I.J. Wireless and Microwave Technologies, **2025, 3, 54-65**
Published Online on June 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijwmt.2025.03.04

Detection of Anomalies Based on User Behavioral Information: A Survey

L. Lanuwabang*
Kalasalingam Academy of Research and Education, and Director, National Institue of Electronics & Information Technology, Kohima 797001, Nagaland, India
Email: lanu@nielit.gov.in
ORCID iD: https://orcid.org/0009-0004-3579-6763
*Corresponding Author

P. Sarasu
Department of CSE, Kalasalingam Academy of Research and Education, Krishnankoil 626126, Tamil Nadu, India
Email: Sarasujivat@gmail.com
ORCID iD: https://orcid.org/0000-0002-0494-6309

Received: 19 March, 2024; Revised: 19 June, 2024; Accepted: 04 March, 2025; Published: 08 June, 2025

Abstract: User and entity behaviour analytics (UEBA) solutions are becoming more and more popular for detecting anomalies since they establish baseline models of typical user behaviour and highlight deviations from them. Modelling normal user behavior and identifying any new behavior that deviates from the normal model user i.e., an attack, which is the main concept of Anomaly Detection (AD) techniques. In this work, a comprehensive review of various AD techniques based on user behavior is presented. Accordingly, this survey is concerted on various techniques employed for AD based on user behavior. Among various research articles, 50 research articles based on AD are considered and categorized based on different parameters, like techniques, publication year, performance metrics, utilized tools, and so on. At last, the research gaps and challenges of this method are illustrated in such a way that a goal for emerging an efficient technique for allowing the effective AD technique is defined.

Index Terms: Anomaly Detection, User Behavior, Deep Learning, Machine Learning, Anomalies

1. Introduction

In security analysis, detecting anomalies from log data is used for analyzing user behavior, which is a significant technique. Anomaly detection relies heavily on sophisticated artificial intelligence and machine learning algorithms. To examine user behaviour trends and spot outliers, models including supervised learning, unsupervised learning, and deep learning are used. UEBA is a cybersecurity solution that employs algorithms and machine learning to find irregularities in the behaviour of a corporate network's endpoints, routers, servers, and users. When there are deviations from routine daily patterns of usage, UEBA strives to identify any odd or suspicious behaviour. For instance, the UEBA system would recognise this as an anomaly and either alert an IT administrator or, if automated processes are in place, automatically disconnect that user from the network if they were regularly downloading files of 20 MB per day but suddenly started downloading files of 4 GB. A hypothetical distributed denial-of-service (DDoS) assault may begin one day when a server in one branch office unexpectedly receives thousands more requests than typical. This kind of activity can go unnoticed by IT managers, but UEBA would detect it and take additional action [1]. User Behavior Analytics (UBA) is one approach for identifying anomalies increase in insider abnormal behavior. The primary and significant detection step in the workflow of cyber security is to examine the user behavior changes based on early signs for additional investigation, and gradually identify the known and unknown attacks [2]. The user behavior analysis considers various factors such as contextual information, duration of sessions, peer group actions, and continuous behaviors [3, 4]. Anomalies are unpredictable and irregular, which generates complexity for the collection of data in real-time scenarios. It is challenging to exactly the discrimination of normal and abnormal events and the description of anomaly depends on global context. Here, the normal or abnormal event is recorded based on various situations, for example, one human tracking the other human’s bank account is considered as an abnormal event, whereas a normal event is tracking on the playground [5, 1]. Considering the nature of the anomalies, it is very complex, if not unfeasible, to assemble a globally comprehensive model, is the first paucity. This is because different subjects may develop novel

and creative methods for creating anomalous data for spoofing attempts, which are not earlier captured by the negative samples of the database utilized in the training stage [6, 7].

In modern days, DL-based generative models attained state-of-the-art presentation in the detection of abnormal events. Due to its advantages over the other Recurrent Neural Networks (RNNs), particularly its solution to the gradient disappearance issue, the Long-Short Term Memory (LSTM) network is the first choice [8, 2]. The k-nearest Neighbor (KNN) algorithm is another popular distance-based method that uses an average distance between its k-nearest neighbors to calculate anomaly scores and another method used is Clustering Based Local Outlier Factor (CBLOF) technique, where anomaly score functions are used for the identification of anomalies depend on clustering. Even though these distance-based methods work in some cases, if the known anomaly duration and how many anomalies are there [9, 10]. Moreover, unsupervised Machine Learning (ML) methods are presented for temporal information for recognizing signs of anomaly behaviors that represent insider attacks [11].

The main objective of the survey is to detect anomalies based on user behavior and it provides a comprehensive summary of the existing methods In the AD domain. This survey considered the analysis based on publication year, various algorithms, performance measures, various datasets, and tools for implementation. Moreover, the performance evaluation measures are considered for determining the function of devised AD-based techniques. The survey is done for the development of research problems, and it provides a direction for a more effective and efficient AD system.

This survey is formulated as follows, wherein part 2 details on the survey of different object recognition approaches, part 3 illustrates the obstacles of existing works, part 4 describes the analysis by considering publication year, used methodology, evaluation measures, datasets used and execution tool and the conclusion of the survey is made in part 5.

2. Literature survey of various AD techniques

The review of several AD techniques is demonstrated in this part. Fig.1. illustrates a type of various AD algorithms. These algorithms are widely classified into five techniques, namely Hidden Markov Model (HMM), ML-based algorithms, DL-based algorithms, Unsupervised learning, and Rule-based analysis. The subtype of ML-based techniques has Principle Component analysis (PCA), Support Vector Machine (SVM), and Isolation forest. A subtype of the DL-based technique is CNN. A concise explanation of the subsequent algorithms are given below:

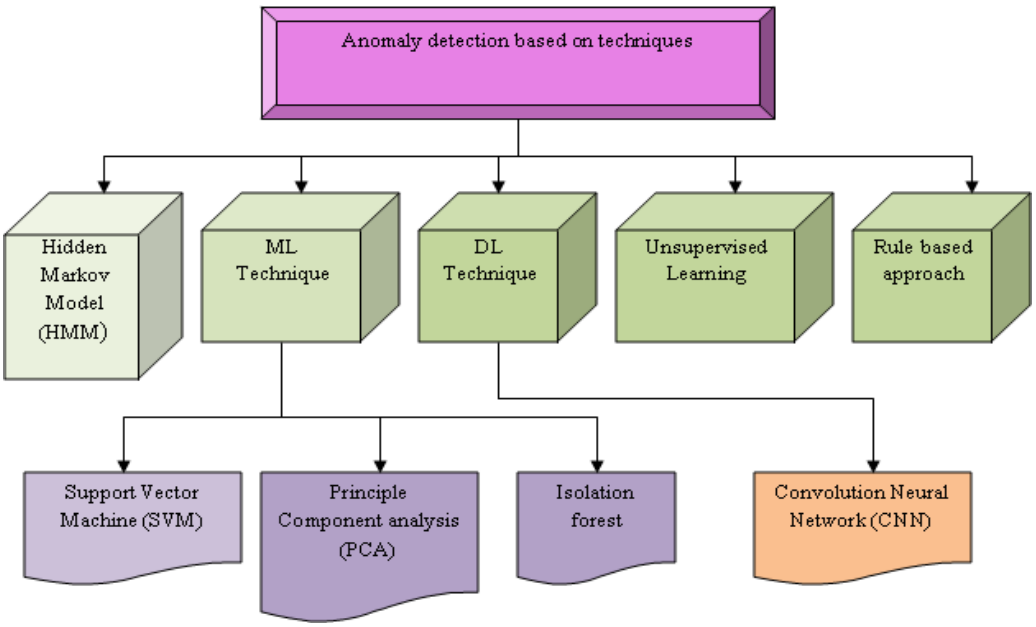

Fig. 1. Anomaly detection-based techniques

2.1 Categorization of AD-based technique

This review selects AD techniques such as HMM, ML-based techniques, DL-based techniques, Unsupervised learning, and Rule-based analysis. The subtype of ML-based techniques has SVM, PCA, and Isolation forest. The subtype of DL based technique is CNN.

2.1.1 HMM technique

Wang, T., *et al.* [12] devised an HMM-based on the Histogram of Optical Flow Orientations (HOFO) for the prediction of abnormalities. This technique was performed in video and it was calculated based on the selected video frame and conventional video frame. Here, anomalies were detected effectively from the normal and abnormal events. However, this technique was not used for multiclass problems. Forkan, A.R.M., *et al.* [13] introduced an HMM technique for daily activity AD, and the result of the technique was fused using a fuzzy rule-based technique for creating the final guess and distribution of accurate context-aware. This technique performed better on complex situations, but this technique failed to process the large-scale dataset. Yamauchi, M., *et al.* [14] introduced an HMM for abnormal event detection, which was used by the home IoT device and it generated multiple sequences of the event through the elimination of some sequences. This technique achieved high detection accuracy and was easy to implement, but it did not include single-operation detection to improve legitimacy.

2.1.2 ML Technique

Kamranfar, P., *et al.*[15] introduced Multiple Instance Learning (MIL) for anomaly detection, which was a key component of various designs and formulation of algorithmic instantiations of this work. This technique formed single-instance learning, which makes it easy to moderate the challenging datasets. However, this technique failed to consider the various strangeness score formulations for abnormality measurement. Mart ń, A.G., *et al.*[16] devised the Symbolic Aggregate approximation (SAX) for AD using temporal information. Here, a single sequence of temporal information was processed and Random Trees embedding was used for each information source. This technique was performed accurately for loss of information but failed to test more information at a time. Parwez, M.S., *et al.*[17] introduced k-means clustering and hierarchical clustering for verifying the correctness of detected anomalies. The major process of ROI was fault avoidance solution and resource allocation. For the training of the data, a neural network and anomaly-free data were identified. The anomalies were detected effectively in traffic networks, but this technique failed to learn the dynamics of users in predicting smart cities.

Hill, D.J., *et al.* [18] developed an AD named AECID to apply ML approaches for performing correlation analysis, and sequence analysis in log data. This execution time of the technique was minimal, but it required more parameters for security purposes. Arashloo, S.R., *et al.* [7] developed the SVM for unseen attackers identified in anomaly detection and the main function of this technique was evaluation situations, statement formation, and solving statement are presented. Here, one class of SVM was performed. This algorithm easily recovered the real-time environmental problems, although this technique had high computational complexity. Feremans, L., Liu, Z., *et al.*[19] devised a K-prototype clustering and k-NN classification from AD based on log information. Here, K-prototype clustering functioned based on the partitioning of data into a number of clusters and the KNN classifier detected the accurate detection results. The technique had more detection accuracy and minimum computational complexity, although this technique failed to collect more anomaly logs for performance improvement.

Hill, D.J. and Minsker, B.S., [20] introduced a data quality assurance and control (QA/QC) for real-time anomaly detection. This model performs faster, scales to a huge amount of data, incremental estimation of data as it becomes accessible, and needs no pre-classification of anomalies. This technique did not use data from numerous sensors for a single-sensor information stream in anomaly detection. Sulayman, I.I.A. and Ouda, A., [21] developed a k-means, HMM, and auto-encoder neural network for the prediction of anomaly through big data utilization. This technique was predicted based on the user's behavior and it had four main steps, rare user functions detection, filtered security action, operation of a user profile, and real-time issue question. This technique performed better in user authentication, although This technique did not combine a number of techniques for increasing data dimension.

Böse, B., *et al.* [22] introduced a Real-time AD In Streaming Heterogeneity (RADDSH) for responding of anomaly behavior. The main application of this technique was detecting the problem of insider threats. This technique failed to multi-tiered detection system subj analysts to operate at the greatest level in the data chain. Cui, J., *et al.* [23] devised a multi-homed abnormal behavior detection model based on fuzzy particle swarm clustering (MAD-FPC), which was performed based on the UEBA platform. The particle swarm optimization algorithm used by the model highly avoided the clustering problem for a local optimal problem. This technique failed to establish a network security protection model to identify unknown abnormal traffic in anomaly detection.

Lukashin, A., *et al.* [2] developed a UBEA architecture for semi-structured data-based anomaly detection. This algorithm had various SIEM sources introduced for constructing features from hybrid data streams. It had a more scalable and effective architecture, but it failed to analyze the various ML methods including exploring seq2seq methods to detect the attacks, and the possibility of utilizing LSTM for AD. Xi, X., *et al.* [4] developed a Compining OCSVM, Replicator neural network (RNN), and Isolation Forest for the prediction of anomalies. A UBA Platform was introduced for the collection of logs, feature extraction, and processing of the experiment. This technique performed better for more complicated and real anomaly detection, but it did not process the sequence of data in UEBA. Hu, J., *et al.* [1] developed a Single-hidden-Layer Feedforward Neural Network for video anomaly detection. Here, optimization was used for solving the regression problem, one class SVM was done for classification and an ensemble technique was used for finding the multiple anomalies detection scores. This algorithm had low computation cost, less running time, more efficiency, and robustness, although it failed to implement the more ensemble technique for improving performance.

i) PCA based Technique

Fernandes Jr, G., *et al.* [24] developed a Digital Signature of Network Segment using Flow analysis (DSNSF) for abnormal behavior detection. The intention of the technique was contextualization and application of Dynamic Time Warping, Ant Colony Optimization, and PCA for AD and pattern detection. This technique performed better in a real-time environment, but it failed to analyze the correlation between the data attributes for improvement of anomaly detection. Hu, J., *et al.* [25] developed a Low-rank based Unsupervised One-class Learning with Ridge Regression (LRUOCL-RR) for video anomaly detection. The one class SVM was used to process the normal functions and increase performance. The computation cost of the technique was minimal but failed to use DL for video AD. Mart ń, A.G., *et al.* [26] developed a one-class SVM for AD based on a fingerprint in user behavior characteristics. Here, the Utilization of the UEBA platform and integration of the algorithm for AD was presented. This technique was scalable, but it did not measure the security level of the technique.

ii) SVM-based technique

Cheng, K.W., *et al.* [27] devised a maximum subsequence search for the identification of anomaly sequences in a video sequence. The proposed search calculation incorporates nearby scores of local anomaly into a worldwide reliable prediction so the beginning and end of a strange occasion are not entirely settled under misleading and missing recognitions. This algorithm achieved effectively that anomaly was predicted in abnormal events, but it failed to detect the moving object-based anomaly detection.

iii) Isolation forest-based technique

Feremans, L., et al. [28] introduced a pattern-based embedding for AD in mixed-type time series. An embedding of a mixed-type time series was created using the approach's regular pattern mining techniques to train an isolation forest. The real-time dataset was effectively processed, although this technique had failed to detect a classifier for AD. Tahir, M., *et al.* [29] developed a local outlier factor (LOF) and One-class SVM, which was profile based in identifying the anomalies. Here, data described user behaviors of various perspectives by various flags. Every profile was composed of information processed over a period of time. Moreover, this technique obtained results that was more scalable and easy to process. This technique did not process such devices as, mobile users, IP cameras, printers, thermostats, multiple base stations, and mobile devices, etc.

2.1.3 DL-based technique

Lee, S., *et al.* [30] developed the Bidirectional Multi-Scale Aggregation Networks (BMAN) for abnormal event detection, which contained two parts: an appearance-motion joint detector and an inter-frame detector. This technique performed better in more challenging tasks, but the configuration of a network was more complex. Chen, W., *et al.* [31] developed a combination of Recurrent Neural Networks and Autoencoders for distributing the rumors anomaly. Here, some feature extraction was used to improve the performance of the detection rate. Moreover, this technique performed effectively for various types of AD, but it was not suitable for more rumors. Li, T., *et al.* [32] introduced the Two-Stream Deep Spatial-Temporal AutoEncoder (Two-Stream DSTAE) for AD. The appearance characteristics are taken from the spatial stream, while the motion patterns are taken from the temporal stream. This technique attained high prediction accuracy for complex data, but it failed to process real-time surveillance video AD.

Li, D., *et al.* [9] introduced a Multivariate AD with GAN (MAD-GAN) for AD, and time series data were processed here. Moreover, this technique easily accessed complex cyber attacks, but it failed to include feature selection for multivariate anomaly detection. Lin, S., *et al.* [33] devised a VAE-LSTM for the prediction of anomaly, which was a hybrid technique based on the unsupervised learning technique. This technique utilized the VAE phase for structuring the local features over a short window and the LSTM phase for calculating long-term correlations in sequence. This algorithm capable of predicting all varieties of anomaly based on multiple time series, although it was failed to use multiple scale windows for accurate AD. Yan, S., *et al.* [34] introduced a two-stream recurrent VAE for moving stream and every stream was performed based on the probabilistic distribution of the normal data in semi-supervised learning approaches. The effectiveness of the proposed technique was performed by a large-scale dataset, although it did not detect a greater number of anomalies.

Henriques, J., *et al.* [35] developed an XGBoost for detecting anomalies in a large number of unlabeled datasets. Here, detection was done by clustering and classification based on the computing environment, clustering carried out by gradient tree boosting algorithm, and classification carried out by XGBoost. This technique regulates complexity, so improves the performance, but it failed to implement more ensemble technologies to improve the performance. Garg, A., *et al.* [36] devised an unsupervised and semi-supervised DL technique for AD in time series data, and a dynamic scoring function was presented for multivariate time series data. F score measurement was utilized to find the detection rate of technique. This method had better generalizability, but it required processing a more challenging dataset for multiple ADs. Zhang, H. and Li, Z., [37] developed a Bayesian-based DL technique for accurate anomaly detection. Initially, input was applied to the anomalous process detection and the quality and amount of data were considered here. Finally, the anomaly was detected and it achieved a better false positive rate and accurate detection rate. Although, this technique failed to analyse the tradeoff between the delivery ratio of data and security for AD.

i) CNN based technique

Ullah, W., *et al.* [38] introduced a Convolutional Neural Network (CNN) for the detection of anomalies from the video frame. A further step of the technique was feature extraction, a Bi-directional long short-term memory (BD-LSTM) method for classification purposes. This technique was processed effectively in complex surveillance scenarios. However, it failed to defeat the issue of subordinate variation for anomaly recognition. Li, N. and Chang, F., [24] developed a Multivariate Gaussian Fully Convolution Adversarial Autoencoder (MGFC-AAE) for the effective detection of anomaly and localization techniques. To achieve more comprehensive detection results, the gradient and optical flow patches were used as inputs for the two-stream framework, which combined the movement and exterior cues to produce the two streams. This technique resolved the error of camera location, but it failed to incorporate LSTM with the adversarial autoencoder (AAE) for conceptualizing the sequential data of video sequences. Li, N. and Chang, F., [39] proposed a anomaly detection technique named Multivariate Gaussian Fully Convolution Adversarial Autoencoder (MGFC-AAE). This model was experimented using datasets like UCSD, UMN, and Avenue. This model solved the camera's position and video scene underlying perspective issues using the multi-scale patch structure. However, this model failed to capture temporal information of the video data.

2.1.4 Unsupervised learning

Le, D.C. and Zincir-Heywood, N., [11] devised an unsupervised ML technique for insider threat detection and it was based on the representation of data with temporal function. This technique more generalized capability for detecting new anomalies in different datasets, but it low convergence speed. Feremans, L., *et al.* [40] introduced a Time series representation for anomaly prediction, which was a novel and unsupervised-based technique. This technique was easy to perform and an open-source tool for implementation and detects the abnormal function in multi-dimensional time series. This technique was required to ensemble technique for more anomaly detection.

2.1.5 Rule-based approach

Jiang, F., *et al.* [41] devised the hierarchical data mining technique for moving object AD and the varied levels of a moving object was tracked based on the spatiotemporal contexts. At every level, normal event rules were discovered based on the frequency-based analysis. This technique was computationally effective and concluded the complex rules, but it failed to upgrade the present models as novel video annotations sequence in, so this method can effectively adapt to visual contextual modifications across a long epoch of time. Fu, Q., *et al.* [42] developed a Finite State Automaton (FSA) for the execution of anomaly, which was an unstructured log analysis technique. The main step of this algorithm was log files were taken from text information, which was fed to the log key and FSA was used for training of log sequence. The training time of the technique was minimal but it failed to add log parameter data for enhancing AD outcome.

Friedberg, I., *et al.* [43] developed signature-based detection methods for AD, which were based on intrusion detection systems. For detecting anomalies keep track of classification events, their dependencies, and their occurrences. As a result, it studies to behave normally over time and information on any actions that are different from the architecture model. This technique achieved high reliable accuracy, but this technique failed to process a lack of data about the resemblance of event classes resulting in unneeded hypotheses that might exceed the system model. Kim, H.K., *et al.* [44] developed a Recency, Frequency, Monetary (RFM) technique for misuse events, which was analyzed based on the incorporation of these three. This technique had less false positive rate, but it had low convergence capability.

2.1.6 Other techniques

Lijun, Z., *et al.* [45] introduced the Intuitionistic Tumbling Windows event calculus (ITWec) for difficult abnormalities detection. The probability of a low-level process connected with a high-level difficult abnormal process was used in the devised technique to begin nonparametrically measuring the probabilities of its survival for each period. The hybridization of the technique had a low execution time, but this technique required to implementation of the TL for an abnormal event. Radoglou-Grammatikis, P., *et al.* [46] developed a Secure and PrivatEsmArtgRid (SPEAR), which was based on Smart Grid (SG). The main goal of the technique was the accomplishment and design of a SIEM method able to predict, correlate cyberattacks, and normalize SG application layer roles. This technique performed good convergence capability, but this technique required the incorporation of more intrusion and anomaly prediction methods in the Big Data Analytics Component (BDAC).

Saebi, M., *et al.* [47] devised a First Order Network (FON) for accurate analysis network and this BuildHON+, for a higher-order network illustration of data resulting from a composite method with different orders of dependencies. This technique attained better results in noisy environments, although it failed to classify various types of anomalies in various nodes. Chong, Y.S. and Tay, Y.H., [48] developed a Spatial convolution for AD, which had two components, that was spatial feature demonstration and another for learning the temporal development of the spatial features. This technique had fast convergence capability, but this technique failed to add active learning – having human feedback for improving prediction.

Colque, R.V.H.M., *et al.* [49] devised a histogram of optical flow orientation and magnitude and entropy for detecting the normality and abnormality of the event. The outcome of the algorithm revealed that handled the various situations and it successfully handled the various situations of anomalies. Although this technique failed to implement the optimization function for enhancing the prediction. Chen, L.J., *et al.* [50] developed an Anomaly Detection Framework (ADF) for real-world environmental sensing systems. This technique contained four phases as Time-Sliced AD, Device Ranking (DR), Real-time Emission Detection (RED), and Malfunction Detection. This algorithm was more effective, because usage of a large-scale dataset, execution time of the technique was high.

Salles, R., *et al.* [51] developed a Hierarchical Temporal Memory for abnormal event detection and it evaluated the detection accuracy and identified the degree of its detection representation events. This technique effectively performed the all-time series for detection, but it failed to perform the real-time application. Cao, N., *et al.* [52] developed a TargetVue for social AD and TargetVue includes three brand-new egocentric glyphs that effectively present features, user's communication activities, and social interactions by visually summarizing their behaviors. The tool of the technique was performed better for the problem of AD, but it did not include the active learning technique for AD.

Xi, X., *et al.* [53] developed a combination of OCSVM, RNN, and Isolation Forest for AD and performed an overview of the UBA platform construct to gather logs, and features extraction and predict anomalous users, which contained potential threats. As a result, to boost performance and robustness, an ensemble of three unsupervised AD algorithms is utilized. However, this technique failed to use a sequence of data in the UBA platform.

Laue, T., *et al.* [54] devised a GLACIER, which was realized in-house functioned Security Information Event Management (SIEM) system for the prediction of an abnormal event. The most particular element of this framework that divides it from comparable arrangements in a market is its user feedback ability. Graphical User Interface (GUI) shows detected anomalies to security personnel, who can provide feedback on them. This technique performed better in security-related anomalies. The comparison of existing methods is provided in Table 1.

Table 1. Comparison table for Existing Methods

Author Name	**Methodology**	**Limitation of the paper**	**Experiment setup**	**Performance parameter**
Lee, S., *et al.* [30]	Bidirectional Multi-Scale Aggregation Networks for Abnormal Event Detection	Less accuracy	Conducted experiments with the datasets such as UCSD, UMN, Avenue Dataset, and Shanghai Tech Dataset.	Abnormality score, AUC.
Chen, W., *et al.* [31]	Unsupervised learning model using RNN and autoencoders.	Less accuracy	Using Sina Weibo, the experiment was conducted.	Accuracy, F1, Precision, Recall, FPR
Mart ń, A.G., *et al.*[16]	Combining user behavioural information at the feature level to enhance continuous authentication systems	High computational cost	UEBA, and TWOS dataset	FAR, FRR, EER, Accuracy, Specificity, Negative Predictive Value (NPV), and F1− score
Skopik, F., *et al.*[18]	Behaviour-based Anomaly Detection in Log Data of Physical Access Control Systems	Cannot be suitable for real-time dataset	The dataset is collected from the real-world installations of Australian company PKE.	-
Yamauchi, M., *et al.* [14]	Anomaly Detection in Smart Home Operation From User Behaviors and Home Conditions such as time and temperature	Privacy is not considered.	Implementation is done collecting data from the home IoT devices.	Misdetection ratio, Detection ratio
Cao, N., *et al.* [52]	a novel visual analysis system is introduced, i.e., TargetVue, which detects anomalous users through an unsupervised learning model and visualizes the behaviors of suspicious users in a behavior-rich context via novel visualization designs and multiple coordinated contextual views.	Failed to compare the results with other techniques.	The experiment is done by using the raw data collected from the Twitter activities.	-
Tahir, M., *et al.* [29]	Profile-based method for identifying anomalous changes in network user behaviors.	High implementation cost	Evaluated through Machine learning dataset	FPR, TPR, Precision, F1-score
Laue, T., *et al.* [54]	Security Information Event Management (SIEM) system for advanced anamoly Detection	Failed to detect complex attacks	Conducted field tests to verify the developments, and to prove the effectiveness.	-
Sulayman, I.I.A. and Ouda, A., [21]	User Modeling via Anomaly Detection Techniques for User Authentication	Failed to consider the large size of the data dimension.	Evaluation is done by three models such as K-means, HMM (Hidden Markov model), Autoencoder neural network.	TPR, TNR
Muliukha, V., *et al.* [55]	Anomaly Detection Approach in Cyber Security for User and Entity Behavior Analytics System Using Unsupervised Model	Failed to analyze the performance of the model.	Conducted machine learning methods such as Isolation forest and Local outlier factor methods to detect the anomalies.	-
Cui, J., *et al.* [23]	Multi-homed abnormal behavior detection algorithm based on fuzzy particle swarm cluster in user and entity behavior analytics	High computational costs	Used the NSL-KDD dataset to verify the effectiveness of the anomaly detection	Accuracy, F1, Precision, Recall.
Zhang, H. and Li, Z., [37]	Anomaly Detection Approach for Urban Sensing Based on Credibility and Time-Series Analysis Optimization Model	High computational complexity	Real-time simulation	FPR, Detection rate

3. Research Gaps and Issues

The issues faced by the literature are detailed in this section. This gap provides a way for future research in developing an efficient method by analyzing the issues that are not addressed and solved by previous methods. The research gaps and issues focused on during the multiple AD technique's analysis are illustrated below. In [12], HMM-based on HOFO was developed for the prediction of abnormalities from the movement of the frame. Here, complex background video anomalies was detected effectively, but this technique failed to be robust for multiclass problems. K-means clustering and hierarchical clustering was introduced in [17] for AD and this algorithm worked better in high traffic networks for the detection of anomalies. However, it was not suitable for the Big data analytics technique, which utilized users' contextual information such as traffic patterns, mobility patterns, social networks, ties, etc. In [18], developed an AD named, AECID approaches for performing correlation analysis, and sequence analysis in log data. The major merits of AD-based detection algorithm; it performed best in susceptible to a convinced type of error, but it required continuous learning mode, which was useful in unstable environments, where no stable behavior representation emerged, it was also difficult to organize a fundamental ageing technique suitably.

In [27], one-class SVM was developed for AD and localization of anomalies and the benefit of this algorithm was incorporating local anomaly scores into a global dependable detection, false/missing detections are likely to be removed, but it failed detection of moving object anomalies. The DL-based technique was introduced in [31] for AD. Rumors and various kinds of misinformation on online social Networks have been identified accurately, but it failed to detect more rumors from a wide range of areas by learning models and it required more training data. MAD-GAN was developed in [9] for AD in time series data. This technique was processed better in complex cyber-attacks and achieved superior performance, but it failed to establish the possible subsequence resolution for time-series analysis by GAN.

CNN was developed for the detection of anomalies from abnormal events. This technique performed better in complex datasets, but it failed to investigate the feature's motion and well-built visual features and it had lower variation features, so it was not efficient for detection [38]. In [11], an unsupervised technique for AD. As a percentage of the data, the temporal representation of the data performed significantly better than the unique extracted data, allowing for effective insider attack detection with very limited investigation budgets and better generalization to new data, but it failed to detect actions of informed attackers and adversarial attacks. Signature-based detection technique was developed in [43], and this technique used real-world datasets provided for AD, although this technique failed to process due to a lack of information from the abnormal event.

4. Analysis and Discussion

The examination for AD based on user behavior using utilized Datasets, based on publication year, devised methodologies, techniques employed, software tools, performance metrics, and performance measures values are illustrated.

4.1 Publication year based assessment

This subsection demonstrates an examination used years in which 50 published research articles for anomaly detection. The publication year-based assessment is portrayed in Fig. 2. Out of the 50 papers are reviewed, more research papers for user behaviour-based AD are available in 2020.

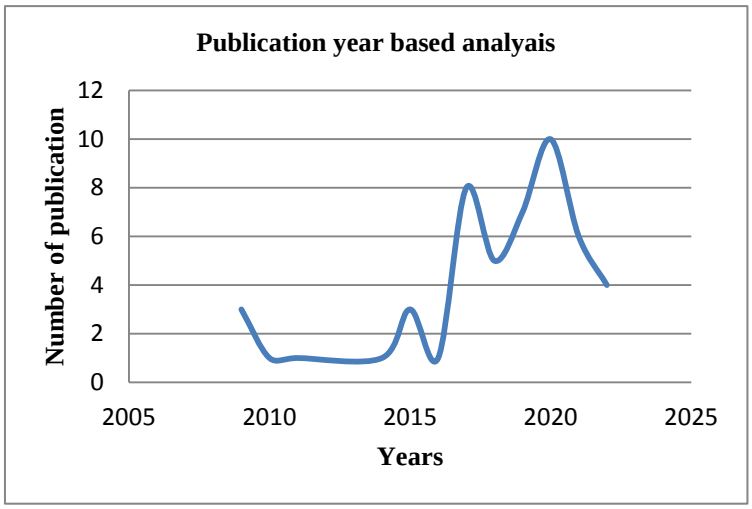

Fig. 2. Publication year-based analysis

4.2 AD Technique based assessment

This section discovers the examination of different approaches employed for AD-based user behavior. Fig. 3 portrays the examination of techniques employed for anomaly prediction. From Figure 3, 48% of research papers reviewed the ML-based technique, 29% of research papers presented DL-related techniques, HMM technique employed the 9% of research articles. 9% of research papers utilized rule-based techniques, and 5% of research articles presented the unsupervised-based algorithm. As a result, ML-related techniques are broadly employed for AD.

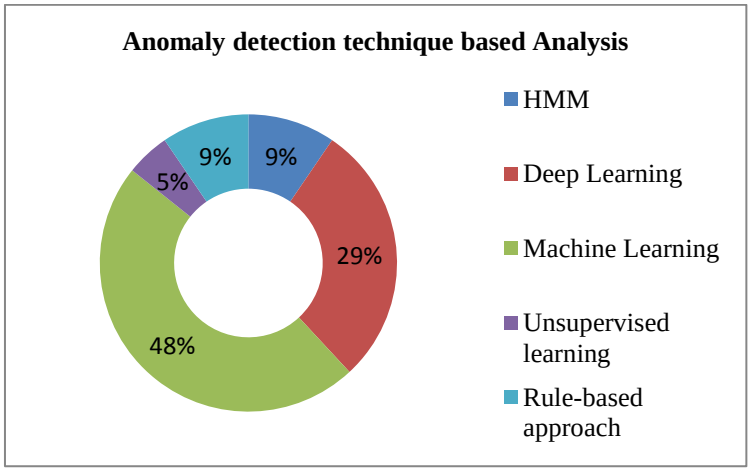

Fig. 3. Anomaly detection technique based analysis

4.3 Execution tool based Assessment

The examination of execution tools for traditional methods from the literary works is established in this part. Fig.4 depicts execution tools-based analysis. The main implementation tools employed in the research articles are MATLAB, Scorpius, Python, and Google Colab GPU environment. From Fig. 4, Python is a commonly used execution tool for anomaly detection.

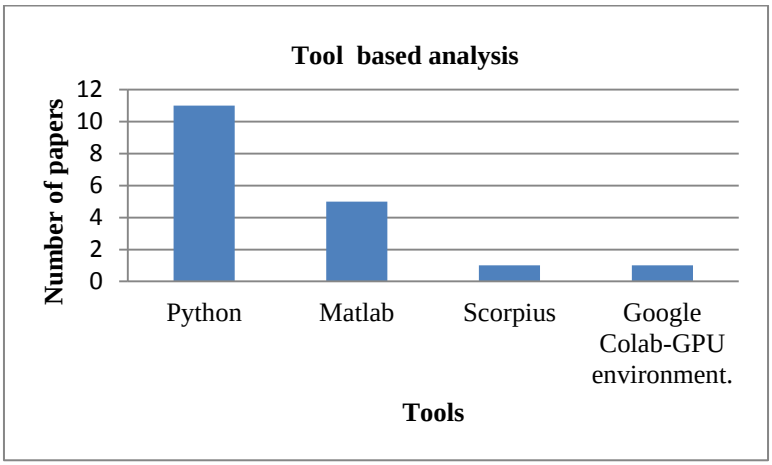

Fig. 4. Execution tool based analysis

4.4 Datasets-based analysis

This subsection signifies the study of examination using the various datasets in the literary articles. The number of datasets utilized for AD is portrayed in Fig. 5. Commonly used datasets in AD are UMN, UCSD, UEBA, Avenue, TWOS, ACODS, CASIA, ADAMS, Airbox dataset, and TIPM. Here, UCSD is frequently a dataset for anomaly detection.

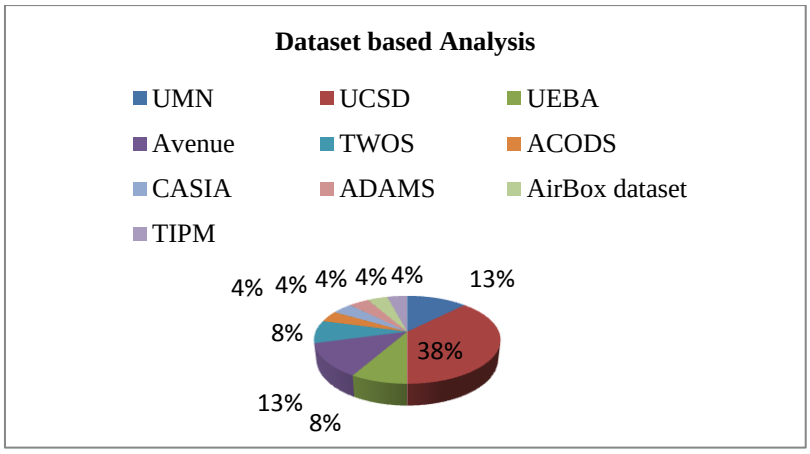

Fig. 5. Dataset based analysis

4.5 Evaluation metrics-based analysis

The various performance metrics employed for AD are represented in Fig. 6, such as Accuracy, AUC, Detection rate, F1 score, precision, TNR, TPR, FAR, and MSE. From these measures, where accuracy is very much time utilized measures for the performance examination of various anomaly prediction approaches. One of the main advantages of using accuracy metrics is that it is relatively simple, faster and effective in measuring the overall performance of the model. Thus, the accuracy is commonly employed in benchmarks and competitions.

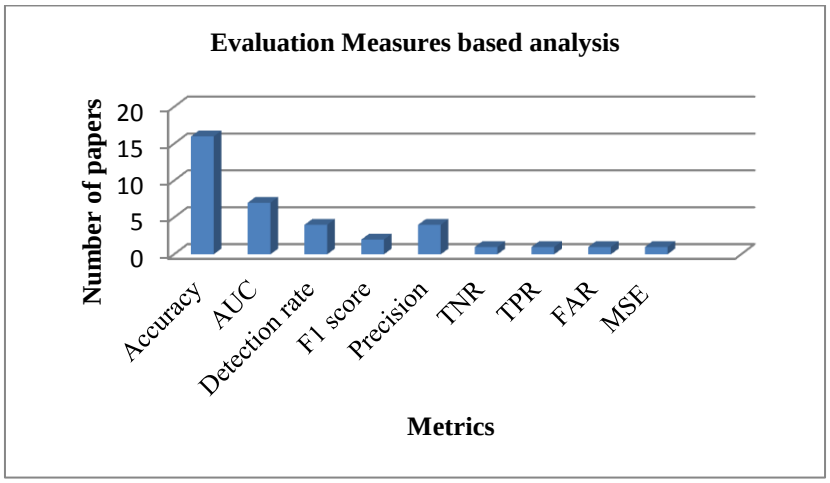

Fig. 6. Evaluation measures based on analysis

4.6. Examination Depends on Accuracy

The performance metrics-based analysis of accuracy in various AD algorithms are portrayed in Table 2. Moreover, the distinct range of accuracy involved in the research articles are 80%- 85%, 85%- 90%, 90% -95%, and 95% -100%. From the examination, it is accomplished that the research papers [31, 26, 46, 7, 53, 22, 23, 2, 4] achieved better accuracy with a range of 90% -95%.

Table 2. Examination depends on accuracy

Accuracy metric range	Research paper
80%-85%	[27]
85%-90%	-
90%-95%	[31, 26, 46, 7, 53, 22, 23, 2, 4]
95%-100%	[12, 24, 18, 32, 49]

5. Conclusion

Object recognition is categorized according to specific criteria, and numerous methods are being adapted. In this survey, AD is classified based on specific criteria, and a number of methods have been presented. The major process of this survey is to review 50 papers on AD into classified based on their years of publication, methodologies, utilized datasets, evaluation measures, software tools, and performance metrics. The process employed for AD are categorized into five types: DL-based algorithms, HMM, ML-based algorithms, Unsupervised learning, and Rule-based analysis are the most frequent ways to get accurate anomaly detection. UMN, UCSD, UEBA, Avenue, TWOS, ACODS, CASIA, ADAMS, Airbox dataset, and TIPM are some of the datasets used. The performance metrics like Accuracy, AUC, Detection rate, F1 score, precision, TNR, TPR, FAR, and MSE. The main implementation tools employed in the research papers are MATLAB, Scorpius, Python, and Google Colab GPU environment. In addition, the gaps and challenges of the research processes based on AD are elaborated more briefly to devise an effectual future scope. For the improvement of novel strategies of anomaly prediction, the main challenges listed can be determined by the researchers in future.

References

[1] J. Hu, E. Zhu, S. Wang, X. Liu, X. Guo, and J. Yin, "An efficient and robust unsupervised anomaly detection method using ensemble random projection in surveillance videos," *Sensors*, vol.19 no.19, p.4145, 2019.

[2] A. Lukashin, M. Popov, A. Bolshakov, and Y. Nikolashin, "Scalable data processing approach and anomaly detection method for user and entity behavior analytics platform," *In Intelligent Distributed Computing XIII*, pp. 344-349, *Springer International Publishing*, 2020.

[3] C. Schwab, "User and Entity Behavior Analytics for Enterprise Security," 2016.

[4] X. Xi, T. Zhang, W. Ye, Z. Wen, S. Zhang, D. Du, and Q. Gao, "An ensemble approach for detecting anomalous user behaviors," *Int. J. Softw. Eng. Knowl. Eng*, vol.28(11n12), pp.1637-1656, 2018.

[5] S. Wang, Y. Zeng, Q. Liu, C. Zhu, E. Zhu, and J. Yin, "Detecting abnormality without knowing normality: A two-stage approach for unsupervised video abnormal event detection," *In Proceedings of the 26th ACM international conference on Multimedia*, pp. 636-644, October 2018.

[6] J. Komulainen, A. Hadid, M. Pietikäinen, A. Anjos, and S. Marcel, "Complementary countermeasures for detecting scenic face spoofing attacks," *In Proceedings of 2013 International conference on biometrics (ICB)*, pp. 1-7, IEEE, June 2013,

[7] S.R. Arashloo, J. Kittler, and W. Christmas, "An anomaly detection approach to face spoofing detection: A new formulation and evaluation protocol," *IEEE Access*, vol.5, pp.13868-13882, 2017.

[8] T. Chen, and C. Guestrin, "Xgboost: A scalable tree boosting system," *In Proceedings of the 22nd acmsigkdd international conference on knowledge discovery and data mining*, pp. 785-794, August 2016.

[9] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S.K. Ng, "MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks," *In Artificial Neural Networks and Machine Learning–ICANN 2019: Text and Time Series: 28th International Conference on Artificial Neural Networks, Munich, Germany, September* 17–19, 2019, *Proceedings, Part IV* (pp. 703-716). *Cham: Springer International Publishing*, September 2019.

[10] D. Kwon, H. Kim, J. S.C., Kim, Suh, I. Kim, and K.J. Kim, "A survey of deep learning-based network anomaly detection," *Clust. Comput*, vol. 22, pp.949-961, 2019.

[11] D.C. Le, and N. Zincir-Heywood, "Anomaly detection for insider threats using unsupervised ensembles," *IEEE Transactions on Network and Service Management*, vol.18 no.2, pp.1152-1164, 2021.

[12] T. Wang, M. Qiao, Y. Deng, Y. Zhou, H. Wang, Q. Lyu, and H. Snoussi, "Abnormal event detection based on analysis of movement information of video sequence," *Optik*, vol.152, pp.50-60, 2018.

[13] A.R.M. Forkan, I. Khalil, Z. Tari, S. Foufou, and A. Bouras, "A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living," *Pattern Recognit*, vol.48 no.3, pp.628-641, 2015.

[14] M. Yamauchi, Y. Ohsita, M. Murata, K. Ueda, and Y. Kato, "Anomaly detection in smart home operation from user behaviors and home conditions," *IEEE Transactions on Consumer Electronics*, vol.66 no.2, pp.183-192, 2020.

[15] P. Kamranfar, D. Lattanzi, A. Shehu, and D. Barbará, "Multiple Instance Learning for Detecting Anomalies over Sequential Real-World Datasets," *arXiv preprint arXiv:2210.01707*, 2022.

[16] A.G. Mart ń, I.M. de Diego, A. Fern ández-Isabel, M. Beltrán, and R.R. Fernández, "Combining user behavioural information at the feature level to enhance continuous authentication systems," *Knowledge-Based Syst*, vol.244, pp.108544, 2022.

[17] M.S. Parwez, D.B. Rawat, and M. Garuba, "Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network," *IEEE Transactions on Industrial Informatics*, vol.13 no.4, pp.2058-2065, 2017.

[18] F. Skopik, M. Wurzenberger, G. Höld, M. Landauer, and W. Kuhn, "Behavior-Based Anomaly Detection in Log Data of Physical Access Control Systems," *IEEE Transactions on Dependable and Secure Computing*, 2022.

[19] Z. Liu, T. Qin, X. Guan, H. Jiang, and C. Wang, "An integrated method for anomaly detection from massive system logs," *IEEE Access*, vol.6, pp.30602-30611, 2018

[20] D.J. Hill, and B.S. Minsker, "Anomaly detection in streaming environmental sensor data: A data-driven modeling approach," *Environ. Model. Softw*, vol.25 no.9, pp.1014-1022, 2010.

[21] I.I.A. Sulayman, and A. Ouda, "User modeling via anomaly detection techniques for user authentication," *In 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 0169-0176, *IEEE*, October 2019.

[22] B. B öse, B. Avasarala, S. Tirthapura, Y.Y. Chung, and D. Steiner, "Detecting insider threats using radish: A system for real-time anomaly detection in heterogeneous data streams," *IEEE Systems Journal*, vol.11 no.2, pp.471-482, 2017.

[23] J. Cui, G. Zhang, Z. Chen, and N. Yu, "Multi-homed abnormal behavior detection algorithm based on fuzzy particle swarm cluster in user and entity behavior analytics," *Sci. Rep*, vol.12 no.1, p.22349, 2022.
[24] G. Fernandes Jr, L.F. Carvalho, J.J. Rodrigues, and M.L. Proença Jr, "Network anomaly detection using IP flows with principal component analysis and ant colony optimization," *J Netw Comput Appl*, vol.64, pp.1-11, 2016.
[25] J. Hu, E. Zhu, S. Wang, S. Wang, X. Liu, and J. Yin, "Two-stage unsupervised video anomaly detection using low-rank based unsupervised one-class learning with ridge regression," *In Proceedings of 2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, IEEE, July 2019.
[26] A.G. Martín, M. Beltrán, A. Fernández-Isabel, and I.M. de Diego, "An approach to detect user behaviour anomalies within identity federations," *Comput. Secur*, vol.108, pp.102356, 2021.
[27] K.W. Cheng, Y.T. Chen, and W.H. Fang, "An efficient subsequence search for video anomaly detection and localization," *Multimed. Tools Appl*, vol.75, pp.15101-15122, 2016.
[28] L. Feremans, V. Vercruyssen, B. Cule, W. Meert, and B. Goethals, "Pattern-based anomaly detection in mixed-type time series," *In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September* 16–20, 2019, *Proceedings, Part I* (pp. 240-256). *Springer International Publishing*, 2020.
[29] M. Tahir, M. Li, X. Zheng, A. Carie, J. Xing, M. Azhar, N. Ayoub, A. Wagan, M. Aamir, L.A. Jamali, and M.A. Imran, "A novel network user behaviors and profile testing based on anomaly detection techniques," *Int J Adv Comput Sc*, vol.10, no.6, 2019.
[30] S. Lee, H.G. Kim, and Y.M. Ro, "Bman: Bidirectional multi-scale aggregation networks for abnormal event detection," *IEEE Transactions on Image Processing*, vol.29, pp.2395-2408, 2019.
[31] W. Chen, Y. Zhang, C.K. Yeo, C.T. Lau, and B.S. Lee, "Unsupervised rumor detection based on users' behaviors using neural networks," *Pattern Recognit. Lett*, vol. 105, pp.226-233, 2018.
[32] T. Li, X. Chen, F. Zhu, Z. Zhang, and H. Yan, "Two-stream deep spatial-temporal auto-encoder for surveillance video abnormal event detection," *Neurocomputing*, vol. 439, pp.256-270, 2021.
[33] S. Lin, R. Clark, R. Birke, S. Schönborn, N. Trigoni, and S. Roberts, "Anomaly detection for time series using vae-lstm hybrid model," *In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4322-4326, May, 2020.
[34] S. Yan, J.S. Smith, W. Lu, and B. Zhang, "Abnormal event detection from videos using a two-stream recurrent variational autoencoder," *IEEE Transactions on Cognitive and Developmental Systems*, vol.12 no.1, pp.30-42, 2018.
[35] J. Henriques, F. Caldeira, T. Cruz, and P. Simões, "Combining k-means and xgboost models for anomaly detection using log datasets," *Electronics*, vol.9 no.7, p.1164, 2020.
[36] A. Garg, W. Zhang, J. Samaran, R. Savitha, and C.S. Foo, "An evaluation of anomaly detection and diagnosis in multivariate time series," *IEEE Transactions on Neural Networks and Learning Systems*, vol.33 no.6, pp.2508-2517, 2021.
[37] H. Zhang, and Z. Li, "Anomaly detection approach for urban sensing based on credibility and time-series analysis optimization model," *IEEE Access*, vol.7, pp.49102-49110, 2019.
[38] W. Ullah, A. Ullah, I.U. Haq, K. Muhammad, M. Sajjad, and S.W. Baik, "CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks," *Multimed. Tools Appl*, vol.80, pp.16979-16995, 2021.
[39] N. Li, and F. Chang, "Video anomaly detection and localization via multivariate gaussian fully convolution adversarial autoencoder," *Neurocomputing*, vol.369, pp.92-105, 2019.
[40] L. Feremans, V. Vercruyssen, W. Meert, B. Cule, and B. Goethals, "A framework for pattern mining and anomaly detection in multi-dimensional time series and event logs," *In New Frontiers in Mining Complex Patterns: 8th International Workshop, NFMCP 2019, Held in Conjunction with ECML-PKDD 2019, Würzburg, Germany, September* 16, 2019, *Revised Selected Papers* 8, pp. 3-20, *Springer International Publishing*, 2020.
[41] F. Jiang, J. Yuan, S.A. Tsaftaris, and A.K. Katsaggelos, "Anomalous video event detection using spatiotemporal context," *Com. Vis. Ima Unders*, vol.115 no.3, pp.323-333, 2011.
[42] Q. Fu, J.G. Lou, Y. Wang, and J. Li, "Execution anomaly detection in distributed systems through unstructured log analysis," *In Proceedings of 2009 ninth IEEE international conference on data mining*, pp. 149-158, IEEE, December 2009.
[43] I. Friedberg, F. Skopik, G. Settanni, and R. Fiedler, "Combating advanced persistent threats: From network event correlation to incident detection," *Comput. Secur*, vol.48, pp.35-57, 2015.
[44] H.K. Kim, K.H. Im, and S.C. Park, "DSS for computer security incident response applying CBR and collaborative response," *Expert Syst. Appl*, vol.37 no.1, pp.852-870, 2010.
[45] Z. Lijun, H. Guiqiu, L. Qingsheng, and D. Guanhua, "An intuitionistic calculus to complex abnormal event recognition on data streams," *Secur Commun Netw*, pp.1-14, 2021.
[46] P. Radoglou-Grammatikis, P. Sarigiannidis, E. Iturbe, E. Rios, S. Martinez, A. Sarigiannidis, G. Eftathopoulos, Y. Spyridis, A. Sesis, N. Vakakis, and D. Tzovaras, "Spear siem: A security information and event management system for the smart grid," *Com. Net*, vol.193, pp.108008., 2021.
[47] M. Saebi, J. Xu, L.M. Kaplan, B. Ribeiro, and N.V. Chawla, "Efficient modeling of higher-order dependencies in networks: from algorithm to application for anomaly detection," *EPJ Data Sci*, vol.9 no.1, p.15, 2020.
[48] Y.S. Chong, and Y.H. Tay, "Abnormal event detection in videos using spatiotemporal autoencoder," *In Advances in Neural Networks-ISNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June* 21–26, 2017, *Proceedings, Part II* 14, pp. 189-196, *Springer International Publishing*, 2017.
[49] R.V.H.M. Colque, C. Caetano, M.T.L. de Andrade, and W.R. Schwartz, "Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos", *IEEE Transactions on Circuits and Systems for Video Technology*, vol.27 no.3, pp.673-682, 2016.
[50] L.J. Chen, Y.H. Ho, H.H. Hsieh, S.T. Huang, H.C. Lee, and S. Mahajan, "ADF: An anomaly detection framework for large-scale PM2. 5 sensing systems," *IEEE Internet of Things Journal*, vol.5 no.2, pp.559-570, 2017.
[51] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," *Neurocomputing*, vol. 262, pp.134-147, 2017.
[52] N. Cao, C. Shi, S. Lin, J. Lu, Y.R. Lin, and C.Y. Lin, "Targetvue: Visual analysis of anomalous user behaviors in online communication systems," *IEEE Transactions on Visualization and computer graphics*, vol.22, no.1, pp.280-289, 2015.

[53] X. Xi, T. Zhang, D. Du, G. Zhao, Q. Gao, W. Zhao, and S. Zhang, "Method and System for Detecting Anomalous User Behaviors: An Ensemble Approach," *In SEKE*, pp. 263-262, 2018.
[54] T. Laue, T. Klecker, C. Kleiner, and K.O. Detken, "A SIEM Architecture for Advanced Anomaly Detection," Open *Journal of Big Data*, vol.6 no.1, pp.26-42, 2022.
[55] V. Muliukha, A. Lukashin, L.V. Utkin, M. Popov, and A.A. Meldo, "Anomaly Detection Approach in Cyber Security for User and Entity Behavior Analytics System," *In ESANN*, pp. 251-256, October 2020.

Authors' Profiles

L. Lanuwabang, a Research Scholar at Kalasalingam Academy of Research and Education, is the Director at National Institute of Electronics & IT under the Ministry of Electronics & IT, Government of India. He holds an M.Tech in Computer Science and Engineering with over 21 years of experience in leadership, teaching, and project management. With 15+ years of expertise in digital forensics and cyber security , he developed tools like the Live Web Acquisition Tool, LEA Helpdesk, and Cyber Crime Case Management Systems. He developed interactive cybersecurity educational apps in 22 Indian and 7 foreign languages. He has promoted entrepreneurship by establishing NIELIT Kohima's Incubation Centre, nurturing 25+ startups. His contributions in curriculum design, 15+ project execution, and digital forensics solutions have significantly supported national growth and enhanced NIELIT's impact.

Prof. Dr. P. Sarasu is Director International and Industry relations, Professor in Computer Science and engg dept at Kalasalingam Academy of Research and Education, Tamilnadu, India. She holds a Ph.D Degree in Computer Science and Engg . She did her masters in Embedded Systems Technology and Bachelors in Computer Science engineering. Her research areas include Chaotic Systems, Cryptography and autonomous vehicle. She played a major role in creating innovation entrepreneurial ecosystem and more than 120 startups are created under her guidance and mentorship. She is continuously involved in mentoring students and faculty members for innovation and entrepreneurship. One patent is granted for her as one of the inventor.

How to cite this paper: L. Lanuwabang, P. Sarasu, "Detection of Anomalies Based on User Behavioral Information: A Survey", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.15, No.3, pp. 54-65, 2025. DOI:10.5815/ijwmt.2025.03.04