IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

Exploring Techniques and Applications for Anomaly Detection in Time Series Data

Temitope, Olubunmi. Awodiji[1], John Owoyemi[2], Edeamah, O. Jonah[3]

Department of Information Security, University of Cumberlands, Kentucky, USA[1]

University of the Cumberlands, Williamsburg, Kentucky[2]

Department of Learning and Resource Center, San Diego Mesa College, San Diego, CA, USA[3]

Abstract: Anomaly detection in time series data is a critical task with numerous applications across various domains. This study presents a comprehensive empirical review that explores the methodologies, evaluation metrics, benchmark datasets, novel techniques, and real-world case studies in anomaly detection. The study begins by examining the different methodologies employed in anomaly detection, including statistical-based methods, machine learning-based approaches, and deep learning-based techniques. Evaluation metrics such as precision, Recall, F1-score, ROC curve, and AUC are discussed, along with commonly used benchmark datasets that serve as standards for evaluating anomaly detection algorithms. Novel techniques and algorithms for anomaly detection in time series data are critically analyzed, including time series decomposition and reconstruction methods, transfer learning and domain adaptation techniques, online and streaming anomaly detection approaches, and ensemble methods. These techniques' strengths, limitations, and potential applications are discussed in detail.

Real-world case studies showcase the practical applications of anomaly detection in different domains. These case studies include anomaly detection in network traffic data, energy consumption patterns, and medical sensor data. The mathematical approaches and algorithms employed in these case studies are examined to provide insights into the specific methodologies used for anomaly detection. Future research directions and challenges in anomaly detection are discussed, highlighting the importance of explainable AI and interpretable models, incorporating domain knowledge and context awareness, privacy-preserving techniques, and integrating anomaly detection with other data analysis techniques. This empirical review contributes to the field by comprehensively analyzing anomaly detection techniques in time series data. The study's findings offer valuable insights into the strengths and limitations of different methodologies and highlight emerging trends in the field. Furthermore, future research directions and challenges provide a roadmap for advancing anomaly detection in time series data analysis.

Keywords: anomaly detection; time series; cybersecurity; neural network; machine learning

I. INTRODUCTION

Anomaly detection in time series data is crucial in various domains, including cybersecurity, finance, industrial IoT, healthcare, and environmental monitoring. It involves the identification of unusual patterns, deviations, or outliers that deviate significantly from the expected behavior of the time series data. The need for effective anomaly detection techniques in time series data has grown rapidly with the increasing availability of large-scale datasets and the emergence of sophisticated data analysis methods. Anomalies in time series data can signify critical events, such as network intrusions, financial fraud, equipment failures, or health abnormalities. Timely detection and accurate identification of such anomalies are essential for proactive decision-making, risk mitigation, and ensuring the integrity and security of systems and processes. While significant progress has been made in the field of anomaly detection, there are still several research gaps and challenges that need to be addressed. Traditional statistical-based methods, such as moving averages, standard deviation analysis, and z-score analysis, have limitations in handling complex and nonlinear time series data. On the other hand, machine learning approaches, including supervised and unsupervised learning algorithms and deep learning models, offer promising avenues for anomaly detection. However, developing robust and scalable techniques to handle the specific characteristics of time series data remains an ongoing research focus. This study aims to explore various techniques and applications for anomaly detection in time series data. It will provide a platform for researchers, practitioners, and domain experts to share their insights, advancements, and case studies in this field. The paper will foster interdisciplinary collaboration and encourage the development of novel algorithms, evaluation metrics, and benchmark datasets to facilitate the progress of anomaly detection in time series analysis. By addressing the challenges and limitations of existing approaches, this paper seeks to contribute to developing more accurate, efficient, and interpretable anomaly detection techniques.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

It will also highlight emerging trends and future directions, such as explainable AI, privacy-preserving techniques, and integration with other data analysis methods. Ultimately, the study aims to advance the state-of-the-art in anomaly detection and promote its application in real-world scenarios, leading to improved decision-making, system reliability, and anomaly detection performance in diverse domains.

II. RELATED WORKS

In the field of machine learning known as "Deep Learning," algorithms and models are modelled after the composition and operation of neural networks seen in the human brain [1]. It entails training multiple-layer artificial neural networks to learn and extract patterns and characteristics from enormous volumes of data [47], allowing the model to make accurate predictions or perform complex tasks.

Convolutional Neural Network (CNN): A convolutional neural network is a deep learning model designed for processing and analysing visual data, such as images or videos [2]. CNNs are particularly effective in tasks like image classification and object recognition. They consist of multiple layers, including convolutional layers that extract features from input images, pooling layers that down-sample the extracted features, and fully connected layers for classification or regression [3].

Support Vector Machine (SVM): SVM is a classification and regression analysis supervised machine learning technique. It operates by determining the best hyperplane for separating distinct classes or predicting continuous values [4]. SVM aims to maximize the margin between the support vectors, which are the data points closest to the decision boundary. Using kernel functions, it can handle both linear and nonlinear classification problems.

Internet of Things (IoT): The Internet of Things (IoT) is a network of actual physical items or "things" that can communicate with one another and with people via connectivity, software, and data exchange [5]. These items could be commonplace gadgets like household appliances, automobiles, or wearable technology. IoT makes it possible for these things to gather and transmit data online, enabling automation, remote monitoring, and cutting-edge data analytics for a range of applications.

Autoencoder: An autoencoder is a neural network for unsupervised learning and dimensionality reduction. It consists of an encoder that compresses the input data into a lower-dimensional representation (encoding) and a decoder that reconstructs the original input data from the encoded representation [6]. Autoencoders are often used for data compression, anomaly detection, and feature extraction tasks.

III. METHODOLOGIES FOR ANOMALY DETECTION

In this study, three classes of methodologies for anomaly detection are examined. These classes are statistical-based methods, machine-learning methods, and hybrid methods.

Statistical-based Method

A statistical-based method refers to an approach that utilizes statistical principles and techniques to analyze data and make inferences or decisions [7]. In the context of anomaly detection, statistical-based methods use statistical measures and assumptions to identify anomalies in time series data. These methods typically use statistical concepts such as mean, standard deviation, variance, and distribution properties to characterize the data and detect anomalies.

By quantifying the typical behavior of the data, statistical-based methods can identify data points or patterns that deviate significantly from the expected or normal behavior. Common statistical-based techniques for anomaly detection in time series data include moving averages, standard deviation analysis, z-score analysis, hypothesis testing, and distribution fitting [8]. These methods establish a baseline or reference behavior based on the statistical properties of the data and flag data points or patterns that deviate beyond a certain threshold.

Moving averages

Moving averages are simple and widely used statistical-based techniques for anomaly detection. They involve calculating the average value of a time series over a specified window or period. Anomalies are detected by comparing data points to the moving average and identifying significant deviations. Moving averages are particularly effective in detecting anomalies that result in sudden shifts or changes in the data [9]. Fig. 1 depicts the anomaly in a moving average.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

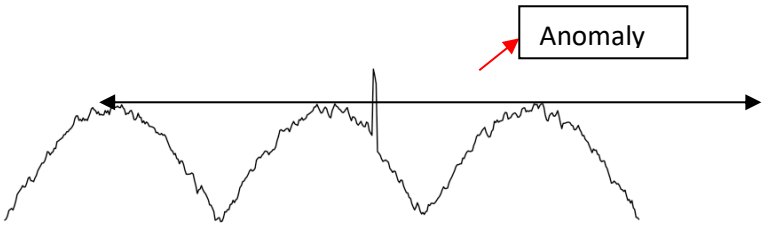

Figure 1. Anomaly Detection in Moving Average

However, the limitation of moving averages is their sensitivity to sudden changes in the time series data. They may not perform well when anomalies occur within the window size or when the data has high variability, volatility, or seasonality. Standard Deviation-based Techniques Standard deviation-based techniques utilize the concept of standard deviation, which measures the dispersion of data points around the mean [10]. By defining a threshold as a certain number of standard deviations from the mean, anomalies can be identified as data points that fall outside this threshold. Standard deviation-based methods are useful in identifying anomalies that exhibit unusually high or low values compared to the mean [11]. However, they assume the data follows a normal distribution, which may not always hold for complex time series data. Additionally, setting an appropriate threshold can be challenging, as it depends on the specific characteristics of the data and the desired trade-off between false positives and false negatives. For example, Fig. 2 below depicts how almost impossible values can be considered anomalous, especially when the outlier deviates too high or low from the mean. This limit can be calculated using the formula:

$$s_d = \sqrt{s^2}x \qquad \text{…Equation 1}$$

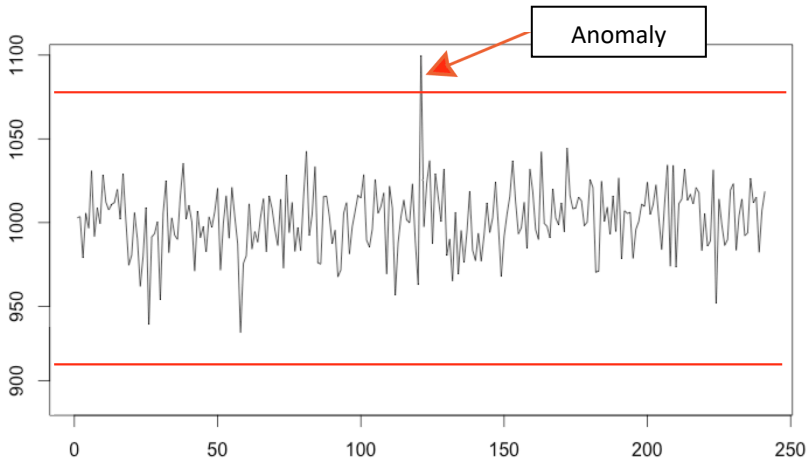

Figure 2. Anomaly Detection in Standard-deviation

Z-score Analysis

The Z-score analysis is similar to standard deviation-based techniques but involves standardizing the data by subtracting the mean and dividing by the standard deviation [12]. This transformation results in z-scores, which indicate how many standard deviations a data point deviates from the mean. Anomalies are identified by setting a threshold on the z-scores. Data points with z-scores above or below the threshold are considered anomalous. The Z-score analysis is advantageous as it accounts for the data's mean and standard deviation, allowing for a more nuanced detection of anomalies [13]. However, like standard deviation-based techniques, the z-score analysis assumes a normal distribution and may not perform well with non-normal data. Determining an appropriate threshold can also be challenging, and domain expertise may be required to set meaningful thresholds. The Z-score is unitless with a mean of 0 and S_d 1. Below is the equation for this method.

$$z = \frac{x - \bar{x}}{s} \qquad \text{…Equation 2}$$

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

IV. MACHINE LEARNING APPROACHES

Machine Learning (ML) algorithms are programs that can learn the hidden patterns from the data, predict the output, and improve the performance from experiences on their own. Different algorithms can be used in machine learning for different tasks, such as simple linear regression that can be used for prediction problems like stock market prediction, and the KNN algorithm can be used for classification problems [14]. Explained below are the three broad classifications of ML.

Supervised Learning Algorithms

Supervised learning algorithms require labeled data, where anomalies are explicitly identified, to train a model. The model learns to differentiate between normal and anomalous patterns based on the provided labels. Various supervised learning algorithms, such as support vector machines (SVM), random forests, and neural networks, can be employed for anomaly detection. Supervised learning approaches can be effective when labeled data is available, allowing the model to learn the specific characteristics of anomalies [15]. However, obtaining labeled data can be time-consuming and costly, especially for rare or complex anomalies. Supervised models may also struggle with detecting previously unseen or novel anomalies that differ significantly from the training data.

Unsupervised Learning Algorithms

Unsupervised learning algorithms operate on unlabelled data and aim to learn patterns inherent in the data without prior knowledge of anomalies [16]. These algorithms identify anomalies as data points that deviate significantly from the learned patterns. Unsupervised approaches, including clustering algorithms (e.g., k-means) and Gaussian mixture models, are widely used for anomaly detection. They can detect anomalies by identifying data points that do not conform to the expected clusters or distributions. Autoencoders, a type of neural network, are also popular unsupervised models for anomaly detection in time series data. They learn to reconstruct the input data and flag instances with high reconstruction errors as anomalies. Unsupervised learning algorithms are advantageous when labeled data is scarce or unavailable. They can discover anomalies without prior knowledge, making them suitable for detecting novel or emerging anomalies [17]. However, unsupervised models may generate false positives or struggle to distinguish between different types of anomalies, especially in complex and high-dimensional time series data.

Deep Learning Models

Deep learning models, such as recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and convolutional neural networks (CNNs), are well-suited for capturing complex temporal dependencies and patterns in time series data [18]. RNNs and LSTMs can retain information over time, effectively detecting anomalies that span multiple time steps. They can learn temporal patterns and identify deviations from the learned behavior. CNNs excel in capturing local patterns and spatial relationships in time series data [19]. Deep learning models have shown promising results in anomaly detection, especially for complex and high-dimensional data. However, they require a significant amount of labeled or unlabelled data for training and may be computationally intensive. The below diagram illustrates the different ML algorithms, along with the categories:

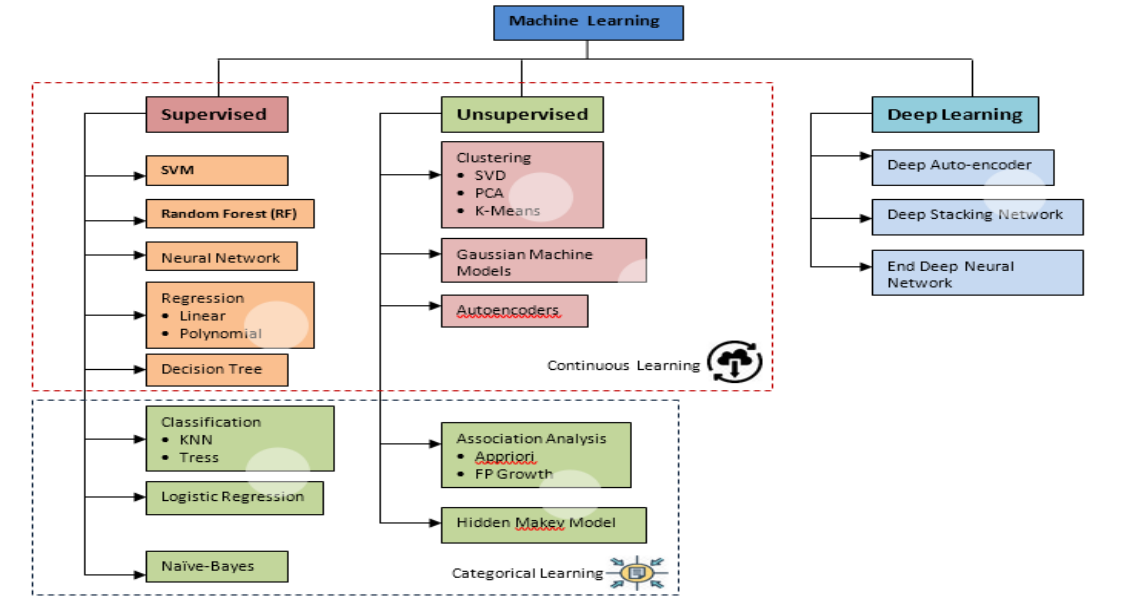

Figure 3. Machine Learning Methods and Classifications

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

V. HYBRID TECHNIQUES (COMBINING STATISTICAL AND MACHINE LEARNING METHODS)

Hybrid techniques combine the strengths of statistical-based methods and machine learning algorithms to enhance anomaly detection performance [20]. For example, statistical methods can be used as a preprocessing step to identify initial candidates for anomalies. These candidates can then be further refined or classified using machine learning algorithms. This hybrid approach benefits from statistical techniques' simplicity and interpretability while leveraging machine learning models' power to handle more complex patterns. Ensemble methods, another hybrid technique, combine multiple anomaly detection models [21]. This includes both statistical-based and machine-learning models to improve overall detection performance. Ensemble methods can leverage the diverse strengths of individual models and provide more robust anomaly detection results [22].

By integrating statistical-based methods with machine learning techniques, hybrid approaches aim to overcome the limitations of each method individually and achieve improved accuracy, interpretability, and robustness in anomaly detection. It is important to note that selecting the appropriate methodology depends on the specific characteristics of the time series data, the availability of labeled or unlabelled data, and the desired trade-offs in terms of accuracy, interpretability, and computational complexity.

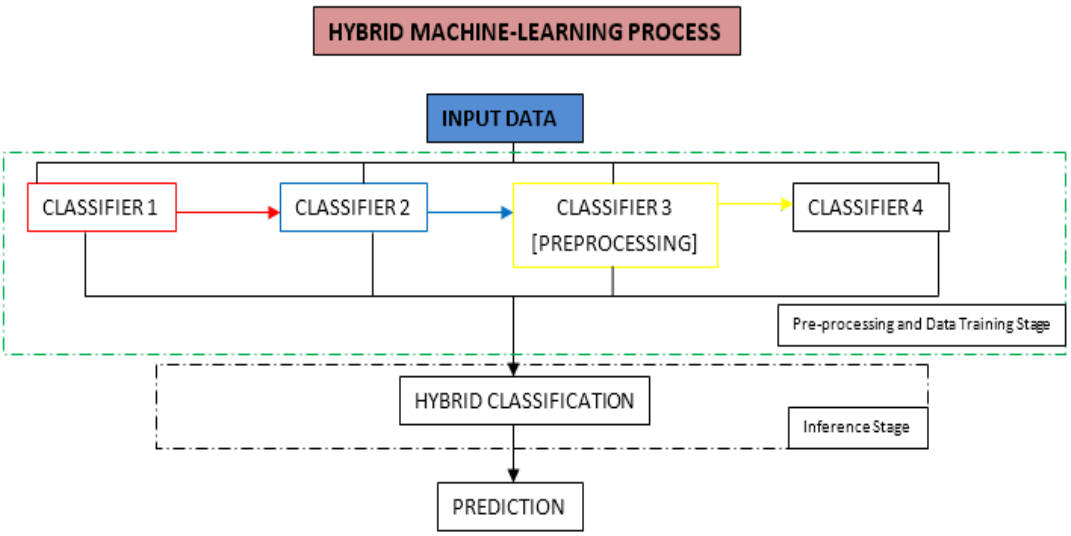

Figure 4. Hybrid Machine Learning Process (Data Training and Inference Stage)

VI. APPLICATION OF ANOMALY DETECTION IN TIME SERIES

The application of anomaly detection in time series can be carried out in various fields, including cybersecurity, industrial IoT and predictive maintenance, finance and stock market analysis, health care monitoring, and environmental monitoring. The application of anomaly detection serves as the most pertinent task of detecting variance and anomaly in a series. In this study, the evaluation of the application would be designated to cybersecurity, industrial IoT, and predictive maintenance. In the same vein, the advantages and challenges of the application in selected fields will be explored.

Application 1: Cybersecurity

Anomaly detection in time series data plays a vital role in cybersecurity by identifying malicious activities, network intrusions, and cyber threats. Time series data in cybersecurity include network traffic logs, system logs, user behaviour logs, and other relevant data sources.

Early Threat Detection

Anomaly detection methods in cybersecurity often utilize statistical and machine learning techniques to identify deviations from normal patterns in time series data. These methods may include:

Statistical-based methods: Moving averages, standard deviation analysis, and z-score analysis can be applied to network traffic or system logs to detect unusual spikes or drops in activity.

Machine learning approaches: Supervised learning algorithms, such as support vector machines (SVM), random forests, or neural networks, can be trained on labeled data to recognize patterns associated with specific types of cyber threats.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

Anomaly detection algorithms are often trained on past data to learn the system's normal behavior to spot threats early. Incoming data is compared to learned patterns throughout the deployment, and substantial variations are reported as probable abnormalities [23]. To handle and evaluate data in real-time, real-time anomaly detection systems may use streaming algorithms such as sliding windows or online learning.

Zero-Day Exploit Detection

Anomaly detection can play a crucial role in identifying zero-day exploits, previously unknown vulnerabilities, or attack methods. Zero-day exploits often evade traditional signature-based detection systems. Anomaly detection methods focus on detecting deviations from normal behavior that could indicate the presence of an exploit [24]. These methods can include:

Unsupervised learning algorithms: Clustering techniques, such as k-means or Gaussian mixture models, can identify unusual patterns or clusters of activity that deviate from the system's normal behavior.
Deep learning models: Recurrent neural networks (RNNs) or LSTM networks can capture temporal dependencies and identify anomalous sequences of events or network traffic patterns that may indicate zero-day exploits.

Reduced False Positives

Reducing false positives is a crucial aspect of effective anomaly detection in cybersecurity. False positives can result from noisy data, legitimate but rare activities, or legitimate variations in the system behavior [25]. Techniques to address false positives include:

Thresholding: Setting appropriate thresholds for anomaly scores or statistical measures to balance sensitivity and specificity.

Ensemble methods: Combining multiple anomaly detection models to improve overall detection performance and reduce false positives.

Feature engineering: Extracting relevant features from the time series data and applying dimensionality reduction techniques to focus on the most discriminative features for anomaly detection.

To address imbalanced data, where anomalies are rare compared to normal instances, techniques like oversampling anomalies, under-sampling of normal instances, or synthetic data generation can be employed to rebalance the dataset and improve the detection of rare anomalies.

Application 2: Industrial IoT and Predictive Maintenance:

Industrial IoT leverages interconnected devices and sensors to monitor and optimize industrial processes. Anomaly detection in time series data is critical for predictive maintenance, enabling the early detection of equipment failures, performance degradation, and abnormal operating conditions.

Preventive Maintenance:

Anomaly detection plays a vital role in predictive maintenance by identifying potential equipment failures or degradation before they cause significant disruptions. Various techniques can be employed in this context, including those explained below.

Sensor-based anomaly detection: Monitoring sensor data from industrial equipment and identifying deviations from normal operating patterns [26]. This can involve statistical methods like moving averages or machine learning models such as SVM or LSTM networks trained on historical sensor data.

Multivariate analysis: Analyzing multiple sensor streams collectively to identify correlated anomalies or abnormalities across different variables. Techniques like principal component analysis (PCA) or multivariate Gaussian models can be used.

Cost Reduction:

Anomaly detection in predictive maintenance helps optimize maintenance schedules, reduce unplanned downtime, and minimize repair costs. By detecting anomalies early, maintenance activities can be planned proactively, reducing the likelihood of equipment failures and associated costs. This requires:

Accurate anomaly detection: Reliable detection of anomalies in time series data, allowing maintenance teams to intervene before the situation worsens.

Predictive modeling: Develop predictive models that estimate the remaining useful life of equipment, enabling proactive maintenance actions.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

Equipment Health Monitoring:
Anomaly detection techniques continuously monitor equipment health and detect deviations from normal behavior. This involves:

Real-time monitoring: Processing streaming sensor data in real-time to detect anomalies and trigger maintenance actions promptly.

Predictive analytics: Combining historical data with real-time information to predict future equipment behavior and identify potential anomalies.
Feature extraction: Extracting relevant features from sensor data, such as statistical measures, spectral analysis, or wavelet transforms, to capture different aspects of equipment health and behavior.

Application Challenges and Research Areas:
Data dimensionality and scalability: Industrial IoT generates massive amounts of high-dimensional time series data. Thus, efficient techniques for handling large-scale data, feature selection, and dimensionality reduction are essential.

Adaptability and robustness: Anomaly detection models need to adapt to changing operational conditions, equipment variations, and evolving attack strategies [27]. Developing adaptive and resilient models is an ongoing research area.

Explainability and interpretability: Interpreting the reasons behind detected anomalies or providing contextual explanations for system behavior is crucial for effective decision-making and trust in the anomaly detection system.

In summary, anomaly detection in time series data for cybersecurity and industrial IoT applications involves a range of statistical, machine learning, and deep learning techniques [28]. Addressing challenges such as imbalanced data, false positives, data noise, and interpretability is crucial to improve the accuracy and effectiveness of anomaly detection systems in these domains.

VII. ANOMALY DETECTION EVALUATION METRICS AND BENCHMARK DATASETS

Evaluation metrics and datasets used in this study are categorized into two distinct classes, whereby each class contains a group of metrics and datasets. Class 1 includes Precision, Recall, and F1-Score metrics; while Class 2 includes Receiver Operating Characteristics (ROC) Curve and Area Under the Curve (AUC). This section provides both quantitative and qualitative analysis of these metrics. The purpose of evaluation metrics herein is to quantitatively assess the performance of anomaly detection techniques in terms of precision, Recall, F1-score, ROC curve, and AUC, providing insights into their effectiveness and trade-offs. Benchmark datasets serve as standardized qualitative references to facilitate fair comparisons and evaluations of different anomaly detection methods, ensuring consistency and reproducibility in the research study.

Precision, Recall, and F1-Score
Quantitative Analysis
Precision: Precision measures the proportion of correctly identified anomalies out of all instances classified as anomalies [29]. It focuses on the accuracy of the positive predictions.

Recall: Recall (also known as sensitivity or true positive rate) measures the proportion of correctly identified anomalies out of all actual anomalies [30]. It focuses on the ability to detect anomalies.

F1-Score: The F1-score combines precision and recall into a single metric, providing a balanced measure of a model's performance [31]. It is the harmonic mean of precision and Recall.

Qualitative Analysis
Precision: A high precision indicates a low false positive (FP) rate, meaning the model is effective at identifying true positive (TP) anomalies and avoiding false alarms.

Recall: A high recall indicates a low false negative (FN) rate, meaning the model is effective at detecting most of the true positive (TP) anomalies present in the data.

F1-Score: The F1-score considers both precision and Recall, providing a balanced view of a model's performance. It is useful when both false positives and false negatives are equally important.

Criticism
Precision, Recall, and F1-score are essential evaluation metrics for anomaly detection, particularly when dealing with imbalanced datasets where anomalies are rare.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

They provide valuable insights into the model's ability to identify anomalies and control false positives. However, precision and recall alone may not provide a complete picture of a model's performance. Fig. 5 is a depiction of how Precision and Recall for Time Series are used in Point-Based and Range-Based Anomalies to measure accuracy. Precision (p) is therefore calculated as:

$$p = TP / (TP+FP) \quad \text{…Equation 3}$$

While Recall (*r*) can be calculated as:

$$r = TP / (TP + FN) \quad \text{…Equation 4}$$

Depending on the specific application, there may be different priorities regarding false positives and false negatives. Therefore, it is crucial to consider other metrics, such as the ROC curve and AUC, for a more comprehensive evaluation.

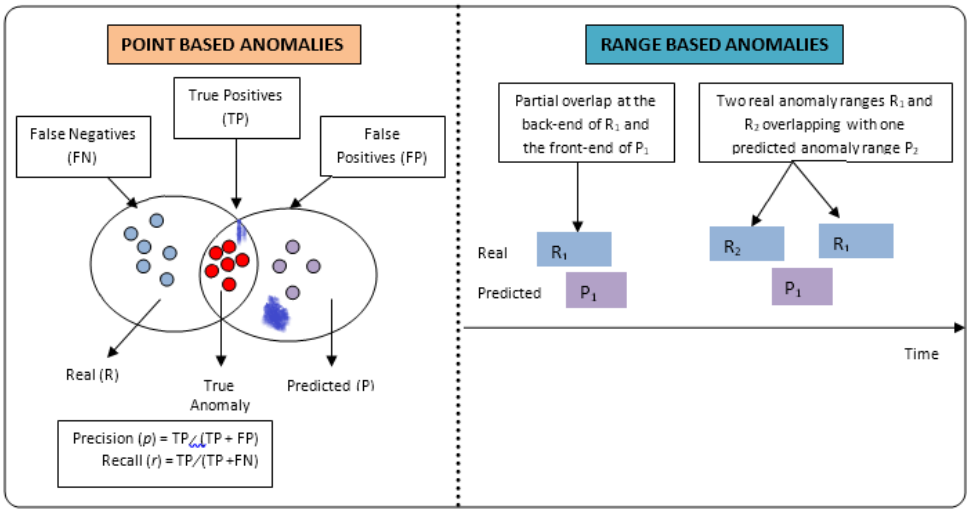

Figure 5 Precision and Recall for Time Series in Point-Based and Range-Based Anomalies

Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC)
Quantitative Analysis

- ROC Curve: The ROC curve is a graphical representation of a model's performance as the discrimination threshold is varied [32]. It plots the true positive rate (Recall) against the false positive rate (1-specificity) at various threshold values.
- AUC: The Area Under the Curve represents the overall performance of the model, providing a single scalar value [33]. It quantifies the model's ability to distinguish between anomalies and normal instances, irrespective of the chosen threshold.

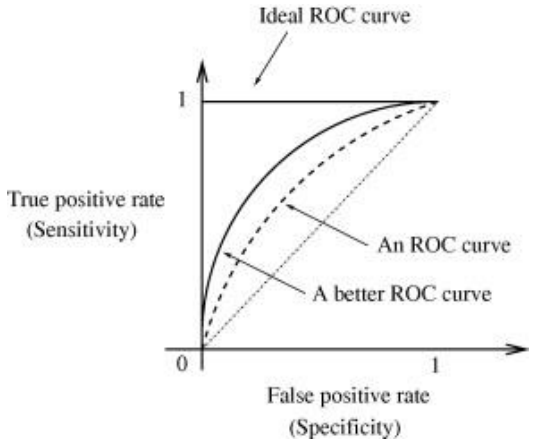

Figure 6. ROC Curve Sample

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

Qualitative Analysis
ROC Curve: The ROC curve allows for visual assessment of the trade-off between true positive rate and false positive rate across different threshold values. It helps in selecting an appropriate threshold based on the desired trade-off in anomaly detection.

AUC: A high AUC indicates a model with better discrimination ability, meaning it has a higher probability of ranking anomalies higher than normal instances.

Criticism
The ROC curve and AUC are widely used in anomaly detection because they are threshold-independent and provide a comprehensive evaluation of model performance [34]. However, the ROC curve and AUC do not consider the imbalance in the dataset and may not provide detailed insights into the model's performance at specific operating points. It is important to interpret the ROC curve and AUC in conjunction with other evaluation metrics, such as precision, Recall, and F1-score, to gain a complete understanding of the model's effectiveness.

Commonly Used Benchmark Datasets for Evaluating Anomaly Detection Techniques:
Quantitative Analysis
Benchmark datasets provide standardized data for evaluating and comparing anomaly detection methods. These datasets often contain labeled instances where anomalies are explicitly identified, allowing for supervised or semi-supervised evaluation.

Qualitative Analysis
Common benchmark datasets include:
KDD Cup 1999: A network intrusion detection dataset widely used for evaluating cybersecurity anomaly detection techniques.

Numenta Anomaly Benchmark (NAB): A collection of time series datasets with labeled anomalies covering various domains such as temperature, CPU usage, and stock prices.

IEEE CIS Fraud Detection: A dataset focusing on credit card fraud detection containing anonymized transaction data with labeled anomalies.

SMD: The Server Machine Dataset, which comprises time series data from servers and includes various anomalies related to hardware, software, and network issues.

Criticism
Benchmark datasets provide a standardized basis for comparing different anomaly detection methods. However, it is essential to recognize those benchmark datasets may not fully represent the complexity and diversity of real-world anomaly detection scenarios [35]. Real-world data often presents challenges like concept drift, varying data distributions, and contextual dependencies that may not be captured by benchmark datasets [36]. It is crucial to validate the performance of anomaly detection techniques on domain-specific datasets or real-world data to assess their effectiveness in practical applications.

In conclusion, precision, Recall, F1-score, ROC curve, AUC, and benchmark datasets are valuable tools for evaluating anomaly detection techniques. While these metrics provide quantitative measures of model performance, it is crucial to consider their qualitative implications and critically analyze their limitations to ensure a comprehensive evaluation of anomaly detection methods.

VIII. NOVEL TECHNIQUES AND ALGORITHMS FOR ANOMALY DETECTION

For novel techniques and algorithms, this study will briefly examine four significant methods, including Time Series Decomposition and Reconstruction methods, Transfer Learning and Domain Adaptation methods, Online and Streaming Anomaly Detection methods, and Ensemble methods.

Time series decomposition and reconstruction methods
Time series decomposition involves breaking down a time series into its constituent components, such as trend, seasonality, and residual [37]. Reconstruction methods aim to reconstruct the time series using these decomposed components.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

In the context of anomaly detection, these techniques can be utilized to identify deviations in the reconstructed time series compared to the original data, indicating the presence of anomalies [38]. Examples of decomposition methods include Singular Spectrum Analysis (SSA), Empirical Mode Decomposition (EMD), and Seasonal and Trend decomposition using LOESS (STL). These methods offer advantages in anomaly detection, such as the following:

Improved anomaly detection: By analyzing individual components, anomalies specific to trend, seasonality, or residual can be identified more accurately.

Enhanced interpretability: Decomposition provides insights into the underlying patterns, allowing for a better understanding and interpretation of anomalies.

Noise reduction: Decomposition techniques can filter out noise or irrelevant patterns, focusing the analysis on the most meaningful components.

Transfer learning and domain adaptation techniques

Transfer learning and domain adaptation techniques involve leveraging knowledge or models learned from a source domain to improve anomaly detection performance in a target domain. In anomaly detection, these techniques can be employed to transfer knowledge about normal patterns or feature representations from a labeled source domain to an unlabelled or sparsely labeled target domain [39]. This allows for better adaptation to the target domain's characteristics, reducing the need for extensive labeled data. Techniques like domain adaptation, domain alignment, and feature adaptation can be applied to facilitate effective anomaly detection in different domains.

Key advantages of transfer learning and domain adaptation in anomaly detection include:

Utilization of existing knowledge: Transfer learning allows the use of labeled data or learned representations from a source domain to enhance the detection of anomalies in the target domain, even with limited labeled data.

Adaptability to new environments: Domain adaptation techniques help adapt anomaly detection models to the target domain's characteristics, addressing the domain shift problem.

Reduction of data labeling effort: By leveraging knowledge from a related domain, transfer learning reduces the need for extensive labeled data in the target domain.

Online and streaming anomaly detection approaches

Online and streaming anomaly detection approaches are designed to detect anomalies in real-time as data streams arrive continuously. These techniques enable the detection and response to anomalies in dynamic and evolving systems without requiring batch processing or access to the entire historical data [40]. Online anomaly detection methods often utilize sliding windows, sequential models, or adaptive algorithms to continuously update and refine anomaly detection models as new data arrives. This allows for timely detection and response to anomalies, making them suitable for time-critical applications [41]. They typically involve the following aspects described below:

Sliding window: Online methods utilize sliding windows to maintain a fixed-sized window of recent data, continuously updating the anomaly detection model as new data arrives [42].

Sequential modeling: Techniques like recurrent neural networks (RNNs) or hidden Markov models (HMMs) capture temporal dependencies and allow for sequential modeling of the data stream [43].

Adaptive algorithms: Online anomaly detection methods often incorporate adaptive algorithms that can dynamically adjust model parameters or thresholds as the data distribution evolves [44].

Benefits of online and streaming anomaly detection include:

Real-time detection: These methods enable timely detection and response to anomalies, reducing the impact of disruptive events.

Efficiency: By processing data in real time and focusing on recent observations, online methods can handle high-velocity data streams efficiently.

Adaptability to dynamic environments: The ability to update the anomaly detection model continuously allows for adaptation to evolving data distributions and changing anomaly patterns.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

Ensemble methods

Ensemble methods involve combining multiple individual anomaly detection models or algorithms to improve overall detection performance. Anomaly detection ensembles aim to leverage the diversity and complementary strengths of different models to enhance accuracy, robustness, and generalization [45]. Ensemble methods can be based on different principles, such as combining outputs from individual models through voting, averaging, or stacking [46]. By aggregating the predictions of multiple models, ensemble approaches can reduce false positives, increase sensitivity to anomalies, and provide more reliable and robust anomaly detection results. These methods can be based on various principles:

Voting: Individual models provide their predictions, and the final decision is made based on majority voting or weighted voting.

Averaging: The predictions of individual models are averaged to obtain a consolidated anomaly score.

Stacking: Individual models' outputs serve as features for a meta-model that learns to make a final decision.

Ensemble methods and anomaly detection ensembles offer several advantages, including:

Improved Detection Performance: Ensemble methods can leverage the collective wisdom of multiple models, capturing a broader range of anomaly patterns and reducing individual model limitations.

Robustness to Variability: Ensemble approaches can mitigate the impact of noise, data variations, and model biases, leading to more reliable and stable anomaly detection results.

Generalization: Anomaly detection ensembles can adapt well to different domains and data characteristics, enhancing their ability to detect anomalies in diverse and evolving systems.

In summary, the novel techniques and algorithms for anomaly detection discussed offer valuable advancements in the field. Time series decomposition and reconstruction methods provide a deeper understanding of the underlying patterns, while transfer learning and domain adaptation techniques leverage knowledge from related domains to improve anomaly detection performance. Online and streaming approaches enable real-time detection and adaptability to dynamic data streams, and ensemble methods harness the strengths of multiple models to enhance detection accuracy and robustness. These techniques contribute to more accurate, efficient, and adaptable anomaly detection systems across various domains and real-world scenarios.

IX. SIMULATED EVENTS AND APPLICATIONS

Case study 1

Anomaly detection in network traffic data

Anomaly detection in network traffic data is crucial for identifying network intrusions and malicious activities. By analyzing patterns and behaviors in network traffic, anomalies indicative of cyber attacks can be detected. The in-depth analysis involves evaluating the effectiveness of anomaly detection algorithms in detecting both known and unknown attack patterns, considering the challenges of high-dimensional and high-volume network data. The evaluative analysis includes assessing the false positive and false negative rates, the ability to handle evolving attack techniques, and the impact on network security and operational costs. Furthermore, the scalability and real-time capabilities of anomaly detection methods should be considered to ensure their applicability in large-scale network environments. By default, the anomaly detection herein can be addressed using the Hybrid Technique employing clustering algorithms (k-means and DBSCAN), statistical methods (Gaussian Mixture Models, Hidden Markov Models (HMM), and machine learning algorithms (Isolation Forest, One-class Support Vector Machines (SVM).

The approach highlighted below can be implemented for this case.

1. Preprocessing: Clean and preprocess network traffic data by removing noise, handling missing values, and normalizing the data.
2. Feature extraction: Extract relevant features from network traffic data, such as packet size, protocol type, source/destination IP addresses, etc.
3. Model selection: Choose an appropriate anomaly detection algorithm, such as clustering-based methods (e.g., k-means), statistical methods (e.g., Gaussian Mixture Models), or machine learning algorithms (e.g., Isolation Forest, One-class SVM).

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

4. Model training: Train the selected anomaly detection model using labeled or unlabeled data, considering normal and anomalous traffic patterns.
5. Anomaly detection: Apply the trained model to identify anomalies in real-time or batch processing by comparing new data instances to the learned patterns.
6. Post-processing: Evaluate and refine the detected anomalies, considering thresholds, domain-specific rules, or expert knowledge.

Case study 2
Anomaly detection in energy consumption patterns
Anomaly detection in energy consumption patterns is valuable for identifying abnormal energy usage, equipment malfunctions, or energy theft. By monitoring and analyzing energy consumption data, anomalies can be detected, leading to improved energy efficiency and cost savings. The in-depth analysis involves examining the accuracy and timeliness of anomaly detection methods in capturing energy-related anomalies, considering factors such as seasonal variations, weather conditions, and occupancy patterns. The evaluative analysis includes assessing the effectiveness of anomaly detection in different types of energy data (e.g., electricity, gas, water), considering the trade-off between false positives and false negatives, and measuring the impact on energy management and sustainability goals. Furthermore, the adaptability of anomaly detection methods to different energy consumption patterns and the scalability to manage large-scale data should be evaluated. In line with the above approach in the first case, the Hybrid Technique is also recommended for this anomaly detection. It has to encapsulate statistical methods (ARIMA, Gaussian distribution-based methods), time series decomposition techniques (STL, EMD), and machine learning models (LSTM, Autoencoders)

The approach highlighted below can be implemented for this case.

1. Data preprocessing: Clean and preprocess energy consumption data by handling missing values, normalizing the data, and considering temporal aspects (e.g., seasonality, time of day).
2. Feature extraction: Extract relevant features from energy consumption patterns, such as average consumption, deviation from the expected consumption, or pattern similarity.
3. Model selection: Choose an appropriate anomaly detection algorithm, such as statistical methods (e.g., ARIMA, Gaussian distribution-based approaches), time series decomposition techniques (e.g., STL, EMD), or machine learning models (e.g., LSTM, Autoencoders).
4. Model training: Train the selected anomaly detection model using labeled or unlabeled energy consumption data, considering normal and anomalous consumption patterns.
5. Anomaly detection: Apply the trained model to detect anomalies by comparing the observed energy consumption to the learned patterns, considering statistical thresholds or model-specific criteria.
6. Post-processing: Evaluate and refine the detected anomalies, considering energy consumption patterns, external factors (e.g., weather data), and domain-specific knowledge.

Case study 3
Anomaly detection in medical sensor data
Anomaly detection in medical sensor data is critical for the early detection of abnormal physiological conditions or adverse events in patients. By monitoring and analyzing data from medical sensors, anomalies can be identified, enabling timely interventions and improved patient outcomes. The in-depth analysis involves examining the sensitivity and specificity of anomaly detection methods in different physiological signals (e.g., heart rate, blood pressure, electrocardiogram), considering variations across patient populations and clinical settings. The evaluative analysis includes assessing the ability of anomaly detection to differentiate between true anomalies and expected variations, evaluating the impact on patient safety and healthcare quality, and addressing privacy concerns and data security in medical data usage. Furthermore, the interpretability and explainability of anomaly detection methods in the medical domain and their integration into clinical workflows should be evaluated.

The approaches for these anomaly detections include:

1. Data preprocessing: Clean and preprocess medical sensor data by handling missing values, noise reduction, and normalization.
2. Feature extraction: Extract relevant features from medical sensor data, such as heart rate variability, abnormal signal patterns, or statistical measures (mean, standard deviation, etc.).
3. Model selection: Choose an appropriate anomaly detection algorithm, such as statistical methods (e.g., z-score, Mahalanobis distance), clustering-based methods (e.g., k-means), or machine learning algorithms (e.g., SVM, Random Forest).

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

4. Model training: Train the selected anomaly detection model using labeled or unlabelled medical sensor data, considering normal physiological patterns and known abnormal conditions.
5. Anomaly detection: Apply the trained model to detect anomalies by comparing new sensor data to the learned patterns, considering statistical thresholds or model-specific criteria.
6. Post-processing: Evaluate and refine the detected anomalies, considering clinical context, expert knowledge, and external factors (e.g., medication
This case study as well follows the trend of the previous ones by deeming fit for a Hybrid technique. Three mixed approaches will aid the anomaly detection here in statistical methods (z-score, Mahalanobis distance), clustering algorithm (k-means), and machine learning algorithms (Support Vector Machines (SVM), Random Forest)

Deductively, a Hybrid Technique is efficacious in anomaly detection as it opens a window of limitless execution, processing, training, and prediction of anomaly sequences in real-time and simulated events. In all these case studies, it is important to critically analyze the limitations and potential biases of the anomaly detection methods used, considering factors such as data availability, data quality, and the assumptions made by the algorithms. The evaluation also includes a comparison of different anomaly detection techniques to determine their relative performance and suitability for specific application domains. Additionally, ethical considerations such as privacy, fairness, and transparency must be carefully addressed when applying anomaly detection in real-world scenarios.

X. EMPIRICAL REVIEW

This paper emphasized several algorithms with advantages, but the general efficacy findings demand more study in the following categories:

Adaptability: There is no one method (or algorithm combination) that completely outperforms all others and resolves all anomaly detection problems [47]. In this study, it is proposed that more research on holistic and hybrid anomaly detection systems combine current strengths for the identification of more varied abnormalities in time series with arbitrary properties to progress the area of anomaly detection.

Dependability: Regardless of unending utmost initiatives, only a small number of algorithms were capable of accurately processing every time series while staying within reasonable time and memory constraints. We also stress the significance of more studies on the stability and scalability of methods for detecting time series anomalies.

Clarity: Most of the study's anomaly detection algorithms required, on average, seven different parameter settings since they were so sensitive to them. The fact that most practical application scenarios lack training data for algorithm tuning makes the situation exacerbated. This lack highlights the critical need for more study(s) on auto-configuring and self-tuning algorithms.

Explainability: Explainable AI and interpretable anomaly detection models are crucial for building trust and understanding in anomaly detection systems. Future directions should focus on developing interpretable models that provide clear explanations for detected anomalies, enabling analysts to understand the underlying reasons and take appropriate actions. This includes research on feature importance analysis, rule-based models, and visualization techniques. Critical evaluation involves assessing the trade-off between model complexity and interpretability, measuring the impact of explainability on decision-making, and addressing challenges in interpreting complex deep learning models.

Incorporation: Future developments should stress the incorporation of domain knowledge and context awareness into the detection process to improve the accuracy and relevance of anomaly detection. Incorporating domain-specific rules, limitations, or expert knowledge to guide the detection process and reduce false positives is an example of this. Contextual information, such as environmental characteristics, temporal trends, or user behavior, can also increase anomaly detection performance. Examining the usefulness of adding domain knowledge, analyzing the scalability of context-aware models, and weighing the trade-off between generalizability and domain specificity are all part of critical evaluation.

Data security and privacy: As data privacy concerns continue to rise, future directions should focus on privacy-preserving techniques for anomaly detection. This includes research on privacy-enhancing algorithms, such as differential privacy, secure multiparty computation, or federated learning, to detect anomalies without compromising sensitive data. Critical evaluation involves assessing the impact of privacy-preserving techniques on detection accuracy, computational efficiency, and scalability. Additionally, considering the legal and ethical implications of data anonymization and protection is crucial in evaluating the feasibility and adoption of privacy-preserving anomaly detection methods.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

Integration: Future directions should explore the integration of anomaly detection with other data analysis techniques, such as predictive modeling, clustering, or outlier detection. This integration can provide a holistic view of the data and improve anomaly detection performance by leveraging complementary information.

Critical evaluation involves assessing the synergy between anomaly detection and other techniques, measuring the impact on detection accuracy and efficiency, and considering the scalability of integrated approaches. Furthermore, exploring ensemble methods that combine multiple detection techniques can enhance the robustness and reliability of anomaly detection systems.

An in-depth critical evaluation of future directions and emerging trends in anomaly detection requires careful consideration of their practical implications, limitations, and potential challenges. Factors such as computational complexity, scalability, real-time processing, and the balance between accuracy and interpretability should be critically analyzed. Additionally, addressing ethical concerns, data privacy, and the impact on decision-making processes are vital aspects of evaluating the feasibility and adoption of these future directions.

XI. CONCLUSION AND RECOMMENDATION

In this study, we have explored the field of anomaly detection in time series data and examined various aspects related to methodologies, evaluation metrics, benchmark datasets, novel techniques, and real-world case studies. Methodologically, statistical-based, machine learning-based, and deep learning-based methods have been discussed in anomaly detection in time series data, highlighting their strengths and limitations.

The study also explored evaluation metrics such as precision, Recall, F1-score, ROC curve, and AUC, emphasizing their significance in assessing the performance of anomaly detection algorithms. Additionally, we examined commonly used benchmark datasets that serve as standards for evaluating the effectiveness of anomaly detection techniques.

While considerable progress has been made in the field of anomaly detection in time series data, several challenges and future research directions remain to be uncovered.

1. Interpretability: Future research should focus on developing explainable AI techniques and interpretable anomaly detection models to enhance the trust, understanding, and adoption of these models in real-world applications.

2. Incorporating Context and Domain Knowledge: There is a need to further explore the integration of contextual information and domain-specific knowledge to improve anomaly detection accuracy and reduce false positives. This involves considering temporal, spatial, and other contextual factors that impact the occurrence of anomalies.

3. Privacy-Preserving Techniques: With increasing concerns about data privacy, future research should address the development of privacy-preserving anomaly detection techniques that can effectively detect anomalies while preserving the confidentiality and integrity of sensitive data.

4. Integration with Other Data Analysis Techniques: Exploring the integration of anomaly detection with other data analysis techniques, such as predictive modeling, clustering, or outlier detection, can lead to more comprehensive and robust anomaly detection systems.

5. Scalability and Real-Time Processing: As the volume and velocity of time series data continue to increase, there is a need for scalable and efficient anomaly detection algorithms that can handle large-scale data streams in real time.

6. Generalization to New Anomaly Types: Anomaly detection techniques should be able to generalize well to new and emerging types of anomalies. Future research should focus on developing algorithms that can adapt and detect anomalies that were not previously encountered during training.

Addressing these challenges and pursuing these future research directions will contribute to the advancement and broader adoption of anomaly detection techniques in time series data analysis.

In conclusion, this study has provided a comprehensive analysis of the methodologies, evaluation metrics, benchmark datasets, novel techniques, and real-world case studies in anomaly detection in time series data. The findings and future research directions discussed in this review pave the way for further advancements in the field and offer valuable insights for researchers and practitioners working in anomaly detection and related domains.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10507

REFERENCES

[1] Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., & Shaalan, K. (2019). Speech recognition using deep neural networks: A systematic review. *IEEE Access*, *7*, 19143-19165.

[2] Byrne, M. F., Chapados, N., Soudan, F., Oertel, C., Pérez, M. L., Kelly, R. & Rex, D. K. (2019). Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during the analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*, *68*(1), 94-100.

[3] Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, *9*, 611-629.

[4] Pruneski, J. A., Pareek, A., Kunze, K. N., Martin, R. K., Karlsson, J., Oeding, J. F., ... & Williams III, R. J. (2023). Supervised machine learning and associated algorithms: applications in orthopedic surgery. *Knee Surgery, Sports Traumatology, Arthroscopy*, *31*(4), 1196-1202.

[5] Sadique, K. M., Rahmani, R., & Johannesson, P. (2018). Towards security on internet of things: applications and challenges in technology. *Procedia Computer Science*, *141*, 199-206.

[6] Bhadoria, R. S., Samanta, S., Pathak, Y., Shukla, P. K., Zubi, A. A., & Kaur, M. (2022). Bunch graph-based dimensionality reduction using auto-encoder for character recognition. *Multimedia Tools and Applications*, *81*(22), 32093-32115.

[7] Zhang, Y., Meratnia, N., & Havinga, P. (2010). Outlier detection techniques for wireless sensor networks: A survey. *IEEE communications surveys & tutorials*, *12*(2), 159-170.

[8] Moustafa, N., Creech, G., & Slay, J. (2017). Big data analytics for intrusion detection system: Statistical decision-making using finite dirichlet mixture models. *Data Analytics and Decision Support for Cybersecurity: Trends, Methodologies and Applications*, 127-156.

[9] Leigh, C., Alsibai, O., Hyndman, R. J., Kandanaarachchi, S., King, O. C., McGree, J. M., ... & Peterson, E. E. (2019). A framework for automated anomaly detection in high frequency water-quality data from in situ sensors. *Science of the Total Environment*, *664*, 885-898.

[10] Silva, A. M. L., da Silva Sousa, F. A., de Freitas Santos, A. R., Machado, V. P., & Santana, A. M. (2023). Method for Inferring the Optimal Number of Clusters with Subsequent Automatic Data Labeling based on Standard Deviation. *International Journal of Advanced Computer Science and Applications*, *14*(3).

[11] An, J., & Cho, S. (2015). Variational autoencoder based anomaly detection using reconstruction probability. *Special lecture on IE*, *2*(1), 1-18.

[12] Lanzante, J. R. (1996). Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *16*(11), 1197-1226.

[13] Collura, T. F., Thatcher, R. W., Smith, M. L., Lambos, W. A., & Stark, C. A. (2009). EEG biofeedback training using live Z-scores and a normative database. *Introduction to quantitative EEG and neurofeedback*, 103-141.

[14] Das, K., & Behera, R. N. (2017). A survey on machine learning: concept, algorithms, and applications. *International Journal of Innovative Research in Computer and Communication Engineering*, *5*(2), 1301-1309.

[15] Hasan, M., Islam, M. M., Zarif, M. I. I., & Hashem, M. M. A. (2019). Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. *Internet of Things*, *7*, 100059.

[16] Pu, G., Wang, L., Shen, J., & Dong, F. (2020). A hybrid unsupervised clustering-based anomaly detection method. *Tsinghua Science and Technology*, *26*(2), 146-153.

[17] Ahmed, M., Mahmood, A. N., & Islam, M. R. (2016). A survey of anomaly detection techniques in the financial domain. *Future Generation Computer Systems*, *55*, 278-288.

[18] Oruh, J., Viriri, S., & Adegun, A. (2022). Long short-term Memory Recurrent neural network for Automatic speech recognition. *IEEE Access*, *10*, 30069-30079.

[19] Fanta, H., Shao, Z., & Ma, L. (2020). SiTGRU: single-tunnelled gated recurrent unit for abnormality detection. *Information Sciences*, *524*, 15-32.

[20] Nguyen, H. D., Tran, K. P., Thomassey, S., & Hamad, M. (2021). Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *International Journal of Information Management*, *57*, 102282.

[21] Kaur, G. (2020). A comparison of two hybrid ensemble techniques for network anomaly detection in spark distributed environment. *Journal of Information Security and Applications*, *55*, 102601.

[22] Rayana, S., & Akoglu, L. (2016). Less is more: Building selective anomaly ensembles. *Acm transactions on knowledge discovery from data (tkdd)*, *10*(4), 1-33.

[23] Cook, A. A., Mısırlı, G., & Fan, Z. (2019). Anomaly detection for IoT time-series data: A survey. *IEEE Internet of Things Journal*, *7*(7), 6481-6494.

[24] Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer networks*, *51*(12), 3448-3470.

DOI: 10.17148/IARJSET.2023.10507

[25] Grill, M., Pevný, T., & Rehak, M. (2017). Reducing false positives of network anomaly detection by local adaptive multivariate smoothing. *Journal of Computer and System Sciences*, *83*(1), 43-57.

[26] Kim, D., Lee, S., & Lee, J. (2020). An ensemble-based approach to anomaly detection in marine engine sensor streams for efficient condition monitoring and analysis. *Sensors*, *20*(24), 7285.

[27] Fiore, U., Palmieri, F., Castiglione, A., & De Santis, A. (2013). Network anomaly detection with the restricted Boltzmann machine. *Neurocomputing*, *122*, 13-23.

[28] Choi, K., Yi, J., Park, C., & Yoon, S. (2021). Deep learning for anomaly detection in time-series data: review, analysis, and guidelines. *IEEE Access*, *9*, 120043-120065.

[29] Ordónez, F. J., de Toledo, P., & Sanchis, A. (2015). Sensor-based Bayesian detection of anomalous living patterns in a home setting. *Personal and Ubiquitous Computing*, *19*, 259-270.

[30] Goldenberg, N., & Wool, A. (2013). Accurate modeling of Modbus/TCP for intrusion detection in SCADA systems. *international journal of critical infrastructure protection*, *6*(2), 63-75.

[31] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, *21*, 1-13.

[32] Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, *21*(4), 498-507.

[33] Lessmann, S., Baesens, B., Mues, C., & Pietsch, S. (2008). Benchmarking classification models for software defect prediction: A proposed framework and novel findings. *IEEE transactions on software engineering*, *34*(4), 485-496.

[34] Paparrizos, J., Boniol, P., Palpanas, T., Tsay, R. S., Elmore, A., & Franklin, M. J. (2022). Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. *Proceedings of the VLDB Endowment*, *15*(11), 2774-2787.

[35] Pranav, M., & Zhenggang, L. (2020). A day on campus-an anomaly detection dataset for events in a single camera. In *Proceedings of the Asian Conference on Computer Vision*.

[36] Pranav, M., & Zhenggang, L. (2020). A day on campus-an anomaly detection dataset for events in a single camera. In *Proceedings of the Asian Conference on Computer Vision*.

[37] Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote sensing of Environment*, *114*(1), 106-115.

[38] Vaheddoost, B., & Aksoy, H. (2019). Reconstruction of hydrometeorological data in Lake Urmia basin by frequency domain analysis using additive decomposition. *Water Resources Management*, *33*, 3899-3911.

[39] Wang, Q., Michau, G., & Fink, O. (2019). Domain adaptive transfer learning for fault diagnosis. In *2019 Prognostics and System Health Management Conference (PHM-Paris)* (pp. 279-285). IEEE.

[40] Ahmad, S., Lavin, A., Purdy, S., & Agha, Z. (2017). Unsupervised real-time anomaly detection for streaming data. *Neurocomputing*, *262*, 134-147.

[41] Dromard, J., Roudiere, G., & Owezarski, P. (2016). Online and scalable unsupervised network anomaly detection method. *IEEE Transactions on Network and Service Management*, *14*(1), 34-47.

[42] Gama, J. (2012). A survey on learning from data streams: current and future trends. *Progress in Artificial Intelligence*, *1*, 45-55.

[43] Cui, R., Liu, H., & Zhang, C. (2019). A deep neural framework for continuous sign language recognition by iterative training. *IEEE Transactions on Multimedia*, *21*(7), 1880-1891.

[44] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, *46*(4), 1-37.

[45] Zhao, Y., & Hryniewicki, M. K. (2018, July). Xgbod: improving supervised outlier detection with unsupervised representation learning. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

[46] Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data*, *7*(1), 1-40.

[47] Temitope, O., Owoyemi, J., & Edeamah, O. (n.d.). *Exploring Techniques and Applications for Anomaly Detection in Time Series Data*.