.

This research emphasizes the importance of systematically analyzing and categorizing machine learning approaches. It investigates models’ assumptions, limitations, and theoretical performance in handling high-dimensional, noisy, and dynamic data environments. Particular attention is given to exploring how hybrid and ensemble approaches might theoretically combine the strengths of individual algorithms to improve overall robustness and accuracy.

The article presents a comprehensive overview of methods and algorithms commonly employed for anomaly detection. These include statistical approaches such as Z-score analysis, Gaussian Mixture Models, and Kernel Density Estimation; popular clustering techniques like K-means, density-based clustering, and self-organizing maps; and advanced deep learning models, including autoencoders, Long Short-Term Memory networks, and Generative Adversarial Networks. Additionally, it explores hybrid and ensemble methods such as Isolation Forests, voting and stacking strategies, and boosting and bagging ensembles, showcasing a broad spectrum of solutions for detecting anomalies.

Within this research, the focus is extended to the interpretability of compound machine learning models, more specifically, neural networks. The investigation tries to theoretically enhance model transparency, enabling domain experts to understand the anomaly detection process. The study also attempts to elucidate how models may be integrated into dynamic environments characterized by streaming data or changing distributions. The main aim of the present study is to further expand the practitioners' knowledge of machine learning model architectures applied for anomaly detection and contribute a theoretical foundation for further studies and applications. These findings can assist in effectively guiding the composition of anomaly detection solutions that are more robust in different areas while addressing important issues, including interpretability, adaptability, and scalability.

Main research findings. The study provides a broad theoretical analysis of machine learning algorithms for anomaly detection, categorizing them into four groups: statistical approaches, clustering techniques, deep learning models, and hybrid/ensemble systems. Each category offers a distinct perspective on anomaly detection, from probabilistic assumptions to neural network power and diverse technique combinations.

Statistical methods are among the most foundational techniques for anomaly detection. They offer a mathematically grounded approach to identifying deviations from expected patterns. Statistical methods are particularly effective when data adheres to well-defined distributions or when the relationships among variables are linear and straightforward.

The Z-score method measures the number of standard deviations a data point deviates from the mean of the dataset [7]. A straightforward way to identify outliers can be provided by assigning a numerical score to each data point; typically, those exceeding a threshold (e.g., ±3 standard deviations) can be outliers. In financial auditing, Z-score analysis can highlight unusually high transactions that deviate significantly from typical transaction values, flagging potential cases of fraud or errors. It can also be utilized in manufacturing. The Z-score method can pinpoint defects or process anomalies, such as production units with dimensions outside normal ranges.

The Z-score method has drawbacks as it assumes a normal distribution of data, a condition which, most times, does not hold true with real-life data. When analyzing skewed or non-Gaussian distributions, such constrictions can be problematic. For instance, the impact of extreme values on the mean and the standard deviation affects the accuracy of the method. Modified forms of the Z-score method that use more resilient statistics, such as the median or the median absolute deviation or MAD, can be utilized to manage it.

Gaussian Mixture Models extend the principles of Z-score analysis by modeling data as a mixture of multiple Gaussian distributions. This probabilistic approach enables GMM to capture more complex data structures, accommodating scenarios where the dataset comprises subgroups with different means and variances. For example, in customer segmentation, GMM can model the purchasing behaviors of distinct groups (e.g., regular buyers and occasional shoppers). Anomalies, in turn, are data points with low likelihoods under the fitted mixture model. This makes GMM effective in applications where data is inherently multimodal, such as biological measurements or user behavior analysis.

GMM faces challenges when applied to high-dimensional datasets. The number of parameters to estimate increases exponentially with dimensionality, leading to computational inefficiencies and potential overfitting. Regularization techniques and dimensionality reduction methods, like principal component analysis (PCA), are often employed to mitigate issues.

Statistical methods can be classified into two categories: parametric methods and non-parametric methods. For example, Z-score and GMM are considered parametric methods because they are dependent on strong assumptions about the distribution from which data is drawn. Although these assumptions facilitate the modeling problem and make these methods efficient in computation, they narrow the usability to data that fits the shape of those distribution curves.

Kernel density estimation (KDE) and histograms are widely used non-parametric methods that operate without assuming an underlying distribution [8]. For example, KDE utilizes kernel functions to smooth observed data points and estimate the probability density function. Its feature is that it offers flexibility for datasets where distributions are complex or unknown. This adaptability makes non-parametric methods particularly valuable in modern anomaly detection scenarios. However, there are challenges to deploying such models in real-time applications because these methods require substantial data to produce reliable estimates and can be computationally demanding for large-scale datasets.

Statistical methods are widely applied in domains with a clear understanding of data distributions and interpretability. Examples include method utilization in healthcare, which identifies abnormal vital signs or lab test results based on established norms; environmental monitoring, which detects unusual weather patterns or pollution levels that deviate from historical averages.

The primary challenges of statistical approaches lie in their reliance on assumptions about data distribution and sensitivity to noise and outliers. As real-world datasets often show non-linear relationships, heavy tails, or multimodality, pure statistical methods will not be effective. In these cases, the necessity for using hybrid approaches or statistical methods with machine learning models arises for improved performance.

A promising area of exploration is a combination of statistical models with machine learning techniques, such as embedding GMMs into deep neural networks or integrating Z-score thresholds with ensemble methods. It helps to make the statistical approaches relevant in more advanced anomaly detection applications.

Clustering techniques are of significant importance in unsupervised methods of anomaly detection. These techniques leverage intrinsic patterns in the data to group points based on their similarities. In effect, they are able to organize similar data points based on natural structures, such as by making a model without providing labels or categories. Such a setup is useful for datasets where the pattern of anomalies is in stark contrast to the pattern of the regular points or where data is not labeled. Clustering methods identify anomalies as points that do not belong to any cluster or are located far from cluster centroids or dense regions.

K-means is one of the simplest and most widely used clustering algorithms. It partitions the dataset into a pre-defined number of clusters (k) by iteratively minimizing the variance within each cluster. In the context of anomaly detection, points that are far from their nearest cluster centroid are flagged as anomalies (Fig. 1).

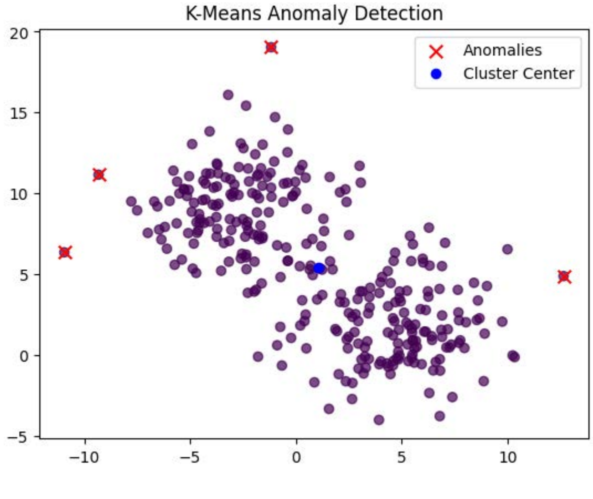

Fig. 1. K-means anomaly detection result

For example, in network traffic analysis, k-means can identify unusual data flow patterns, such as significant, infrequent file transfers that deviate from regular usage patterns. Another example is retail analytics, in which k-means can detect customers with atypical purchasing behavior that may indicate fraudulent transactions. It may look like k-means is simple and efficient, but this method has notable limitations. It assumes clusters are spherical and equally sized, which may not hold for many real-world datasets. Moreover, it is sensitive to the choice of k and initialization, which can lead to suboptimal clustering. Variants like k-means++ have been developed to improve initialization to address these issues, while other clustering techniques may be used for more complex data structures.

The density-based clustering algorithm called DBSCAN groups data points like the human eye. In essence, the algorithm identifies dense regions of data points as separated by sparser areas, and outliers, or anomalies, are defined as low-density areas that are not a part of any cluster. Clusters of different shapes and sizes improve the application in geospatial databases or industrial monitoring systems. Recent research demonstrated its utility in environmental applications, such as identifying anomalous weather patterns [9]. Compared to k-means, DBSCAN excels in detecting clusters of arbitrary shapes and automatically determining the number of clusters, making it adaptable for complex data distributions [10].

Hierarchical clustering builds a tree-like structure (dendrogram) to represent nested groupings of data points. There are two main approaches: agglomerative clustering, which starts with each data point as its cluster and merges them iteratively, and divisive clustering, which begins with all data points in a single cluster and splits them iteratively.

In anomaly detection, hierarchical clustering identifies outliers as points that either form clusters or appear at distant branches of the dendrogram. The method is particularly effective for visualizing cluster relationships and detecting anomalies in small to medium-sized datasets. The downside is the lack of a definitive stopping criterion, which complicates the process of determining the appropriate number of clusters. The method of hierarchical clustering assumes that each data point should be assigned to a cluster, and in case of a high proportion of anomalies, effectiveness declines.

Self-organizing maps (SOMs) are neural network-based methods that project high-dimensional data onto a low-dimensional grid while preserving the topological structure of the data [11]. Points that poorly fit the grid or significantly differ from their neighbors are identified as anomalies. SOMs are widely used for visualizing and analyzing complex datasets, including genomic data and financial transactions. Recently, SOMs were integrated with deep learning approaches, enhancing their scalability and accuracy for detecting anomalies in large-scale and intricate datasets.

Clustering methods face several challenges in anomaly detection. As the number of dimensions increases, distance becomes less meaningful, making it harder to identify clusters or anomalies. Dimensionality reduction techniques, such as PCA or t-SNE, are often required to preprocess data. Methods like DBSCAN and k-means depend on parameter choices (e.g., k, epsilon) that significantly impact performance. Clustering methods can be sensitive to noise in the dataset, mistaking noisy data points for anomalies or distorting cluster formation.

Deep learning models represent a transformative advancement in anomaly detection, offering unparalleled capabilities in handling complex, high-dimensional, and unstructured data. These models leverage the power of artificial neural networks to learn intricate patterns and relationships within data, enabling the identification of subtle anomalies that traditional methods might overlook. The flexibility and scalability of deep learning have positioned it as a key tool for anomaly detection in diverse applications, from image and video analysis to sequential data like time series and text.

Autoencoders, a type of unsupervised neural network, excel in encoding input data into a simplified representation and subsequently reconstructing it. High reconstruction error can be associated with anomalous data instances outside the learned pattern. For instance, autoencoders can help analysts study network traffic by determining if there are some unusual data exchanges that may be an indication that an intruder is present. Autoencoders are suited to the task of working on higher-level representations of data types like images, videos, or genomic datasets. Possessing such enhanced features, the networks, in particular the variational autoencoders VAEs who specialize in probabilistic modeling, offer not only anomaly but detection but also quantifying uncertainty [12].

LSTMs, a form of recurrent neural network, are adept at capturing long-term dependencies in sequential data, making them ideal for time-series analysis to detect anomalies, such as abrupt or gradual deviations in data patterns. In contexts like IoT, LSTMs can monitor sensor outputs to detect anomalies such as unusual spikes in readings. In the financial sector, these networks are invaluable for spotting atypical trading patterns or fraudulent activities using historical data. However, LSTMs do demand significant computational resources and are particularly sensitive to the settings of their hyperparameters [13].

CNNs, primarily utilized for spatial data interpretation, have shown effectiveness in anomaly detection within visual data by extracting and analyzing hierarchical features. In healthcare, CNNs are instrumental in identifying anomalies like tumors from MRI scans or irregularities in X-ray images. In manufacturing, they help in inspecting product images to detect flaws such as scratches or misalignments. Extensions like 3D CNNs broaden their applicability to dynamic contexts like monitoring video feeds in surveillance or traffic systems [14].

Generative Adversarial Networks (GANs) and VAEs are generative models that have gained widespread acceptance as tools for anomaly detection [15]. In a typical GAN, there exist two networks: the first is referred to as the generative network, which generates the data, while the second is called the discriminative network, which is used to classify the data into either real or synthetic. This approach has been useful in applications like computer security, where it can be used to detect fraud or spot defects in products. VAEs use deep learning neural networks and probabilistic methods and are capable of making independence and providing reconstruction alongside uncertainty about the resulting reconstruction. They treat abnormal signals as those that are highly unlikely to exist within the distribution that is modeled.

GAN and VAE models are very effective in capturing and modeling complex data but tend to be unstable during training, requiring intricate designs and careful parameter setting. As deep learning continues to mature, its application in anomaly detection systems should also broaden, thus providing more advanced solutions to complex detection problems.

To capitalize on the advantages of both approaches, hybrid deep learning models combine deep learning architectures with conventional techniques. For instance, when combined with statistical methods like Z-score analysis, autoencoders can make anomaly detection more effective by providing interpretable scores and reconstruction errors. Similarly, temporal patterns and density-based anomalies can be examined by combining LSTMs with clustering techniques such as DBSCAN.

Hybrid approaches have proven effective in domains where data complexity and variability require multiple layers of analysis. For example, a hybrid model might combine LSTM's temporal analysis with GANs in fraud detection to simulate standard transaction patterns [16].

One of the most significant advantages of deep learning is that it excels at handling complex datasets, such as videos, images, and text, where traditional methods struggle. Also, neural networks can be tailored to specific use cases, with architectures and layers optimized for different data types and patterns. Importantly, deep learning models automatically learn relevant features, saving time and improving accuracy, unlike traditional methods that require manual feature engineering.

The feature of deep learning models is that they require high computational power by themselves to perform, which often includes not only CPUs but also GPUs and memory. This combination makes it challenging to train and deploy them. Automatic feature detection using deep learning models in a dataset can also be ineffective because they are often criticized as “black boxes”, making it difficult to explain why a particular data point is classified as anomalous.

Future research in deep learning for anomaly detection focuses on improving interpretability, reducing data requirements through semi-supervised or unsupervised learning, and enhancing the scalability of models to handle massive datasets in real-time. Advances in hybrid models and domain-specific adaptations promise to expand the applicability and effectiveness of deep learning in increasingly complex and dynamic anomaly detection scenarios.

Hybrid and ensemble methods combine the strengths of multiple machine-learning approaches to achieve more robust and accurate anomaly detection. These methods address the limitations of individual techniques by integrating diverse algorithms or combining different types of data representations, leading to improved performance, generalization, and adaptability across various domains.

Integration of clustering algorithms and neural networks is a good practice. Clustering algorithms, such as k-means or DBSCAN, can group data into clusters to identify potential anomalies, while a neural network, like an LSTM, analyzes sequential dependencies within the clusters. This approach is commonly used in time-series analysis, such as detecting anomalies in IoT sensor data or energy consumption patterns.

Hybrid models are often customized for specific applications. For instance, in healthcare, clustering methods might identify subgroups of patient records, and a generative adversarial network (GAN) could simulate normal patterns within these groups to highlight unusual cases indicative of rare diseases [17].

Ensemble methods combine multiple models to enhance accuracy, reduce overfitting, and increase robustness. The two primary strategies in ensemble learning are bagging and boosting, but their adaptations for anomaly detection take on specialized forms: isolation forests, voting and stacking, boosting-based ensembles, and bagging ensembles for diversity.

Isolation Forests continue to be an effective ensemble-based anomaly detection method, leveraging multiple decision trees to isolate data points swiftly. The isolation forests method shines in high-dimensional data contexts and maintains computational efficiency, making it a top choice for real-time applications such as fraud detection or intrusion monitoring [18].

Voting ensembles, which aggregate predictions from a blend of statistical, clustering, and deep learning models, are complemented by stacking ensembles. These utilize a meta-model to optimize prediction combinations for enhanced anomaly detection accuracy. For instance, stacking ensembles might integrate outputs from autoencoders, k-means, and Gaussian Mixture Models (GMMs) to effectively pinpoint anomalies in complex manufacturing processes [19].

Recent advancements in gradient boosting techniques like XGBoost and LightGBM have spotlighted their adaptability in anomaly detection, particularly when dealing with imbalanced datasets where anomalies are infrequent. These methods excel by iteratively refining focus on hard-to-classify instances, improving the identification of anomalies in contexts such as product quality control [20].

Bagging methods remain robust, leveraging multiple models, each trained on random data subsets to increase noise and data variability tolerance. Enhanced algorithms like Random Forests are extensively applied in structured data anomaly detection scenarios, including monitoring credit card transactions or loan applications [21].

Hybrid and ensemble methods continue to evolve, proving invaluable in overcoming individual method limitations and achieving superior anomaly detection accuracy. These methods aid in mitigating model overfitting and reducing bias towards particular algorithms or datasets. Ensemble methods like Isolation Forests demonstrate impressive scalability and effectiveness across diverse datasets and application domains. Still, hybrid and ensemble methods have challenges. Combining hybrid or ensemble systems can be computationally intensive and require significant expertise. The combination of multiple algorithms often results in harder to interpret systems, posing challenges in sensitive applications like healthcare or finance. Hybrid systems require careful tuning of parameters across different components to achieve optimal performance.

Hybrid and ensemble methods are widely used in cybersecurity, healthcare, manufacturing, and energy sectors. In cybersecurity, advanced persistent threats are detected by combining anomaly scores from clustering, statistical analysis, and neural networks. Identifying rare diseases or adverse events in healthcare can be performed by integrating patient clustering, predictive models, and statistical anomaly scoring. For manufacturing, monitoring industrial processes can involve a combination of sensor data clustering with neural network-based anomaly detection.

Enhancing automation, scalability, and interpretability is key to the future of hybrid and ensemble approaches in anomaly detection. To lessen the need for human interaction, frameworks that automatically choose and combine the best algorithms for a specific dataset are being developed. The goal of explainable AI (XAI) developments is to increase the interpretability of ensemble and hybrid systems, guaranteeing responsibility and confidence in crucial applications.

Furthermore, it is anticipated that combining hybrid techniques with cutting-edge technologies like edge computing and federated learning would increase suitability in remote and resource-constrained settings. Hybrid and ensemble approaches will continue to be essential in addressing dynamic and increasingly complicated anomaly detection problems by utilizing these developments.

Conclusions and prospects for further research. The research was focused on theoretical approaches to some of the machine learning methods that use algorithms for the purpose of detecting anomalous events. The interest in the research is in understanding the basic ideas behind their design and construction, as well as their advantages and disadvantages. The research includes a variety of clustering, statistical, deep learning approaches, and hybrid/ensemble methods, evaluating their effectiveness across various domains and data types.

Statistical methods are used in the classification of deviations. They apply probabilistic models and outlier detection approaches to look for exceptional instances in structured datasets. Clustering algorithms provide insights into group-based anomaly detection. They show effectiveness in unstructured or spatial data scenarios, though they often demand meticulous parameter tuning. Deep learning models, such as autoencoders, LSTMs, and GANs, show good results in managing high-dimensional and complex datasets. Yet, they face challenges such as high computational demands and the necessity for large training datasets. Hybrid and ensemble methods have the advantages of multiple approaches to boost robustness and accuracy, mitigating the shortcomings of individual models and adapting effectively to dynamic and multifaceted anomaly detection tasks.

The study highlights the critical importance of choosing suitable models based on the specific properties of the data and the demands of the application domain. Statistical methods serve well for preliminary anomaly detection and quick assessment. In contrast, clustering techniques and deep learning models better detect less evident patterns. Hybrid and ensemble approaches provide a versatile solution, delivering flexibility and enhanced performance across various scenarios.

The article highlights the common issues that can be faced in the process of anomaly detection. Some of them include interpretability, scalability, and the ability to apply tired models across various domains. The theoretical findings of the topic provide a strong basis for addressing relevant problems in this area in the future, contributing to the advancement of model development, algorithm combination, and practical use.

Future research directions in anomaly detection should focus on developing methods that minimize computational demands, improve model interpretability, and adapt to changing data patterns. In this regard, there seems to be a good chance of pursuing semi-supervised and unsupervised learning approaches along with hybrid models and ensemble techniques to address the existing problems. In addition, the incorporation of emerging technologies such as federated learning and edge computing into the systems of anomaly detection may broaden their scope of application to distributed and resource-constrained environments.

This research contributes to the understanding of anomaly detection and its ability to address significant scientific and practical issues in the fields of industrial systems, cybersecurity, healthcare, and finance. As a vital tool for spotting uncommon and important events in more complicated data environments, machine learning-driven anomaly detection will continue to develop thanks to the insights gained.

REFERENCES

1. Pang, G., Shen, C., Cao, L., & Hengel, A. (2021). Deep learning for anomaly detection: A review. ACM Computing Surveys, 54 (1), 1–38. https://doi.org/10.1145/3439950.
2. Wu, H., Schafer, T.L.J., & Matteson, D.S. (2024). Trend and variance adaptive Bayesian changepoint analysis & local outlier scoring. arXiv. https://doi.org/10.48550/arXiv.2011.09437.
3. Zheng, H., Yu, X., & Li, J. (2021). Density-based neural networks for anomaly detection. Advances in Neural Information Processing Systems (NeurIPS).
4. Yang, C., Lan, S., Huang, W., Wang, W., Liu, G., Yang, H., Ma, W., & Li, P. (2022). A transformer-based GAN for anomaly detection. Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-031-15931-2_29.
5. Chabchoub, Y., Togbe, M., Boly, A., & Chiky, R. (2022). An in-depth study and improvement of Isolation Forest. IEEE Access. https://doi.org/10.1109/ACCESS.2022.3144425.
6. Noorchenarboo, M., & Grolinger, K. (2025). Explaining deep learning-based anomaly detection in energy consumption data by focusing on contextually relevant data. *Energy and Buildings*, 328, 115177. https://doi.org/10.1016/j.enbuild.2024.115177.
7. Rashid, F., Khan, R., & Qureshi, I. (2023). A comprehensive review of the Altman Z-score model across industries. SSRN. https://doi.org/10.2139/ssrn.5044057.

8. Amador Luna, D., Alonso-Chaves, F. M., & Fernández, C. (2024). Kernel density estimation for the interpretation of seismic big data in tectonics using QGIS: The Türkiye–Syria earthquakes (2023). Remote Sensing, 16 (20), 3849. https://doi.org/10.3390/rs16203849.
9. Wibisono, S., Anwar, M., Supriyanto, A., & Amin, I. (2021). Multivariate weather anomaly detection using DBSCAN clustering algorithm. *Journal of Physics: Conference Series*, 1869, 012077. https://doi.org/10.1088/1742-6596/1869/1/012077.
10. Li, Y., Wang, J., Zhao, H., Wang, C., & Shao, Q. (2024). Adaptive DBSCAN clustering and GASA optimization for underdetermined mixing matrix estimation in fault diagnosis of reciprocating compressors. *Sensors*, 24 (1), 167. https://doi.org/10.3390/s24010167
11. Licen, S., Astel, A., & Tsakovski, S. (2023). Self-organizing map algorithm for assessing spatial and temporal patterns of pollutants in environmental compartments: A review. Science of The Total Environment, 878, 163084. https://doi.org/10.1016/j.scitotenv.2023.163084.
12. Gorman, M., Ding, X., Maguire, L., & Coyle, D. (2023). Convolutional autoencoders for anomaly detection in semiconductor manufacturing. IEEE AI for Cybersecurity (AICS), 1–6. https://doi.org/10.1109/AICS60730.2023.10470831.
13. Arifin, S., Wijaya, A., Nariswari, R., Yudistira, A., Suwarno, Faisal, F., & Wihardini, D. (2023). Long short-term memory (LSTM): Trends and future research potential. *International Journal of Emerging Technology and Advanced Engineering*, 13, 24–35. https://doi.org/10.46338/ijetae0523_04.
14. Saifullah, S., Yuwono, B., Rustamaji, H.C., Saputra, B., Dwiyanto, F.A., & Dreżewski, R. (2023). Detection of chest X-ray abnormalities using CNN based on hyperparameter optimization. Engineering Proceedings, 56 (1), 223. https://doi.org/10.3390/ASEC2023-16260.
15. Zenati, H., Foo, C.S., Lecouat, B., Manek, G., & Chandrasekhar, V.R. (2019). Efficient GAN-based anomaly detection. arXiv. https://doi.org/10.48550/arXiv.1802.06222.
16. Bousmina, A., Selmi, M., Rhaiem, M., & Farah, I. (2023). A hybrid approach based on GAN and CNN-LSTM for aerial activity recognition. Remote Sensing, 15, 3626. https://doi.org/10.3390/rs15143626.
17. Purandhar, N., Ayyasamy, & Poruran, S. (2022). Classification of clustered health care data analysis using generative adversarial networks (GAN). Soft Computing. https://doi.org/10.1007/s00500-022-07026-7.
18. Hariri, S., Kind, M.C., & Brunner, R.J. (2021). Extended isolation forest. IEEE Transactions on Knowledge and Data Engineering, 33 (4), 1479–1489. https://doi.org/10.1109/TKDE.2019.2947676.
19. Lazzarini, R., Tianfield, H., & Charissis, V. (2023). A stacking ensemble of deep learning models for IoT intrusion detection. Knowledge-Based Systems, 279, 110941. https://doi.org/10.1016/j.knosys.2023.110941.
20. Vozza, M., Polden, J., Mattera, G., Piscopo, G., Vespoli, S., & Nele, L. (2024). Explaining anomaly detection in additive manufacturing via boosting models and frequency analysis. Mathematics, 12, 3414. https://doi.org/10.3390/math12213414.
21. Schonlau, M., & Zou, R. (2020). The random forest algorithm for statistical learning. *The Stata Journal: Promoting Communications on Statistics and Stata*, 20, 3–29. https://doi.org/10.1177/1536867X20909688.