IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 August 2023, accepted 7 September 2023, date of publication 4 October 2023, date of current version 20 October 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3321800*

SURVEY

# A Comprehensive Survey of Machine Learning Methods for Surveillance Videos Anomaly Detection

**NOMICA CHOUDHRY[1,2], JEMAL ABAWAJY[1], SHAMSUL HUDA[1], AND IMRAN RAO[3]**
[1]Faculty of Science, Engineering and Built Environment, Deakins University, Burwood VIC 3125, Australia
[2]Department of Computer Science, National University of Modern Languages, Rawalpindi 44000, Pakistan
[3]Blue Brackets Technologies, Islamabad 45730, Pakistan

Corresponding author: Nomica Choudhry (choudhryn@deakin.edu.au)

**ABSTRACT** Video Surveillance Systems (VSSs) are used in a wide range of applications including public safety and perimeter security. They are deployed in places such as markets, hospitals, schools, banks, shopping malls, offices, and smart cities. VSSs generate a massive amount of surveillance data, and significant research has been published on the use of machine learning algorithms to handle surveillance data. In this paper, we present an extensive overview and a thorough analysis of cutting-edge learning methods used in VSSs. Existing surveys on learning approaches in video surveillance have some drawbacks, such as a lack of in-depth analysis of the learning algorithms, omission of certain methodologies, insufficient critical evaluation, and absence of recent learning algorithms. To fill these gaps, this survey provides a thorough examination of the most recent learning algorithms for anomaly detection. A critical assessment of the algorithms including their strengths, weaknesses, and applicability as well as tailored classifications of anomaly types for different domains are provided. Our study also offers insights into the future development of learning techniques in VSS, positioning itself as a valuable resource for both researchers and practitioners in the field. Finally, we share our thoughts on what we learned and how it can help with new developments in the future.

**INDEX TERMS** Machine learning, anomaly detection, video surveillance systems, supervised learning methods, unsupervised learning methods, semi-supervised learning methods.

## I. INTRODUCTION

THE rapid progress in closed-circuit television (CCTV) technology, along with advancements in its underlying infrastructure [53] – including network, storage, and processing hardware – has led to the surge in surveillance cameras globally. The projected worldwide market for surveillance cameras is estimated to reach a value of US $ 39.13 billion [106] by 2025, with a compound annual growth rate (CAGR) of 8.17% from 2021 to 2025. These surveillance cameras generate a lot of surveillance data. To make any sense of this data, it not only needed to be properly managed but also required to be continuously monitored and analyzed to detect anomalies. These anomalies, also known as abnormalities [59], novelties [231], and outliers [30], are described as instances of unusual characteristics or the occurrence of typical features in unexpected places or moments.

In the context of surveillance systems, a variety of video anomalies may need identification, including atypical motions, uncommon behaviors, or items that are out of place in a particular setting. Video anomaly identification is a specialized area within the broader realm of understanding behavior patterns [179]. We grouped these anomalies based on their characteristics and described them in Table 1.

However, given the immense volume of real-time video generated in video surveillance systems [195], manual analysis to detect anomalies by human operators [53] is not only inefficient and costly but also nearly impossible. This situation creates a significant demand for automated

The associate editor coordinating the review of this manuscript and approving it for publication was Xianzhi Wang.

© 2023 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

and intelligent methods to analyze video footage and detect anomalies.

Machine Learning (ML) [36] has emerged as a powerful approach for automated video anomaly detection. Extensive research has been conducted to explore and devise various anomaly detection methods. These learning methods utilize machine learning techniques that can be categorized into three main types: supervised, unsupervised, and semi-supervised. Figure 1 provides a visual representation of these techniques, illustrating the breadth and depth of our survey coverage.

Moreover, it is well known that the effectiveness of an ML method relies heavily on input data quality, [158], which forms the basis for learning normal and anomalous behaviors. Due to the dependence of results yielded by learning methods on the input data, scientists and researchers seek standardized data sets to compare their suggested methods with existing ones. Table 2 presents commonly used datasets for video anomaly detection purposes.

The research on anomaly detection using ML techniques is constantly evolving and undergoing significant changes. Anomaly detection methods using ML techniques have been surveyed for their accuracy, efficiency, and comparison in the recent past [13], [24], [158], [188], [212]. We discuss some of these survey papers and critically analyzed them in the following section.

### A. RELATED WORK

Kiran et al. [101] start by introducing the concepts of anomaly detection and its importance in various application domains, such as surveillance, quality control, and healthcare. They discuss the challenges of detecting anomalies in videos, including the complexity of video data, high computational requirements, and the need for efficient and accurate models. The paper analyzes various deep learning architectures, such as AEs, CNNs, RNNs, and their variations, that have been used for unsupervised and semi-supervised anomaly detection. The authors discuss the advantages and limitations of each method, as well as the specific scenarios in which they are most effective. Although the study emphasizes unsupervised and semi-supervised deep learning techniques for video anomaly detection, it does not investigate supervised methods or other cutting-edge learning strategies that might be relevant to the reader's interests.

Raja and Sharma [158] aim to present a comprehensive review of recent advancements and future prospects in the domain of anomaly detection in surveillance video. The authors present a well-structured analysis, focusing on various aspects such as detection methods, algorithms, and available data sets. The paper also introduces a new taxonomy for crowd analysis and anomaly detection, which contributes to the organization of existing works in the field. However, the authors have not provided an in-depth analysis of the strengths and weaknesses of specific algorithms and methods.

Jebur and Hussein [91] has analyzed and summarized the deep learning techniques employed in anomaly detection (AD) for video streaming. The researchers have divided anomaly detection into two primary categories including one classification, which relies on the quantity of frames utilized in the detection process, while the other focuses on the count of anomalies within a given scene. They have evaluated the efficacy of several well-known deep learning methods for detecting anomalies and have organized them according to their network type and architectural design. Additionally, they provide a detailed list of benchmark data sets and evaluation criteria employed to assess the efficiency of these methods. However, they don't critically analyze each method for anomaly detection and consequently do not provide the user with the idea of which method best suit their problem domain.

Anoopa and Salim [13] present eight separate methods for identifying anomalies in videos, half of which incorporate deep learning approaches that use artificial neural networks to extract features from input videos automatically. Rest of them are classified under representation learning, which extracts valuable information from input videos using prior domain knowledge and applies dimensionality reduction to decrease computational complexity. The paper investigates deep learning methodologies, including the incremental spatio-temporal learning strategy, unsupervised spectral mapping, multi-layer perceptron, recurrent neural networks, and genetic algorithm-based enhancement techniques. The research examines various representation learning techniques, including optical flow-based convolutional auto-encoders, low-dimension descriptor-based detection, joint video representation based on local motion, and low-rank dictionary learning. However, the paper does not provide in-depth explanations of these methods, limiting readers' understanding of their practical implications.

Zhang and Zhang [221] carried out an extensive review of various human action feature representation methods, targeting a wide range of datasets. While their primary emphasis is on human action recognition utilizing both hand-designed features and deep learning techniques, the survey does not cover the latest developments in deep learning, such as Generative Adversarial Networks (GANs) and Autoencoders (AEs), which are employed in surveillance applications.

Javed and Jalil [90] presents a survey paper on the field of video forensics. The authors aim to review the literature, focusing on data extraction and forgery/counter forgery techniques. They present observations and reporting on current video data extraction, video forgery, and enhancement techniques, along with a discussion on copy-move detection. The authors also present a compilation of challenges faced by users and researchers in video forensics, as well as a discussion on different products used for evidence searching. They highlight the need for a better understanding of Deep Learning (DL) theory to determine the optimal number

**TABLE 1.** Types of video anomalies.

| Anomaly | Sub-Types and Explanation |
|---|---|
| **Motion-based** ([88]) | **Loitering** ([199])<br>People or vehicles lingering in a particular area for an unexpected period. |
| | **Directional Deviation ([107])**<br>Motion that diverges from the anticipated or standard route |
| | **Speed Variation ([100])**<br>Uncommon alterations in the velocity of moving objects or persons |
| **Crowd-based** ([89]) | **Crowd Formation ([11])**<br>Unexpected assembly or dispersal of a large number of individuals |
| | **Panic ([11])**<br>Disordered or erratic movement patterns, often signaling an emergency |
| **Object-based** ([88]) | **Abandoned Objects ([129])**<br>Unsupervised items left in public areas |
| | **Removed Objects ([232])**<br>Items taken or stolen from their initial position |
| | **Object Interactions ([232])**<br>Atypical interactions between individuals and objects, such as vandalism, tampering, or alteration |
| **Point-Based** ([6]) | **individual data points that differ from the norm [30]**<br>Single object differ from the norm |
| | **Removed Objects ([45])**<br>Vehicle parked in a no-parking zone |
| | **Object Interactions ([37])**<br>Individual breaking into a restricted area |
| **Contextual** ([38]) | **Unusual events are considered anomalous within context**<br>Anomalies are not outliers independently but become anomalous with context |
| | **Physical altercations judged with context ([146])**<br>Vehicle driving unusually slow in a high-speed traffic zone |
| | **Unusual situation with context ([50])**<br>Person wearing heavy winter clothing in the middle of summer |
| **Temporal** ([96]) | **Sudden Spikes** ([82])<br>Sudden increase in activity or behavior |
| | **Periodic patterns ([171])**<br>Changes in behavior or activity that occur on a regular, repeating basis |
| | **Long-term trends ([171])**<br>Gradual change in behavior or activity over a longer period |
| **Behavioral** ([162]) | **Unusual Gestures or Postures ([1])**<br>Body language raising suspicious intent |
| | **Behavior intending aggression ([41])**<br>Hostile actions or physical altercations that indicating potential for violence |
| | **Trespassing or Intrusion ([223])**<br>Unauthorized breach of restricted areas |

of layers and the number of convolutional, recurrent, and pooling layers. They also discuss the challenge of limited datasets, which can result in decreased accuracy with DL techniques. The emergence of Artificial Intelligence (AI) in the Internet of Things (IoT) poses a challenge for forensics investigations due to the complexity of IoT networks and the vast amounts of data produced by heterogeneous devices with finite memory, power, and processing capabilities. The authors also discuss the need for real-time processing of videos, which is currently a challenge for DL forensic endeavors. In a related work, Abawajy et al. [3] studied increasing threat of malware in Mobile-Internet of Things applications on edge computing platforms is studied, with a focus on identifying deceptive app behavior and exploring a broader range of permissions. This research utilizes behavioral analysis and a two-layer detection approach to

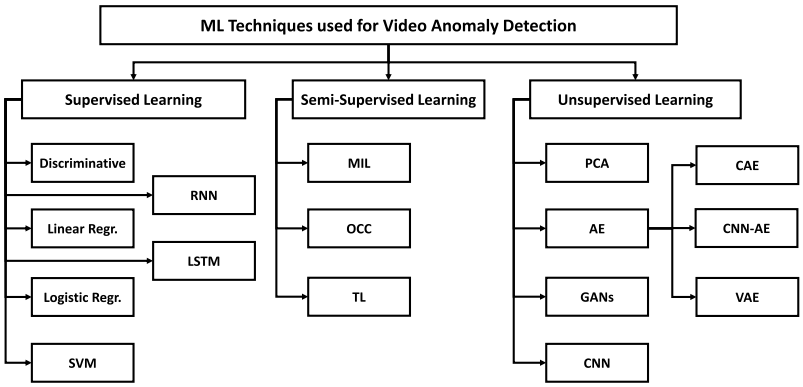


**FIGURE 1.** Machine learning (ML) techniques used by anomaly detection methods covered in this survey.

**TABLE 2.** Dataset used for video anomaly detection.

| Dataset | Total Videos | Training Videos | Testing Videos | Anomalies |
|---|---|---|---|---|
| UMN [155] | 22 | 11 | 11 | Crowd-based |
| UMCD [19] | 20 | Not Given | Not Given | Motion-based |
| UCSD Ped1 [113] | 70 | 34 | 36 | Crowd-based |
| UCSD Ped2 [113] | 28 | 16 | 12 | Crowd-based |
| Subway [7] | 1 | - | - | Point-based |
| Avenue [128] | 37 | 16 | 21 | Point-based |
| Street Scene [159] | 81 | 46 | 35 | Object-based |
| Behave [29] | 4 | 2 | 2 | Behavioral |
| UCF Crime [184] | 1900 | 1610 | 290 | Motion-based |
| ShanghaiTech [123] | 437 | 330 | 107 | Motion-based |

outperform other methods in detecting real mobile malware IoT data.

Samal and Zhang [167] proposed model that integrates an ABP-embedded Swin transformer with YOLOv3, offering a promising enhancement over traditional YOLOv3 by combining the benefits of both Swin transformer and convolutional techniques. Attention-based pooling provides a nuanced input analysis, while the Swin transformer improves scalability and focus. Strategic placement of the Swin transformer post the Darknet-53 block makes the model more device-friendly. The integration of ABP directly addresses disruptions caused by the linear embedding layer at the start of the Swin transformer, ensuring better feature interpretation. The model's practicality is demonstrated through its training on 13,500 annotated obscene images and testing on an additional 3,000. However, sourcing from copyright-free pornography sites may raise ethical questions.

Abbas and Hasan [4] introduces a deep learning method to predict the future citation impact of academic articles in the realm of Informetrics. Using temporal features, this method aims to rank articles based on potential popularity. When compared with existing models on various evaluation metrics, this new model generally outperforms them, especially in the context of predicting future network behavior. Practical applications range from identifying key influencers in social networks to anomaly detection. However, challenges like data biases, complexity, and lack of transparency exist. The authors recognize these limitations and propose directions for future research to enhance the model's effectiveness and interpretability.

Khan and Li [98] present the Growth-based Popularity Predictor (GPP) model, designed to predict and rank web content on online social media. The model considers the unpredictable nature of people's interest in online posts, which can either grow or decay at various rates. The GPP model's effectiveness was evaluated using three real datasets: Movielens, Facebook-wall-post, and Digg, and its performance was measured against four information-retrieval metrics. Results indicate the model can enhance prediction accuracy by mapping scores to cumulative predicted item rankings. Notably, the GPP model can forecast both already popular items and emerging popular items. The datasets used in this study offer insights into user interactions, modeled as both monopartite and bipartite networks, helping to predict item popularity.

Khan and Li [99] addresses the significant challenges related to the stability analysis of power systems and acknowledges the potential for further improvements. Transient disturbances in power systems are typically examined through critical contingency simulations. Proper assessment

of these disturbances is crucial for ensuring consistent power supply and preventing generating units from desynchronizing. This study aims to create a swift and efficient online tool for transient stability assessment, capable of categorizing system operations and pinpointing critical system generators during instability periods. The proposed solution is a machine learning framework that utilizes a multi-feature hybrid network, leveraging Phasor Measurement Unit (PMU) measurements to monitor system transient stability in real-time. The framework has proven to be both fast and precise, making it suitable for stability monitoring applications in power systems.

In a recent study closely related to our work, Jebur and Hussein [91] provides an overview of several real-world applications of deep anomaly detection. However, the study only provides a general description of selected categories, which makes it difficult to understand depth of approaches employed by contemporary methods and their fundamental concepts. In contrast, our analysis seeks to provide an in-depth understanding of the most recent deep detection techniques, emphasizing their key aspects, inherent capabilities, application areas, and constraints in relation to video anomaly detection.

### B. RESEARCH GAP – THE NEED OF THIS SURVEY

Although a number of studies have been dedicated to the categorization and survey of the learning methods in this area, a few studies provide a comprehensive scoping review of the field. Table 3 highlights the research gap and the need for our work and summarizes the scope of our survey along with the limitations of the other recent surveys on anomaly detection in surveillance.

In the supervised learning, the study by Roka et al. [162] noticeably lacks coverage on techniques such as Discriminative, Linear Regression, and Logistic Regression. Meanwhile, the works of Kiran et al. [101] and Jebur and Hussein [91] also do not delve into foundational supervised techniques like Discriminative, Linear Regression, and Logistic Regression, among others. The work by Rezaee and Rezakhani [161], while comprehensive in some areas, neglects methodologies like Discriminative and Linear Regression. Similarly, Suarez et al. [182] and Sreenu et al. [179] also fall short in thoroughly examining the gamut of supervised techniques. On the other hand, Behniafar et al. [24] omits a few critical supervised approaches such as Discriminative and SVM.

In the semi-supervised learning paradigm, there is a discernible gap in the literature. The works of Roka et al., Kiran et al., Sreenu et al., and Behniafar et al. do not appear to touch upon any of the listed methodologies, including MIL, OCC, and TL. Even though Jebur and Hussein and Rezaee and Rezakhani do explore TL, they overlook MIL and OCC.

Turning our attention to unsupervised methodologies, Roka et al.'s work, while including techniques like DNN, has gaps in exploring autoencoder-based methods such as AE, CAE, and CNN-AE. Similarly, Kiran et al. seems to sidestep GANs, while Jebur and Hussein skips over architectures like CAE. On a parallel note, Rezaee and Rezakhani doesn't delve deeply into generative methodologies such as GANs or specialized autoencoders like CAE. Both Suarez et al. and Sreenu et al. appear to have a similar limitation, missing out on autoencoder variants like CAE. Lastly, Behniafar et al. also misses an opportunity to discuss the nuances of autoencoder architectures like CAE and CNN-AE.

Also, there are a wide range ML methods that can be employed for anomaly detection, and choosing an appropriate learning method for the given anomaly detection application is a challenge. Therefore, there is a need for a detailed review of the learning methods w.r.t. their underlying learning technique and their strengths and drawbacks. This work fills this research gap and provides further insight by identifying the future work required.

In sum, our work stands out as a holistic survey, bridging the gaps found in various studies by offering an encompassing examination of methods across all three paradigms: supervised, semi-supervised, and unsupervised.

### C. METHODOLOGY

We employ a scoping review methodology to survey the existing literature on anomaly detection in surveillance systems providing a broad overview of the key concepts, theories, sources, and evidence available. In contrast to systematic reviews, which involve a rigorous assessment of the quality and weight of the included studies, scoping reviews typically do not assess the quality or weight of the literature considered. Instead, it focuses on presenting a comprehensive view of the research landscape. To gain access to a more comprehensive research available on the topic, we select following prominent representative search website used for academic and research purposes: (a) ACM (www.acm.org/dl), (b) The Internet Archive (www.archive.org), (c) ResearchGate (www.researchgate.net), (d) Semantic Scholar (www.semanticscholar.org), (e) Google Search (www.google.com), and (f) Google Scholar (scholar.google.com).

#### 1) IDENTIFICATION

We employed keyword combinations including ''anomaly detection'', ''surveillance'', and ''machine learning'' to search for relevant existing work. Our emphasis was on peer-reviewed research papers, cited research theses, and articles published on reputable university, organization, and company websites. A total of 839 articles were downloaded and scrutinized, later managed using reference software (Mendeley Desktop version for Windows).

#### 2) EXCLUSION

Among these, duplicates were subsequently eliminated through automated or manual processes. We also excluded articles published before year 2015, with exceptions made for seminal and highly-cited works. Articles that were not cited or only self-cited were also excluded from our consideration. The remaining articles underwent relevance

**TABLE 3.** Comparison of our survey to the related work.

| Reference | Roka et al. ([162]) | Kiran et al. ([101]) | Jebur and Hussein ([91]) | Rezaee and Reza-khani ([161]) | Suarez et al. ([182]) | Sreenu et al. ([179]) | Behniafar et al. ([24]) | Our Work |
|---|---|---|---|---|---|---|---|---|
| **Supervised** | | | | | | | | |
| Discriminative | X | X | X | X | X | X | X | ✓ |
| Linear Regr. | X | X | X | X | X | X | X | ✓ |
| Logestic Regr. | X | X | X | X | X | X | X | ✓ |
| SVM | ✓ | X | X | ✓ | X | X | X | ✓ |
| RNN | X | X | ✓ | X | X | ✓ | ✓ | ✓ |
| LSTM | X | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ |
| **Semi-Supervised** | | | | | | | | |
| MIL | X | X | X | X | X | X | X | ✓ |
| OCC | X | X | X | X | X | X | X | ✓ |
| TL | X | X | ✓ | ✓ | X | X | X | ✓ |
| **Unsupervised** | | | | | | | | |
| DNN | ✓ | X | X | ✓ | ✓ | ✓ | ✓ | ✓ |
| CNN | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | X | ✓ |
| GANs | ✓ | ✓ | ✓ | X | ✓ | ✓ | X | ✓ |
| AE | X | ✓ | ✓ | X | ✓ | X | X | ✓ |
| CAE | X | ✓ | X | X | X | X | X | ✓ |
| CNN-AE | X | ✓ | X | X | X | X | X | ✓ |
| VAE | X | ✓ | X | X | X | X | X | ✓ |

assessment based on their titles and abstracts, leading to the exclusion of an additional unwanted studies.

#### 3) ELIGIBILITY

Following this screening, 339 articles were chosen for thorough reading and analysis. After a comprehensive review of the full texts, 177 studies were ruled out due to their lack of adherence to eligibility criteria concerning techniques used for anomaly detection and relevance with our study.

#### 4) INCLUSION

Ultimately, the screening process culminated in the inclusion of 162 studies. Of these, 75% are from the last four years, and over 30% are from the past one year alone. Out of all the papers we reviewed, 15% are literature surveys and reviews. The year-wise categorization of these papers can be found in Figure 2.

### D. CONTRIBUTIONS OF THE PAPER

Our objective is to pinpoint the constraints, inadequacies, and disadvantages of prevalent video anomaly detection approaches. We offer an exhaustive evaluation of anomaly detection methods and classify anomaly types tailored to the dataset for specific application domains. To our knowledge, no prior research has presented such an in-depth categorization of anomalies and their corresponding optimal datasets. This paper not only scrutinizes recent academic contributions but also critically assesses them. The emphasis of this paper lies in the current tendencies in the field, and its primary contribution differs from earlier reviews in this domain. The key contributions of this review encompass:

1) Surveillance anomaly detection techniques has been classified into three main groups: supervised learning systems, semi-supervised learning systems, and unsupervised learning systems.
2) A novel categorization of various types of anomalies, enabling a more systematic review approach to anomaly detection and analysis.
3) Provide a thorough evaluation of distinct anomaly detection approaches concerning the datasets, which can act as a valuable asset for professionals and researchers alike.
4) Explain in detail the applications related to surveillance for anomaly detection methods.
5) In-depth analysis of the drawbacks of anomaly detection methods related to surveillance.
6) rovide a thorough assessment of distinct anomaly detection approaches concerning the datasets, which can act as a valuable asset for professionals and researchers alike.

### E. ORGANIZATION OF THE PAPER

This paper is organized as follows: Section II reviews the supervised learning methods. Section III surveys the

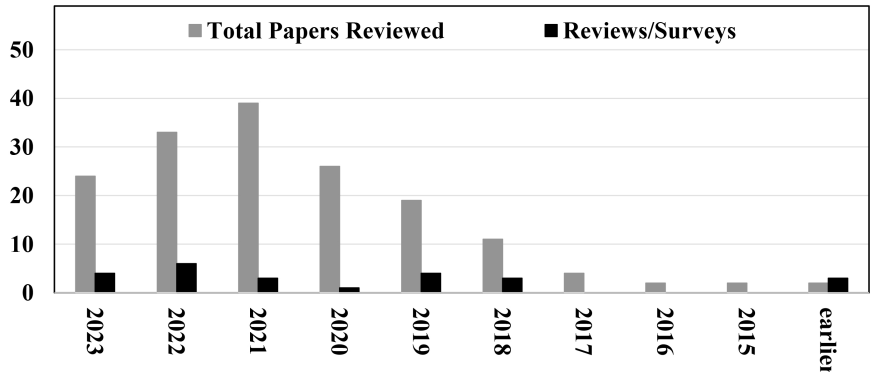


**FIGURE 2.** Publication year-wise distribution of the papers reviewed in this survey.

semi-supervised learning methods explored, while Section IV concentrates on literature survey of the unsupervised learning-based anomaly detection methods. Section V presents the key findings of this survey and provides concluding observations.

## II. SUPERVISED LEARNING METHODS

Supervised learning [194], [198], [200] employs a training dataset to instruct models to produce the expected results. The said dataset consists of inputs and their corresponding accurate outputs, facilitating the model's learning process over a period. The model's accuracy is evaluated through a metric known as the loss function. The model adjusts accordingly until the error reaches an acceptably low level. The popular supervised methods and the state-of-the-art research in the field are listed in Table 4.

### A. SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine, or SVM [157], [166], [192], is a type of supervised learning algorithm predominantly used for classification tasks, although it can also handle regression tasks. The SVM algorithm functions by representing each data item as a point in an n-dimensional space, where 'n' denotes the number of features. The value of each feature corresponds to the value of a specific coordinate. Classification is achieved by identifying the optimal hyperplane that distinctly separates the two classes. Among the many potential hyperplanes that could be chosen to segregate the two classes of data points, the goal is to find the one with the maximum margin, which is the greatest distance between the data points of both classes. By maximizing this margin distance, the algorithm can classify future data points with increased confidence. These hyperplanes serve as decision boundaries, classifying data points based on which side of the hyperplane they fall on. The dimensionality of the hyperplane is contingent upon the number of features; for instance, with two features the hyperplane is a line, while with three it becomes a two-dimensional plane. Visualizing the hyperplane grows more complex as the number of features exceeds three.

Ma and Sun [130] presents a novel model, SVM-L, developed for anomaly detection in network traffic. This model uses an innovative approach of treating raw URLs as natural language, which are then converted into mathematical vectors using statistical laws and natural language processing techniques. These vectors are subsequently used as training data for the kernel SVM classifier. The authors leverage the dual formulation of kernel SVM and linear discriminant analysis (LDA) to propose an optimization model for hyper-parameter adjustment. The simplicity of the resulting one-dimensional problem and the application of the golden section method for its resolution are commendable. The model was tested on three different datasets, demonstrating the potential applicability of this novel approach. The use of real-world datasets, from a well-known Chinese internet company and campus network traffic, adds credibility to their findings. However, the model's requirement of a pre-existing corpus of URLs for data transformation and training is a significant limitation. This necessitates the ongoing collection of emerging abnormal URLs, which might not be feasible in certain scenarios. Also, the paper does not discuss how the model performs in situations where the URLs don't exhibit characteristics included in the training data. The authors' proposal to address this issue includes adding new abnormal URLs to the training samples, updating lexicons, and retraining the classifier to detect new types of abnormal URLs. This process seems to require a substantial ongoing effort, potentially limiting the model's scalability and practical applicability. Furthermore, it would be beneficial if the authors had compared their approach with other existing methods in the literature.

Yang et al. [211] presents an in-depth exploration of One-Class Support Vector Machines (OCSVM), a leading approach for anomaly detection in machine learning, and its application in Internet of Things (IoT). The authors acknowledge the inherent challenges with conventional OCSVMs, notably significant memory requirements and computational expense, which can be prohibitive in real-world deployments with large training sets. To address this, they propose two approaches, OC-Nyström and OC-KJL, which incorporate Nyström and Gaussian Sketching techniques, clustering, and

**TABLE 4.** Popular supervised learning methods surveyed.

| Method | Ref | Anomalies | Dataset | AUC |
|---|---|---|---|---|
| Linear Regr. | [225] | Point-based | Opening Price of Baoxin Energy | X |
| | [103] | Point-based | Mauna Loa Observatory Meteorological Data | X |
| | [173] | Crowd-based | PETS2009 | X |
| | [79] | Object-based | Imperial College Healthcare NHS Trust | X |
| SVM | [83] | Object-based | 1998 DARPA BSM [120] | 81.8% |
| | [164] | Point-based | Online sources | 93–98% |
| | [20] | Motion-based | UMCD [19] | 71.23% |
| | [136] | Crowd-based | UMN [155], UCF [184] | 97% |
| | [81] | Object-based | SAIAZ | 68% |
| | [137] | Motion-based | Images | 8.41% |
| LSTM | [197] | Object-based | Gearbox and Helicopter Test Flight | 0.57% |
| | [191] | Object-based | UMN [155], UCF [184], Avenue [128] | 1.77% |
| | [165] | Point-based | Network Traffic Data | 90.5% |

gaussian mixture models. These methods demonstrate substantial speedups in prediction time and space across various IoT settings, without compromising detection accuracy. The authors use a diverse range of IoT devices for testing, including multi-purpose and simpler devices, and consider both benign and malicious anomalies. Despite the promising results, the authors recognize limitations such as model drift, where a device's normal behavior may change over time, potentially leading to false positive detections. Additionally, the authors highlight the need for output from these models to be actionable and suggest potential areas for future research, including the exploration of different data representations to further improve efficiency.

#### 1) APPLICATIONS

SVMs are widely employed in various fields including surveillance, due to their ability to handle high-dimensional data and provide accurate classification results. Examples of SVM applications [17], [78] in surveillance comprise:

- Object detection [51]: SVMs are capable of detecting and categorizing objects within a scene, including people, vehicles, and various other items. This is especially beneficial for identifying unauthorized individuals, overseeing traffic patterns, and pinpointing suspicious objects or actions.
- Motion detection [220]: SVMs can be employed to identify motion patterns within video feeds. By examining alterations in pixel intensities over time, SVMs can recognize moving objects and classify them as needed. This is advantageous for detecting unauthorized entries, monitoring the movements of individuals or objects, and assessing crowd dynamics.
- Physical Violence Detection [214]: SVMs can play a crucial role in detecting physical violence in surveillance systems. The core advantage lies in their ability to understand signs of physical violence that may not be readily apparent to human observers, thus enabling timely intervention.
- Behavior analysis [169]: SVVM, as a supervised learning model, can handle high dimensional data and generate robust classification models based on input features, making it an excellent tool for behavior pattern recognition. It works by mapping input data to a high-dimensional feature space where a hyperplane is constructed to optimally segregate different behaviors. This makes it particularly suitable for complex tasks like emotion recognition, activity tracking, or customer behavior prediction where distinct behavioral patterns need to be identified from rich and diverse datasets.
- Mass Management [169]: SVM can be employed for estimating crowd density and analyzing crowd flow in public areas or at gatherings, providing valuable insights for managing crowds and ensuring safety.

#### 2) DRAWBACKS

SVM-based classifiers have gained popularity and shown effectiveness in many applications, including surveillance. Nevertheless, there are some potential challenges associated with employing SVM-based classifiers in surveillance scenarios:

- Limited flexibility [68]: SVM-based classifiers might exhibit less flexibility compared to other machine learning algorithms. Their primary focus is identifying the optimal linear boundary between classes, which could limit their ability to capture more intricate relationships among features.
- Sensitivity to parameter tuning [68]: SVM-based classifiers come with multiple hyperparameters that require proper tuning to achieve the best performance, including the kernel function and regularization parameter. Suboptimal selection of these parameters could negatively impact the classifier's effectiveness.

- Kernel choice matters [66]: Picking the right kernel affects SVM's performance, but finding the best one can be tough.
- Sensitive to settings [124]: The performance of SVM depends on the chosen parameters, and it's hard to figure out the best values.
- Needs a lot of memory [68]: SVM requires storing large amounts of data, making it memory-intensive for big data sets.
- Works best for two groups [68]: SVM mainly handles two-group problems, although more groups can be managed using other strategies.
- No probability explanation [68]: SVM doesn't give a probability-based reason for its decisions, which might be a problem in some cases.
- Limited interpretability [174]: SVM-based classifiers are sometimes considered as black-box models, as they don't offer clear explanations for their decisions. In surveillance applications, comprehending the rationale behind a specific decision or classification is crucial, which could be a limitation with these classifiers.

In summary, SVM-based classifiers can prove valuable in surveillance applications; however, it is essential to use them carefully and in conjunction with other methods to address their limitations.

### B. RECURRENT NEURAL NETWORK (RNN)

Recurrent Neural Networks (RNN) [141] are a form of supervised deep learning well-suited for handling sequential data. RNNs differ from feed-forward neural networks due to their looped connections, enabling them to retain hidden states across time steps. As a result, RNNs excel at learning patterns and dependencies within sequences.

Murugesan and Thilagamani [141] suggests an anomaly detection technique for surveillance videos. They called their proposed method multi-layer perception recurrent neural network (MLP-RNN). Their primary focus is on improving accuracy and reducing computational time. Even though they have use their method on various datasets and have compared the results, the proposed method have some limitations. The paper requires more detail on architecture, training procedures, and other contributing factors, along with visualizations and examples. Additionally, a clearer explanation of evaluation metrics used for performance assessment would help readers better understand the method's strengths and weaknesses.

Gautam and Henry [64] introduce an RNN-based deep learning method for anomaly detection in time-series data, by examining various LSTM and GRU combinations and sequence directions, with the Adaptive Gradient (AdaGrad) optimizer proving most suitable. The model's precision is slightly lower than other classifiers like Random Forest, Naive Bayes, KNN, and Ensemble methods, but it utilizes fewer features for each sub-dataset. However, the authors don't analyze the computational complexity or discuss the interpretability of their model. Information on training time and resource requirements would be useful, and addressing the ''black-box'' nature of deep learning models could enhance the paper's value.

#### 1) APPLICATIONS

RNNs can be employed in numerous applications related to surveillance, some are:

- Anomaly detection [141]: RNNs have the ability to learn and represent the standard operations or patterns within a supervised environment or system. By recognizing any variations from the anticipated behavior, RNNs can aid in pinpointing anomalies or possible risks in monitored information. This is valuable for uncovering unexpected events, pinpointing unwarranted access, detecting trespassers, or observing questionable actions.
- Video analysis [141]: RNNs are effective in evaluating video content within monitoring systems, enabling object detection, event categorization, and identification of actions or movements. By examining video frames in a sequence, RNNs can grasp time-related data and correlations, resulting in a more comprehensive understanding of the scenario.
- Crowd analysis [11]: RNNs can be utilized for examining the actions of individuals in communal areas, like airports, railway terminals, or sports arenas. Through evaluating the motion tendencies and interpersonal dynamics, RNNs contribute to detecting unusual activities, calculating the number of people in a space, and forecasting possible areas of overcrowding.
- Object tracking [11], [67]: RNNs can be integrated with additional deep learning techniques, such as Convolutional Neural Networks (CNNs), to monitor items or people in video footage. By leveraging the time-related relationships in the information, RNNs can assist in preserving the continuity of object identities throughout the sequence.
- Audio surveillance [47]: RNNs can be employed for handling and examining audio information gathered from monitoring systems. They can identify particular sounds like gunshots, glass shattering, or distinguish distinct noises or vocalizations that could signal potential hazards or unlawful actions.
- Predictive analytics [28]: RNNs can be utilized for forecasting upcoming occurrences by examining previous monitoring data. This enables security staff to gain insights into possible threats or incidents, empowering them to implement preemptive strategies to avert or minimize potential issues.
- Multi-modal data fusion [92]: RNNs can be used to combine and evaluate information from various inputs or forms, including video, audio, and sensor data, offering a more complete view of a supervised setting.

#### 2) DRAWBACKS

Overall, while RNNs [115], [145], [156] can be effective in surveillance systems, there are the following potential drawbacks of using RNN-based classifiers in surveillance:

- Computationally expensive [149]: RNNs necessitate substantial computational resources for training and execution, particularly when working with extensive datasets. This can pose a considerable obstacle for surveillance systems that must process vast quantities of data in real time.
- Complicated activation function [149]: When employing ReLU or tanh as activation functions, processing extensive sequences becomes highly time-consuming.
- Limited memory [77]: RNNs possess a restricted memory capability, which may hinder their ability to recognize long-range relationships within the information. This limitation can pose challenges for monitoring systems that require pattern detection over prolonged duration, such as tracing an item or individual's movement throughout a time span.
- Noise Intolrence [116]: RNNs can be vulnerable to noises in input information, a typical concern in monitoring situations where external elements, like lighting conditions or weather, can impact the quality of captured images or videos.
- Inflexibility with variable-length sequences [215]: Although RNNs can manage sequences of varying lengths, designing them to excel across a broad spectrum of sequence durations can be demanding. This factor may influence their effectiveness in surveillance systems, where the duration of relevant events may vary.

### C. LOGISTIC REGRESSION

Logistic regression [138] is a statistical method used to analyze and predict the relationship between a binary outcome (two possible results) and one or more input features or variables. In simple words, it helps to determine the probability of an event occurring (e.g., success or failure, yes or no) based on certain factors or characteristics. Logistic regression works by converting these probabilities into a continuous range of values (usually between 0 and 1) using a mathematical function called the logistic function. This allows for easier interpretation and prediction of the binary outcome based on the given input features.

Wang and Cherian [200] put forth a different model-based strategy using logistic regression. Their approach is based on the idea that among the numerous deep features created from short video segments, at least one feature can accurately describe the video. They theorize that a hyperplane might separate this unknown yet distinguishing feature from the rest. However, their SVM-based pooling method struggles with large datasets, and their approach's performance drops when the number of features surpasses the training data.

Guo et al. [71] introduce a deep learning approach centered on discriminative multiple spatial-spectral feature fusion. Their proposed method relies on a CNN-based technique for Hyperspectral Image (HSI) classification. They train shallow layers to collect comprehensive information, which is then expertly combined with semantic details extracted from deeper layers. Nevertheless, their model could be enhanced by incorporating a more extensive array of spatial spectral features.

#### 1) APPLICATIONS

Logistic regression is widely employed in monitoring systems due to its ability to predict the likelihood of a specific event or result. Various uses of logistic regression in surveillance contexts are as follows:

- Fraud Detection [138]: Logistic regression enables the estimation of the likelihood of a transaction or activity being fraudulent by utilizing historical records and associated risk elements. This assists financial institutions in identifying and averting fraudulent behavior.
- Security Surveillance [16]: Logistic regression is capable of estimating the chances of a security compromise or breaches, considering behavioral trends and additional risk aspects. This allows security professionals to identify and mitigate possible security occurrences.
- Traffic Surveillance [150]: Logistic regression is use to model the probability of traffic incidents and rules violations based on environmental and behavioral factors. This data helps in pointing out areas where risk is more and introduce specific measures to decrease the frequency of accidents and violations.

#### 2) DRAWBACKS

Logistic regression has the following limitations when used for surveillance applications:

- Assumes independence of observations [109]: In logistic regression, one of the key assumptions is the independence of observations, also known as the independence of errors. This means that the residuals (errors) of the predicted responses should not be correlated. Each observation in the dataset should represent a separate and distinct case. The problem with violating this assumption is that it can lead to biased estimates of the regression coefficients and the standard errors, which, in turn, can lead to incorrect inference.
- Overfitting [163]: If there are fewer observations than features, using logistic regression is not recommended, as it can result in overfitting.
- Logistic regression may not be able to grasp intricate relationships [163] as effectively as more powerful and compact algorithms like Neural Networks can.
- Limited to binary outcomes [61]: Logistic regression is constrained to modeling binary outcomes, which may not be appropriate for all surveillance applications. For instance, it might not be useful for predicting the severity of a disease.

- Limited to linear decision boundaries [61]: Logistic regression is a linear classifier, meaning that it uses a linear decision boundary to separate classes. This linearity is in the parameters of the model - it assumes a linear relationship between the logit of the outcome variable and the predictor variables. With the help of feature engineering, such as creating polynomial or interaction terms, logistic regression can capture non-linear relationships as well. However, this process involves manual feature creation and might not always be straightforward or possible for high-dimensional data.

### D. LINEAR REGRESSION

Linear Regression [103], [200], [212] is a basic statistical method that helps to understand the relationship between a dependent variable and one or more independent variables. In simpler terms, it's a way to predict an outcome (like a price or a score) based on certain factors or inputs (such as size or time). Linear regression works by finding the best-fitting straight line, called the regression line, that represents the relationship between the variables. The goal is to minimize the differences or errors between the actual data points and the predicted points on the regression line.

In surveillance, linear regression can be applied to analyze and predict relationships between various factors and outcomes. In simple terms, it helps to estimate how one or more features impact a specific outcome, like the number of people in a room or the temperature of an area, or noise levels. By finding the best-fitting straight line that represents these relationships, linear regression can assist in predicting outcomes and trends in surveillance situations. This information can then be used to make informed decisions, improve security, and optimize the performance of surveillance systems.

#### 1) APPLICATIONS

Linear regression has numerous uses in the realm of surveillance, focusing on the examination and forecasting of data trends to improve security and observation initiatives. Some of these applications consist of:

- Linear regression is capable of examining and forecasting crowd density [212] by considering multiple factors, including time, location, and associated events. This data is valuable for managing public areas, devising security strategies, and efficiently distributing resources during events with high attendance.
- Anomaly detection: Linear regression can be utilized to identify unusual trends or actions [210] in monitoring data by examining the connections between various factors. The model can detect deviations from known patterns, which may indicate potential security risks or questionable activities.
- Object tracking: In surveillance systems [104], linear regression can be employed to estimate the motion of entities or people. The model can project future paths by assessing the historical locations of a subject, thereby facilitating more effective tracking and observation.
- To guarantee optimal performance, surveillance systems necessitate routine updates. Linear regression can be utilized to predict [185] maintenance needs, taking into account aspects such as usage, degradation, and environmental factors. This enables the optimization of maintenance planning and reduces system downtime.
- Camera placement optimization: Linear regression is useful for establishing the connection between camera positioning and coverage effectiveness. Examining elements such as the quantity of cameras, their placement, and their viewing range, the model can estimate the ideal camera locations to enhance coverage while reducing areas with limited visibility.
- Traffic Monitoring: Linear regression [139] can be employed to monitor traffic and congestion by considering aspects such as time, weather conditions, and road structure. This data can aid in more efficient traffic management, alleviating congestion, and enhancing overall transportation effectiveness.

#### 2) DRAWBACKS

Linear regression [132] can have several drawbacks when used for surveillance applications, including:

- Non-linear relationships [185]: In the context of surveillance, using linear regression could be problematic when the relationship between the independent and dependent variables is not linear. Linear regression is an algorithm based on a linear approach where a change in the predictor leads to a proportional change in the outcome variable. If this assumption is violated (i.e., the relationship between the variables is nonlinear), the model might lead to inaccurate predictions and conclusions.
- High dimensionality [58]: High-dimensionality in data can pose substantial challenges for linear regression models in surveillance. The term ''high-dimensionality'' usually refers to a situation where the number of features (independent variables) in a dataset is quite large. Issues arrises are Overfitting, Multicollinearity and Curse of dimensionality.
- Incomplete data [132]: Surveillance data could be incomplete or have missing values. Linear regression necessitates comprehensive data for every variable in the model, complicating its usage in surveillance applications.
- Assumption of normality [185]: Linear regression also presumes that residuals (the discrepancies between predicted and actual values) follow a normal distribution. If this assumption is not satisfied, the results might be unreliable.
- Causation versus correlation [26]: In surveillance applications using linear regression, it's important to exercise

caution when interpreting correlation between variables. A key tenet is that correlation does not imply causation; a relationship between variables, such as the number of security cameras and reported crimes, does not necessarily mean one directly causes the other. Multicollinearity, a condition where independent variables are highly correlated, can confound the individual effects of each variable, complicating interpretation and prediction. Spurious correlation, where two variables appear related but have no causal connection or are linked through a third variable, can also mislead analyses. Therefore, careful examination of data relationships is essential in surveillance studies using linear regression.

- Limited to continuous variables [183]: Linear regression is a statistical modeling technique that predicts a continuous dependent variable based on one or more independent variables. Therefore, it's indeed limited to dealing with continuous variables as the output. However, the independent variables in a linear regression model can be either continuous or categorical. When categorical variables are used, they are typically converted into dummy variables (also known as indicator variables) for inclusion in the model. A dummy variable is a binary variable that indicates whether a certain category of a categorical variable is present or not.

### E. LONG SHORT-TERM MEMORY (LSTM)

Long Short-Term Memory (LSTM) [121] is a type of RNN architecture designed to handle sequential data and effectively capture long-range dependencies within it. Unlike standard RNNs, LSTMs are better at remembering information from earlier time steps due to their unique cell structure that helps mitigate the vanishing gradient problem. This problem occurs when gradients become too small during training, causing RNNs to struggle with learning long-term dependencies.

LSTMs consist of memory cells and three gates: input, forget, and output gates. These gates regulate the flow of information within the LSTM cell, allowing it to selectively remember, forget, or update the cell's internal state based on the input sequence. As a result, LSTMs can learn complex patterns and dependencies in time series data, making them highly suitable for various applications, including natural language processing, speech recognition, and time series prediction.

Ullah et al. introduces a highly effective anomaly detection framework in their study [190]. This framework employs a deep features-based approach with minimized time complexity. Features are extracted from a sequence of frames based on space and time. These characteristics are then inputted into an already trained convolutional neural network. The deep features that are extracted are further processed by a multi-layered Bi-directional LSTM model, efficiently pinpointing anomalies and unusual patterns. The proposed application domain focuses on detecting anomalies in intricate surveillance scenarios within smart cities.

Hussain and Muhammad proposes an MVS framework in their study [85], where multi-view videos are divided into segments based on the presence of humans and vehicles. These segments, accompanied by a timestamp, are stored in a look-up table. After processing, deep features are extracted and sent to an LSTM, which is trained to generate probabilities according to informative and non-informative categories. The final summary consists of data exhibiting the highest probability of informativeness. However, their current approach utilizes a complicated and resource-demanding CNN model. This could be substituted with a streamlined deep-learning approach that provides comparable or even improved accuracy.

#### 1) APPLICATIONS

LSTM networks, a type of RNN, are highly useful in surveillance systems due to their ability to process sequential data and recognizing patterns over time. They can effectively detect anomalies or unusual activities by learning from temporal dependencies in surveillance video sequences. Major applications are:

- Crowd Behavior Recognition [179]: LSTM networks are capable of examining group actions within surveillance footage, enabling the detection of potential hazards like congestion, stampedes, or aggressive events. This proves valuable in public areas such as railway stations, malls, and sports arenas.
- Gesture Recognition [75]: LSTM networks possess the ability to identify human gestures in video surveillance data, proving beneficial for uses such as sign language interpretation, activity recognition, and behavioral assessment.
- Fall Detection [32]: Utilizing LSTMs for detecting falls or mishaps in surveillance footage is particularly advantageous in overseeing elderly or susceptible individuals within care facilities.
- Recognition systems [46]: By integrating LSTM networks with convolutional neural networks, recognition in surveillance footage becomes possible, aiding in the identification of recognized offenders or lost individuals.
- Observation on traffic behavior: [9]: LSTMs can be utilized for evaluating traffic trends and forecasting traffic bottlenecks or collisions, thereby enhancing traffic control and security measures.
- Predictive Maintenance [55]: LSTM networks can anticipate potential equipment malfunctions or servicing requirements by examining sensor information from monitoring systems, minimizing downtime and boosting overall system dependability.
- Detection of unauthorized access [121]: By examining movement patterns, LSTM networks contribute to the identification of unauthorized entry into protected zones or computer infrastructures, strengthening the safety measures of both physical and digital settings.

#### 2) DRAWBACKS

LSTMs gained popularity for their ability to address the vanishing gradient issue, but they do not eliminate it entirely. This is because data must still pass from one cell to another during evaluation, and the cell structure has become more complex due to additional features, such as forget gates. These intrinsic limitations of LSTM impose some drawbacks when used for surveillance systems including:

- Hardware Inefficiencies [8]: These networks demand significant resources and time for training before they can be applied in real-world scenarios. LSTMs require high memory bandwidth because of the linear layers in each cell, which can lead to hardware inefficiencies.
- Enhanced Long-Term Memory [8]: As data mining advances, there is a growing need for models capable of retaining information for longer periods than LSTMs.
- Impact of Weight Initialization [8]: LSTMs are sensitive to different random weight initialization, which can cause them to behave similarly to feed-forward neural networks. Therefore, they tend to perform better with smaller weight initialization.
- Interpretability [140]: LSTMs face challenges in comprehending the rationale behind their predictions, which may pose issues in surveillance systems where clarity and interpretability are crucial for maintaining accountability and confidence.
- Resource Intensive Nature [151]: A major drawback of using LSTM in surveillance is its computational intensity, particularly for real-time applications. Training and implementing LSTM models, especially on high-resolution video data, can require significant computational resources and time, which may not be feasible for real-time surveillance systems.

### F. SUMMARY OF SUPERVISED LEARNING METHODS

In this section, we provided a comprehensive overview of several supervised methods, emphasizing both their applications and inherent limitations. The limitations and strengths of supervised learning methods are listed in Table 5.

SVM are lauded for effectively classifying high-dimensional data, especially in object and motion pattern recognition. However, their shortcomings include sensitivity to parameter tuning and a limited capacity for interpreting intricate feature relationships. RNNs are adept at processing sequential data, finding their niche in video analysis and anomaly detection, yet they grapple with computational demands, especially for expansive datasets. Logistic Regression, ideal for predicting binary outcomes, is frequently applied in fraud detection and traffic monitoring but faces challenges stemming from presumptive independent observations and potential overfitting. LSTM networks, designed to process sequences with extended dependencies, are instrumental in recognizing crowd behaviors. However, they often encounter hardware inefficiencies due to their intensive resource demands. Finally, Linear Regression, employed for tasks like crowd density forecasting, is primarily effective with continuous variables but struggles with discerning causation from mere correlation and handling non-linear relationships. Each method, while potent in specific contexts, presents unique challenges that warrant consideration.

## III. SEMI-SUPERVISED LEARNING METHODS

Semi-supervised learning plays a significant role in surveillance applications, as it offers a practical approach to anomaly detection and pattern recognition. Obtaining a large volume of labeled data for surveillance scenarios can be both challenging and time-consuming, semi-supervised learning leverages the available labeled examples alongside a more extensive set of unlabeled data to train models effectively. This method not only improves the accuracy and efficiency of detecting irregularities, but also enhances the overall performance of the surveillance system, ensuring a more robust and reliable security solution. Popular semi-supervised learning methods in the field are listed in Table 6.

### A. MULTIPLE INSTANCE LEARNING (MIL)

Multiple Instance Learning (MIL) [34] is a unique form of supervised learning that handles ambiguously labeled data. Unlike traditional supervised learning where each instance in the training set is explicitly labeled, in MIL, instances are grouped into bags, each of which is assigned a label. If a bag is labeled as positive, it means there's at least one positive instance in it, but we don't know which one. On the contrary, if a bag is labeled as negative, all instances within it are negative. This kind of learning framework is valuable in situations where it's hard or impractical to obtain precise instance-level labels, and it finds applications in diverse fields like image classification, text categorization, and medical diagnosis.

The paper by [184] presents a MIL method for detecting anomalies in surveillance videos. This approach utilizes two bags of video segments and assumes that each bag has at least one anomalous instance. A deep MIL ranking loss is employed during training, which penalizes the classifier if the positive bag's highest score is lower than that of the negative bag. The proposed technique demonstrates across a range of benchmark datasets for anomaly detection in surveillance videos. This work also presents a novel dataset for detecting anomalies in traffic surveillance videos, featuring diverse abnormalities like accidents, traffic jams, and road infractions. The technique comes with multiple advantages such as efficient anomaly detection, noise-resistant annotations, and scalability. Nonetheless, it also exhibits some constraints like limited compatibility, high computational requirements, and rigidity towards bag size. A more flexible MIL algorithms that can be capable of handling different durations of anomalous events and varying levels of anomaly is more desirable.

Zhou and Song [230] introduces a semi-supervised strategy for anomaly detection leveraging Autoencoders (AEs).

**TABLE 5.** Summary of supervised learning methods surveyed.

| Method | Findings | Limitations |
|---|---|---|
| **SVM** | SVM is effective for classification with high-dimensional data including objects, motion patterns, identifies violence, and analyzes behavior | • Limited flexibility in capturing intricate relationships among features.<br>• Sensitivity to parameter tuning, especially kernel choice.<br>• Requires careful parameters for optimal performance.<br>• Memory-intensive for large datasets.<br>• Works best for two-group problems.<br>• Lack of probability-based explanations for decisions.<br>• Limited interpretability in surveillance contexts. |
| **RNN** | RNN excels at handling sequential data and patterns. It's useful for anomaly detection, video analysis, crowd analysis, and object tracking. | • Computationally expensive for extensive datasets.<br>• Activation functions like ReLU or tanh can lead to time-consuming processing.<br>• Limited memory for recognizing long-range relationships.<br>• Vulnerable to noise in input data.<br>• Challenges in training with noisy or incomplete data.<br>• Lack of transparency and explainability.<br>• Inflexibility with variable-length sequences. |
| **Logistic Regr.** | Logistic regression is useful for predicting binary outcomes based on input features. It finds applications in fraud detection, security surveillance, and traffic monitoring. | • Assumes independence of observations.<br>• Overfitting if fewer observations than features.<br>• Limited to linear decision boundaries.<br>• Limited to continuous variables.<br>• Challenges with non-linear relationships.<br>• Incomplete data can pose challenges.<br>• Assumption of normality in residuals.<br>• Causation versus correlation must be carefully considered. |
| **LSTM** | LSTM is powerful for processing sequential data with long-range dependencies. It's used for crowd behavior recognition, gesture recognition, fall detection, and more. | • Hardware inefficiencies due to resource demands.<br>• Enhanced long-term memory requirements for evolving data mining.<br>• Sensitivity to weight initialization.<br>• Interpretability challenges.<br>• Resource-intensive nature for real-time applications. |
| **Linear Regr.** | Linear Regression is used for Crowd density forecasting, Anomaly detection, Object Tracking an d Monitoring | • Limited to continuous variables.<br>• Causation versus correlation prolem.<br>• Non-linear relationships. |

The procedure involves extraction of features from videos by using encoding network, which are then fed into an AE for reassembly. The divergence in the reconstruction, noted as the reconstruction error, acts as an anomaly score, with larger discrepancies hinting at a higher probability of anomalies. The study conducts an in-depth analysis of the experimental outcomes and provides valuable understanding about the strengths and weaknesses of the proposed technique. A notable limitation of this method is its substantial computing resource requirements for training and testing the model. Furthermore, the approach's interpretability is restricted, posing a challenge in real-life applications

**TABLE 6.** Popular semi-supervised learning methods surveyed.

| Method | Anomalies | Dataset | AUC |
|---|---|---|---|
| MIL [96] | Temporal | UCF [184], ShanghaiTech [123] | 79.49% |
| MIL [216] | Behavioral | XD-Violence [204] | 82.17% |
| MIL [184] | Pint-based | UCF [184], the Dataset [184] | 75.41% |
| MIL [111] | Point-based | the Dataset [111] | 78.2% |
| OCC [148] | Behavioral | UCSD Ped1 [113], Avenue [128], UMN [155] | 78.30–88.60% |
| OCC [114] | Motion-based | NSL-KDD, CIC-IDS2017, MAWILab | 95.71–99.66% |

where understanding the model's decision-making process is crucial.

#### 1) APPLICATIONS

The following examples of MIL usage highlight its adaptability in tackling diverse issues in the realm of surveillance, especially when confronted with weakly labeled or noisy information.

- Temporal Activity recognition [84]: Temporal Activity Recognition utilizing MIL is a dynamic application in the field of surveillance. Instead of identifying activities based on individual frames or instances, MIL enables us to group sequences of frames into 'bags' and assign labels to these bags. In the context of temporal activity recognition, a bag would be labeled as positive if it contains at least one instance of the targeted activity, even if the exact frame or sequence exhibiting the activity isn't precisely identified. If the bag is labeled as negative, it means that none of the instances within the bag display the targeted activity. This approach is especially useful in surveillance scenarios where activities of interest are dispersed over time, and pinpointing the exact frames representing the activity can be challenging. For instance, in monitoring suspicious activities in a public area, an activity such as an unattended bag may only be significant when observed over a certain duration of time.
- Anomaly detection [97]: Utilizing MIL in surveillance data facilitates the detection of abnormal events by categorizing normal activities as negative bags and irregular activities as positive bags. In this context, a bag represents a sequence of video frames or the attributes derived from these frames. The objective is to identify the specific instances within the bag that contribute to its classification as an anomaly.
- Weakly supervised localization [34]: When faced with coarse labels in video streams, MIL can be employed to localize objects or regions of interest effectively.
- Adaptive compression: By leveraging the content of the surveillance feed, MIL can enhance video compression algorithms, leading to improved storage and transmission efficiency of video data. This optimization enables more effective utilization of resources in terms of video storage and transmission requirements.
- Privacy protection [34]: With MIL, models can be trained to automatically obscure or conceal sensitive details, like faces or license plates, within surveillance footage. This strategy ensures adherence to privacy norms by protecting personal identities and sensitive information within the captured video.
- Predictive analytic [44]: Through the analysis of historical surveillance data, MIL can assist in forecasting and mitigating future incidents, including criminal activity or traffic accidents. By leveraging patterns and insights derived from past data, MIL aids in predicting and preventing potential occurrences for enhanced safety and security.
- Customizable alerts [117]: Through training, MIL can acquire the capability to identify and detect particular events or behaviors of interest. When such incidents are recognized in real-time, MIL can promptly send alerts to security personnel, enabling timely response and action to address the detected occurrences.
- Integration with other systems [184]: MIL can be utilized to merge data from various surveillance systems such as video, audio, and sensors, resulting in a holistic comprehension of the monitored environment. By fusing information from multiple sources, MIL enhances the overall situational awareness and provides a more comprehensive understanding of the surveillance context.

#### 2) DRAWBACKS

Despite its many benefits, there are also some potential downsides to employ MIL as a machine learning strategy. Some of the primary limitations of MIL encompass:

- Task/Prediction problem on Instance level vs Bag Level [80]: In certain situations, such as finding objects in images (for content search, for example), the goal isn't to categorize groups (bags), but to classify single items (instances). The group label indicates whether the item is in the image or not. It's also important to note that a method's ability to classify bags may not accurately reflect its ability to classify individual items. For instance, when examining negative bags, just one False Positive can lead to the misclassification of the entire bag. However, in positive bags, this doesn't change the label and shouldn't impact the overall group-level loss.

- Label ambiguity [34]: Label ambiguity naturally occurs in weak supervision. In MIL, this ambiguity varies based on the problem's assumptions. With standard MIL, there's no confusion in negative bag labels. More relaxed MIL assumptions introduce additional ambiguity sources, such as label noise and differing label spaces for instances and bags.
- Lack of standard benchmarks [69]: The absence of widely accepted benchmarks and evaluation metrics for MIL makes it difficult to compare the performance of different algorithms and approaches, limiting progress in the field.
- Limited ability to handle rare classes [34]: MIL might encounter difficulties in dealing with classes that have limited representation in the labeled data. The lack of sufficient examples for these rare classes can hinder the candidate models from learning an accurate and representative model, leading to lower classification accuracy.

### B. ONE-CLASS CLASSIFICATION (OCC)

One-Class Classification (OCC) [172], also known as unary or single-class classification, is a machine learning approach that focuses on recognizing the patterns of a single class, treating all other patterns as outliers. This technique is particularly useful in situations where the data for one class (the ''normal'' class) is abundant, but the data for the other class (the ''anomalous'' class) is scarce or non-existent.

In the context of surveillance, one-class classification can be particularly beneficial. For instance, consider a surveillance system set up to monitor a secured area where intrusion is rare. The system has abundant data on the normal scenario, i.e., when no intrusion is happening. However, data on the anomalous scenario, i.e., an intrusion, is limited or even absent due to the rarity of such events. In such cases, a one-class classifier can be trained on the 'normal' data to recognize typical activity patterns within the secured area. Once trained, it can classify any deviation from these normal patterns as an anomaly, thereby detecting potential intrusions. This allows the system to effectively monitor for unusual activities, despite the lack of explicit 'anomalous' training data.

Gautam and Mishra [63] introduce a novel method for one-class classification, integrating deep learning and kernel techniques. However, the study could be enriched by a more comprehensive comparison of the newly proposed method with established techniques, encompassing an in-depth discourse on the pros and cons of each approach.

#### 1) APPLICATIONS

OCC [23] is a machine learning approach that is designed to identify and classify anomalies by learning the defining features of a particular class, usually representing as the ''normal'' behavior. It aims to detect instances that deviate from this learned norm. OCC has proven to be advantageous in diverse surveillance applications including:

- Binary Classification Problem [172]: In a simple binary classification problem involving two groups (positive and negative), a standard machine learning algorithm tries to distinguish between them and create a model that can correctly classify new, unseen examples from both groups. In a situation where a computer needs to tell apart two groups, it can struggle if one group has many more examples than the other. This is called class imbalance and can make the computer biased towards the bigger group. When the imbalance is extreme, it's hard to accurately identify the smaller, more important group. One-class classification (OCC) can help solve this problem.
- Anomaly detection [172]: The utilization of OCC allows for the representation of typical behaviors or actions within a specific setting. By pinpointing occurrences that diverge from the established standard, this technique enables the real-time recognition of uncommon or dubious activities, which could potentially avert security risks or violations.
- Cyber Security [60]: Within the realm of cyber-security, OCC proves valuable for scrutinizing network traffic. By familiarizing itself with habitual network operations, this method can identify irregular traffic behaviors or unsanctioned access efforts, thereby thwarting possible cyber intrusions.
- Crowd behavior analysis [172]: OCC can be employed to simulate standard crowd movements and recognize abnormal behavior, like unexpected distress or aggression. Such insights can assist law enforcement in foreseeing and averting incidents that threaten public safety.
- Unattended object detection [95]: OCC can be used to pinpoint items that are out of place in a supervised setting, like unattended bags or parcels, which may present security hazards.
- Motion pattern analysis: By understanding customary movement patterns within an observed zone, this technique can be used to discern anomalies in motion, such as lingering individuals or unauthorized access to restricted areas.
- Video tampering detection [143]: Additionally, OCC can be applied to detect irregularities in video data, like atypical pixel arrangements or alterations in video attributes, thereby maintaining the credibility of surveillance recordings.

#### 2) DRAWBACKS

Despite the utility of One-Class Classification (OCC) in identifying anomalies in data, there are certain potential limitations when implementing it in real-world scenarios. Here are some of those constraints:

- Learning From imbalance data [31]: One major limitation of OCC is that it does not utilize outlier instances (positive cases) during training and instead discards them. This indicates that a reverse modeling approach (e.g., treating the positive case as normal) might be worth considering concurrently. Moreover, the one-class classifier could serve as an input for a group of algorithms, with each algorithm leveraging the training dataset in distinct ways.
- Dependence on training data [31]: OCC focus on a single class of interest, and create difficulties in forming a comprehensive decision boundary, as it's trained on a single class. This could potentially make it harder to differentiate between diverse instances within that class.
- Noise Sensitivity [172]: OCC algorithm is prone to noise within the data, possibly resulting in imprecise anomaly detection. It may necessitate the use of pre-processing techniques to get rid of noise and outliers before applying OCC algorithms.
- Limited interpretability with feature selection [202]: Additionally, feature selection could indeed become a challenging task in OCC. As it deals with only one class, the features to be chosen are based solely on the minority class, which may not always offer a comprehensive understanding of the entire data landscape, as would be the case in a traditional binary or multi-class scenario.

### C. TRANSFER LEARNING

Transfer learning [12], [22] is a powerful technique in machine learning, which allows a pre-trained model from one task is used as a foundation for a second, related task. The key idea here is to harness the wealth of knowledge accumulated from the first task, typically rich in data, and apply it to enhance the learning process of a second task, which might be data-sparse.

This method is especially advantageous in the realm of surveillance. Surveillance situations are often characterized by unique settings, and gathering sufficient labeled data for training can be a daunting, expensive, or even unfeasible task. By applying transfer learning, surveillance systems can benefit from pre-existing models that were trained on similar tasks or environments, thereby conserving both time and resources.

#### 1) APPLICATIONS

Transfer learning [122] has various applications in surveillance, particularly in tasks that involve object recognition, activity recognition, and anomaly detection. Leveraging pre-trained models from related tasks can improve the performance of models in surveillance with limited training data or computational resources. Here are some applications of transfer learning in surveillance:

- Object detection [12]: TL is use to enhance the precision of object recognition models in surveillance, aiding in the identification of people, vehicles, or objects. Models pre-trained on extensive datasets like ImageNet or COCO, such as YOLO, Faster R-CNN, or SSD, can be fine-tuned to detect specific objects pertinent to a surveillance scenario.
- Public safety [22]: Transfer learning can be used in public safety applications to detect unusual events, such as fires, accidents, or other emergencies, allowing for faster response by emergency services.
- Anomaly detection [122]: The detection of anomalous events or activities is crucial in surveillance systems. Transfer learning provides a potent tool for anomaly detection, an essential task in many domains such as cyber security, fraud detection, health monitoring, and industrial fault detection. Anomaly detection involves identifying unusual or suspicious patterns that deviate from typical behavior.
- Computer Vision [178]: TL is used in pre-trained models from large-scale image recognition tasks for specific applications such as object detection, facial recognition, and image analysis.
- Crowd analysis [25]: Transfer learning can be highly effective for crowd analysis, a task that involves assessing and understanding the behavior, movement, and other characteristics of crowds. This is typically a challenging task due to the variability in crowd density, behavior, and environmental factors.

#### 2) DRAWBACKS

While transfer learning can offer numerous advantages in several surveillance use-cases, it's not without drawbacks too. Below are some of the principal constraints:

- Domain discrepancy [222]: Domain discrepancy is a critical challenge in transfer learning, where the source and target domains are different in terms of data distribution or feature space. This discrepancy may lead to a decline in the performance of the model when applied to the target task.
- Negative transfer [93]: Negative transfer is a scenario in transfer learning where the knowledge transferred from a source task adversely affects the performance of the model on the target task. This situation typically arises when the source and target tasks are dissimilar or unrelated.
- Privacy concerns [206]: Pre-trained models by using teansfer learning may raise privacy concerns. Techniques such as differential privacy and federated learning are increasingly being used to ensure privacy preservation. These methods aim to provide a balance, allowing models to learn useful representations for transfer learning while ensuring sensitive information remains confidential.
- Computational resources [144]: Transfer learning, can help reduce the need for large amounts of labeled data, obtaining a suitable pre-trained model or the data necessary for fine-tuning, does pose some challenges when

it comes to resource utilization. Pre-trained models, particularly those used for complex tasks like image recognition or natural language processing, can be quite large, requiring significant computational power and memory for both training and inference.

- Adaptability and Generalization [93]: Adjusting a pre-trained model based on a small dataset could result in overfitting, where the model becomes too tailored to the training data and performs poorly on unfamiliar data. This can pose significant issues in surveillance scenarios, where the model is expected to cope with diverse conditions like fluctuating light levels, different camera positions, or alterations in object visuals.
- Not good for large datasets [52]: Transfer learning may not yield significant improvements for tasks with larger data sets. While transfer learning can enhance the performance of machine learning models, its impact may be less pronounced for tasks involving larger data sets. Conventional learning methods begin with random weights and adjust them until convergence is reached. In contrast, transfer learning starts with a pre-trained model. However, the presence of larger data sets results in a greater number of iterations, which can diminish the importance of the initial weights.

### D. SUMMARY OF SEMI-SUPERVISED LEARNING METHODS

We provided a detailed overview of several semi-supervised methods, highlighting both their applications and limitations, listed in Table 7.

The table provides insights into various semi-supervised methods, underlining their applicability and inherent limitations. MIL emerges as an adept technique in pinpointing anomalies within surveillance videos by using labeled bags of video segments. Its proficiency extends to managing inconsistent or absent labels and varied camera angles, making it ideal for activities recognition, anomalies detection, object localization, and outcome prediction across different levels. However, MIL grapples with challenges like prediction discrepancies between instance and bag levels, ambiguous labels, an absence of standardized benchmarks, constraints in managing infrequent classes, and complexities in addressing intricate anomalies spanning multiple bags or instances. Furthermore, the method's efficiency is swayed by the choice of bag or instance representation and the similarity measure.

OCC zeroes in on single-class pattern recognition, relegating all other patterns as outliers. Its strength lies in anomaly detection, especially when 'normal' class data is profuse but 'anomalous' class data is sparse. Yet, OCC is not without its limitations. The method struggles with learning from imbalanced data, is heavily reliant on training data, exhibits noise sensitivity, and offers restricted interpretability, especially in feature selection.

Transfer Learning is recognized for leveraging a model pre-trained on one task to bolster the learning for a related task. Given the scarce labeled data and unique settings in surveillance, this technique is invaluable. Still, it's not without its caveats. Transfer learning can be hampered by domain discrepancies, the phenomena of negative transfer, privacy-related issues, the demand for extensive computational resources, challenges with adaptability and generalization, and its unsuitability for vast datasets.

## IV. UNSUPERVISED LEARNING

Unsupervised learning is a method in machine learning that discovers patterns and structures in data independently, without the need for labelled samples. Generally, unsupervised learning techniques are effective for anomaly detection when labeled data is scarce or too costly to acquire. In anomaly detection, unsupervised learning methods can identify unusual behavior or data points based solely on input features. One of the potential category of unsupervised learning is dimensionality reduction. This approach identify data points with high reconstruction errors when projected onto a lower-dimensional subspace. Principal Component Analysis (PCA) and Autoencoders (AEs) are two popular dimensionality reduction methods. PCA is used to find the lower-dimensional subspace, while AEs are neural network structures trained to reconstruct input data. In this paper, we will be focusing on Dimensionality Reduction Methods only.

### A. PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables, known as principal components. The goal is to simplify the complexity of high-dimensional data while retaining trends and patterns. It does this by transforming the data into fewer dimensions, which are constructed as linear combinations of the original variables. The first principal component captures the most variance in the data, and each subsequent component accounts for as much of the remaining variance as possible, under the constraint that it is orthogonal (uncorrelated) to the preceding components. By embracing PCA, one can efficiently and persuasively analyze complex datasets, thereby enabling more informed decision-making and enhancing overall understanding.

Arivudainambi et al. [15] presents traffic classification system by combining PCA with ANN, which can classify greater number of attacks accurately in a significantly reduced time frame. They have used dimensionality reduction approach, which minimizes data size and, consequently, the amount of data requiring training. However, it is essential to acknowledge that reducing the number of variables in a dataset may also lead to a decrease in result accuracy, as anticipated. Furthermore, it is worth noting that as the network size expands, the number of network weights increases, potentially leading to overfitting.

**TABLE 7.** Summary of semi-supervised methods surveyed.

| Method | Findings | Limitations |
|---|---|---|
| **MIL** | MIL is a way of finding anomalies in surveillance videos by using bags of video segments with labels. MIL can deal with noisy or missing labels and different camera angles. MIL can be used for recognizing activities, detecting anomalies, locating objects, and predicting outcomes at different levels. | • Task/Prediction problem on Instance level vs Bag Level<br>• Label ambiguity<br>• Lack of standard benchmarks<br>• Limited ability to handle rare classes<br>• Difficulties in handling complex anomalies involving multiple bags or instances.<br>• Sensitivity to the choice of bag or instance representation and similarity measure is another challenge in MIL. |
| **OCC** | OCC recognizing the patterns of a single class, treating all other patterns as outliers, which is useful for detecting anomalies where data for the 'normal' class is abundant but 'anomalous' class data is scarce | • Learning From imbalance data<br>• Dependence on training data<br>• Noise Sensitivity<br>• Limited interpretability with feature selection |
| **Transfer Learning** | Transfer Learning is a powerful machine learning technique where a pre-trained model from one task enhances the learning process of a related task, which is particularly useful in surveillance due to limited labeled data and unique settings | • Domain discrepancy<br>• Negative transfer<br>• Privacy concerns<br>• Computational resources<br>• Adaptability and Generalization<br>• Not good for large datasets |

Xiao and Huang [208] introduce an innovative online robust method, named Online Moving Window Robust Principal Component Analysis (OMWRPCA), designed to handle both slowly and abruptly fluctuating subspaces. This cutting-edge approach integrates hypothesis testing into an efficient online PCA framework, enabling the identification of change points within the underlying subspace. As a result, the low-rank subspace and sparse errors are concurrently predicted. The authors assert that OMWRPCA is a pioneering algorithm that not only detects change points but also computes RPCA in an online manner. This breakthrough technique is the first of its kind to manage gradually and abruptly shifting spaces online. Nonetheless, the results indicate that OMWRPCA performs well in the context of gradually changing subspaces, while its performance may be hindered when subspaces change abruptly, as the updates are not accurately applied. When the new subspace differs significantly from the original one, online updates may not be the most suitable choice, and the update process may take some time to complete.

#### 1) APPLICATIONS

PCA [152] is a dimensionality reduction technique that has been widely used in various fields, including surveillance systems. In the context of video surveillance, PCA can be employed for several purposes, such as object recognition, background modeling, and compression. Here are some specific applications of PCA in surveillance:

- Background subtraction [72]: Background subtraction is a fundamental step in video surveillance systems, as it allows for the separation of foreground objects (e.g., moving people or vehicles) from the static background. PCA can be employed to model the background by capturing the most significant variations in the scene. The low-dimensional representation of the background can then be used to identify and subtract the background, allowing for efficient foreground object detection.
- Scaling [27]: One of the assumptions underlying PCA is that the data is appropriately scaled and centralized, which can pose a challenge under certain conditions. If the data is not scaled appropriately, the extracted principal components may not accurately represent the inherent patterns within the data. This potential misrepresentation of data can be a notable limitation of PCA.
- Anomaly detection [152]: PCA can be employed to learn the normal patterns in a scene by capturing the principal components of the training data. By comparing the projections of new data onto the principal components, the surveillance system can identify deviations from normal patterns, which may indicate an anomaly or suspicious activity.

- Data compression: Video surveillance systems generate large amounts of data that need to be stored and transmitted. PCA can be used to reduce the dimensionality of the data while preserving the most important information. This compression can significantly reduce storage and transmission requirements, making the surveillance system more efficient.
- Visualize Multimedia data [27]: In the context of surveillance, PCA is employed to visualize multidimensional data, which can facilitate the analysis and interpretation of complex patterns within the collected information. This technique helps in reducing the dimensionality of the dataset while preserving the essential features, making it easier to identify potential security threats, anomalies, and trends in the surveillance data.
- Face recognition [131]: PCA, often referred to as Eigenfaces in this context, can be used for face recognition in surveillance systems. By extracting the most significant features of a face, PCA creates a lower-dimensional representation that can be used for comparison and identification. This technique is particularly useful for recognizing individuals in real time and can be applied to access control or monitoring of restricted areas.

#### 2) DRAWBACKS

Although PCA does offer many advantages for dimensionality reduction, there are some potential drawbacks to consider when employed for anomaly detection. These are:

- Loss of Interpretability in independent variables [86]: After applying PCA, the original independent variables are transformed into Principal Components, which are linear combinations of the original variables. These Principal Components may be less interpretable and understandable compared to the original variables. This is because they do not directly correspond to the initial features, making it more challenging to relate them to real-world surveillance attributes or events.
- Standardization of data [86]: In the context of anomaly detection for surveillance, standardizing the data before implementing PCA is crucial to ensure optimal results. For instance, when analyzing surveillance data, the feature set may include measurements in various units, such as kilograms, light years, or millions. The variance scale in such a dataset can be quite large, which can affect the PCA results. If the data is not standardized before applying PCA, the principal components may be biased towards features with high variance, leading to inaccurate conclusions and potentially missing important anomalies in the surveillance data.
- Information Loss [105]: In the context of anomaly detection for surveillance, it is important to carefully select the number of Principal Components to retain after applying PCA. While Principal Components aim to capture the maximum variance among the features in a dataset, choosing an insufficient number of Principal Components can lead to information loss compared to the original set of features. Information loss can impact the effectiveness of anomaly detection in surveillance, as it may cause the model to miss important patterns or trends that could be indicative of unusual or suspicious activities. To minimize this risk, it is essential to balance the need for dimensionality reduction with the preservation of key information from the original features.
- Dependence on parameter tuning [105]: A key requirement of PCA is deciding the number of principal components to utilize for reducing dimensions. Determining the ideal count can be tricky, and varying selections may lead to differing outcomes when it comes to tasks like anomaly detection. This variability poses a challenge in the effective application of PCA.

### B. AUTO-ENCODERS (AES)

Autoencoders (AEs) [48], [57] are a specific type of neural network used primarily for learning efficient codings of input data, often with the intent of dimensionality reduction. Structurally, an AE is composed of an encoder and a decoder. The encoder maps the input data to a lower-dimensional representation, also known as a latent space or bottleneck layer, while the decoder aims to reconstruct the original input data from this encoded representation. The primary goal of an AE is to minimize the reconstruction error, thereby forcing the model to retain as much of the meaningful data as possible in the encoded representation. AEs have been widely used for various tasks such as anomaly detection, denoising, and feature extraction, particularly in the fields of image and text analysis. The sub-types of AE are listed below and a list of popular AE methods we surveyed are summarized in Table 8.

- **Contractive AE (CAE)**: As discussed in [2], [201], and [218], a variant of AE known as the Contractive Autoencoder (CAE) introduces a regularization term within the loss function. This addition encourages the model to learn smooth and resilient feature representations. By making the encoding less sensitive to minor fluctuations in the training dataset, CAE achieves improved stability. This is accomplished by incorporating a regularizer or penalty term that is derived from a specific cost or objective function.
- **Convolutional AE (CNN-AE)**: The Convolutional AE (CNN-AE) [87], [119], [134] represents an advancement of convolutional neural networks, primarily used in the field of image reconstruction. These networks leverage optimal filters to minimize reconstruction errors. Once trained, CNN-AEs can be employed on various input types for effective feature extraction.
- **Variational AE (VAE)**: Variational AE (VAE) The Variational AE (VAE) [213] constitutes a generative model that learns a continuous latent space representation of data via encoding and decoding processes. In a VAE, the encoder outputs parameters of a probability distribution, typically a mean and variance of a Gaussian distribution. Then, a sample from this distribution is taken and

passed through the decoder to generate outputs. This probabilistic approach allows for more flexibility and enables VAEs to generate new data that's similar to the training data - a property that has made VAEs particularly popular for tasks like image generation, anomaly detection, and other applications where not just learning but also generation of new data is desired. Unlike AEs, which produce a single value for each encoding dimension, VAEs generate outputs in the form of probability distributions.

Chang et al. [39] suggest a method that employs a convolutional AE to individually encapsulate the spatiotemporal reconstruction data. This method divides the reconstruction process into spatial and temporal components. The spatial part rebuilds the most recent single frame, while the temporal segment accepts a series of frames as input. This process results in an optical flow, generated by the RGB disparity between the input and output. Consequently, unusual activities, characterized by irregular appearance and movement, trigger significant reconstruction errors. To ensure a more concise data representation, the gap between the data depiction and the cluster centers is minimized using two deep k-means clusters.

Tien and Huang [186] present a method based on supervised learning and AE for device type identification and detect anomalies in Internet of Things (IoT) devices. Anomalies are spotted within the packets emitted by the devices, and the identification of the devices is achieved via supervised learning on the collected packets. Following this, unsupervised learning techniques, including SVM, isolation forest, and AEs, are employed for reducing dimensionality. Nevertheless, the suggested method is specifically tailored for particular device types, indicating that additional studies are required to assess its efficiency across different device types and malicious behaviors in various IoT-oriented scenarios.

Aamir et al. [2] present an image classification approach that relies on feature and abstract representation. They build an enhanced feed-forward layered structure grounded on Convolutional Autoencoders (CAEs). These CAEs are arranged in layers, wherein encoding and decoding operations are conducted. The reconstruction error is progressively reduced by each CAE layer, ultimately leading to the identification of informative features. However, the proposed scheme heavily relies on features; feeding the system with irrelevant features could impact classification results. Additionally, memory consumption in their scheme is tied to representation learning, which may pose a challenge if applied to big data-based applications.

Liu et al. [125] propose a two-phase learning framework based on stacked AEs (SCAE) for video classification. They establish separate stacked convolutional AEs for image, audio, and text, which are the fundamental modalities in any video. The results from these AEs are then merged and input into another Multimodal Stacked Convolutional Autoencoder (MSCAE). In this two-step model, the initial phase prioritizes maintaining semantic relationships within each modality, whereas the subsequent phase targets revealing connections between semantics across different modalities.

In their work, the authors Santos et al. [168] present a semi-supervised training methodology for deep networks that fuses the capabilities of CNNs and AEs. The essence of their model is to learn image features while concurrently harnessing both supervised and unsupervised learning methods. It proves effective in circumstances where data is partially labeled. They test the effectiveness of 2D CNNs using different surveillance videos as their case study. Nevertheless, the model's analysis is confined to immediate appearances of features obtained from separate frames, omitting any form of temporal tracking. Furthermore, the approach is resource-intensive; it might not be the best fit in scenarios with restricted processing capabilities, even though it offers potential for anomaly detection.

Pawar and Attar [147] propose a deep learning model that combines CNN-AE, LSTM, and RBF networks, called CNN-SEQ2SEQ-RBF. This approach employs a 2-dimensional CNN-AE model for feature learning, while a sequence-to-sequence LSTM-based network identifies temporal statistical correlations. Lastly, an RBF network is utilized for one-class classification. However, their training process only involves normal data, with both normal and anomalous data used during the testing phase, which raises questions about the validity of their results.

#### 1) APPLICATIONS

AEs [49], [76], [110] have a wide range of applications, including image generation, data compression, denoising, anomaly detection, image inpainting, recommender systems, drug discovery, and text generation. Their ability to learn a lower-dimensional latent representation of complex data makes them versatile tools for various tasks across different domains.

- Anomaly detection [110]: AEs are adept at learning normal behavior patterns and highlighting anything that deviates from this as an irregularity. This ability makes them an excellent tool in surveillance systems where the goal is to detect unusual actions, such as identifying suspicious activities or tracking the health status of IT infrastructures.
- Video compression [73]: AEs can be utilized to effectively reduce the size of video data in real-time while maintaining high quality. This technology is beneficial for optimizing storage and transmission of surveillance data, especially in systems with limited bandwidth or storage capacity.
- Background subtraction [170]: Background subtraction using AEs is a technique that leverages the power of AE neural networks to separate the static background from the dynamic foreground in surveillance videos. By training the AE on a dataset of video frames containing both

**TABLE 8.** Popular AE methods surveyed.

| Method | Anomalies | Dataset |
|---|---|---|
| VAE [57] | Object-based | UCSD Ped1 [113], UCSD Ped2 [113], Avenue [128] |
| VAE [65] | Motion-based | UCSD Ped1 [113], Subway [7] |
| VAE [209] | Crowd-based | UCSD Ped1 [113], UCSD Ped2 [113] |
| CAE [54] | Object-based | UCSD Ped1 [113], UCSD Ped2 [113], Avenue [128] |
| VAE [196] | Motion-based | MW Motorsport |
| VAE [153] | Contextual | SUN RGB-D and SUNCG |
| DAE [189] | Motion-based | UCF [184], HMDB51 and Youtube action |
| DAE [39] | Object-based | UCSD Ped2 [113], Avenue [128], ShanghaiTech [123] |
| DAE [112] | Point-based | UCSD Ped1 [113], UCSD Ped2 [113], Avenue [128], UMN [155] |
| AE [229] | Point-based | UCSD Ped1 [113], UCSD Ped2 [113], Avenue [128], Subway [7] |

background and foreground information, the network learns to encode and decode the input frames accurately. During inference, the trained AE is used to reconstruct the input frames. By comparing the reconstructed frames with the original frames, the differences can be measured. The areas where significant differences occur are considered as the foreground, while the regions with minimal changes represent the background.

- Object tracking [88]: y learning the specific features and patterns present in images or videos, AEs can help identify objects or human actions. For example, they could detect the presence of a particular individual entering a facility or pick up certain movements like a person collapsing, engaging in a fight, or leaving behind an unattended item.
- Activity recognition [49]: AEs can be used to learn features from video data for recognizing and classifying human activities or gestures. This can aid in detecting unusual behavior or identifying specific actions that may be relevant to security issues.
- Image restoration [203]: In instances where parts of an image are missing or obscured, AEs can be employed for image inpainting, which involves completing the image based on learned patterns.
- Data compression [56]: VAEs can be employed for lossy data compression. By learning a lower-dimensional latent representation of the data, VAEs effectively compress the data while preserving most of the important information. This can be particularly useful for image, audio, or video compression.
- Denoising [40]: VAEs can be used to denoise images, audio, or other types of data. By learning the underlying structure of the data, VAEs can reconstruct the original, clean data from the noisy observations, effectively filtering out the noise.
- Recommender systems [181]: VAEs can be employed in collaborative filtering for recommender systems. VAEs can predict users' preferences for items they have not interacted with, allowing for personalized recommendations.

#### 2) DRAWBACKS

Despite their numerous benefits, several drawbacks are associated with using AEs in surveillance, including:

- One of the main limitations of AEs is their limited capacity to handle high-dimensional [133] and intricate data. AEs consist of an encoder network that compresses the input data into a lower-dimensional representation, followed by a decoder network that attempts to reconstruct the original input from this compressed representation. This compression-decompression process works well for simple patterns and regularities, but it struggles to capture and represent the intricate relationships present in complex data.
- One substantial worry revolves around the possibility of achieving less-than-ideal reconstruction quality, as highlighted in [118] study. Variational Autoencoders (VAEs) are designed to optimize the log-likelihood data's lower limit, which could possibly lead to unclear or fuzzy outcomes.
- Another potential downside pertains to the difficulties encountered when trying to understand the learned representations. VAEs are capable of extracting intricate and abstract features from the input data, which might pose a challenge for humans to comprehend. This could limit the model's practical utility, especially in surveillance applications, where human interpretation plays a crucial role.
- Privacy is indeed a significant concern when using VAEs in surveillance applications [228]. The requirement for large volumes of data to train these models effectively can lead to the handling of sensitive information about individuals. If such data is not managed with the utmost care, it can result in privacy breaches and various security risks.
- False alarms are another considerable issue associated with AEs in surveillance systems [193]. In a surveillance context, this can cause operators to become overwhelmed by the volume of false alarms, potentially causing them to miss genuine threats.

- The vulnerability of AEs to adversarial attacks is an additional concern [226]. Malicious actors can take advantage of AE weaknesses by modifying input data to evade detection, which may lead to serious security breaches in surveillance systems. Ensuring the robustness of these models against such attacks is crucial for maintaining the integrity of the surveillance applications they support.
- Overfitting [180] is a common problem when using AEs. This issue arises when the model learns the training data too well, to the point where it captures not just the underlying patterns, but also the noise or random fluctuations in the data. This makes the model highly specialized to the training data, leading to high accuracy on the training set but poor performance on new, unseen data. In essence, an overfitted AE fails to generalize well, which is crucial for the model's effectiveness in real-world applications. This overfitting can be exacerbated if the model is overly complex or if there's not enough training data. Regularization techniques are often used to mitigate this issue.
- Reconstruction Error [2]: The primary disadvantage linked to traditional CAE is the increased reconstruction error encountered when encoding and decoding input features within the network. This limitation in the CAE's functioning results in its inability to delve into the finer details within the input features, causing it to overlook valuable information. Consequently, the features extracted by the CAE do not accurately represent all input features, leading to the classifier's ineffectiveness in addressing classification challenges efficiently.
- Sensitivity to Hyper-parameters [142]: AEs, like many machine learning models, have a known issue of being highly sensitive to hyperparameters. This sensitivity means that the model's performance is heavily dependent on the initial settings for various parameters, such as the learning rate, the number of hidden layers and units, and regularization parameters. Choosing inappropriate values for these hyperparameters can lead to suboptimal results, including slow or unstable training, and poor model performance. This necessitates a thorough hyperparameter tuning process, which can be computationally expensive and time-consuming. Consequently, the challenge of finding the right set of hyperparameters can make the implementation of AEs somewhat complex and resource-intensive.

### C. GENERATIVE ADVERSARIAL NETWORKS (GANS)

A Generative Model [14], [21], [207], [227] is designed to learn the joint probability distribution and utilizes Bayes Theorem to predict conditional probabilities. Generative classifiers, such as Naive Bayes, Bayesian Networks, Markov Random Fields, and Hidden Markov Models (HMM), exemplify this approach.

Generative Adversarial Networks (GANs) represent a class of neural networks capable of generating new data samples resembling a provided training dataset. GANs comprise two interacting neural networks: the generator (G) and the discriminator (D), engaged in a competitive game-theoretic setting. The generator's role is to produce counterfeit data samples, while the discriminator's goal is to differentiate between genuine and fabricated samples. A list of popular GAN techniques explored in this paper can be found in Table 9.

Liu and Li [126] present an innovative approach called Single-Objective Generative Adversarial Active Learning (SO-GAAL) for anomaly detection, based on a competitive interaction between a generator and a discriminator. The generator leverages random noise as its input, which, under the guidance of the discriminator, generates meaningful outliers that mimic actual data. Consequently, the SO-GAAL discriminator can pinpoint these outliers and create a boundary to distinguish potential anomalies from genuine data. Moreover, the GAAL framework is extended from one generator (SO-GAAL) to multiple generators with unique objectives (MO-GAAL), assisting the generator in avoiding problems related to mode-collapse. However, this method does not provide a time-saving advantage for smaller datasets and can be quite demanding in terms of resources. The resource requirements of MO-GAAL scale linearly with the data size, making it less suitable for large-scale datasets due to its significant resource demands.

Shin and Cho [175] propose a Voice Activity Detection (VAD) classifier modeling method that employs a Generative Adversarial Network (GAN) in a supervised learning manner. Their primary goal is to address the issue of insufficient labeled data through transfer learning. Given GAN's ability to generate data absent from the dataset, the model can learn from data that is not real but closely resembles actual data. This approach helps mitigate the problem of limited labeled data.

Nevertheless, the significance of labeled data, which is crucial for the VAD classifier's learning process, cannot be overlooked. Even a small amount of labeled abnormal data can greatly impact the system's overall performance. In their suggested method, the GAN discriminator must extract features from both real and fake data, which might impede effective feature extraction.

Gui and Sun [70] attempt to offer a comprehensive review of various GANs, focusing on algorithms, theory, and software packages. They provide an in-depth introduction to the motivation, mathematical formulations, and architecture of commonly used GAN algorithms. Additionally, they discuss GANs in conjunction with other machine learning algorithms, such as semi-supervised learning, transfer learning, and reinforcement learning. They compare the similarities and differences among various techniques employing GANs. The authors have also tak about various application domians of GANs along with research challenges.

**TABLE 9.** Popular GAN learning methods surveyed.

| Method | Anomalies | Dataset | AUC |
|---|---|---|---|
| GAN [176] | Object-based | UCSD Ped1 [113], UCSD Ped2 [113], Subway [7] | 90–96% |
| GAN [160] | Point-based | UCSD Ped1 [113], UCSD Ped2 [113], UMN [155] | 95.5–99% |
| GAN [53] | Object-based | UCSD Ped2 [113], Avenue [128], ShanghaiTech [123] | 70.9–97.2% |
| GAN [62] | Point-based | UCSD Ped1 [113], UCSD Ped2 [113], UMN [155] | 94.6–99% |
| GAN [10] | Behavioral | UCSD Ped1 [113], UCSD Ped2 [113], UMN [155], Hajj | 79.63–98.1% |
| GAN [18] | Object-based | UMCD [19] | 97.2–95.7% |
| GAN [224] | Object-based | UCSD Ped2 [113], ShanghaiTech [123] | 75.2–95.3% |
| GAN [219] | Object-based | MNIST | Not Given |
| GAN [42] | Motion-based | UCSD Ped1 [113], Avenue [128] | 87.87% |
| GAN [43] | Object-based | UCSD Ped1 [113], Avenue [128], ShanghaiTech [123] | 74.5–81.6% |
| 3D-CNN-GAN [176] | Motion-based | UCSD Ped1 [113] | 86.4–94.4% |

#### 1) APPLICATIONS

GANs have found extensive applications across various domains, with emerging uses in surveillance systems as well. Some specific applications of GANs in surveillance include:

- Data augmentation: GANs can generate realistic synthetic data for training surveillance systems. By creating additional training samples that mimic real-world situations, GANs can enhance the resilience and generalizability of models, especially when there is limited or imbalanced labeled data.
- Anomaly detection [33]:GANs can be utilized to learn the distribution of regular patterns within a scene. The generator is trained to produce lifelike samples, while the discriminator is trained to distinguish between real and generated samples. If the discriminator can easily classify a new input as real or generated, it is likely to represent normal behavior; otherwise, it may suggest an anomaly or unusual activity.
- Image inpainting [35]: Surveillance footage often contains missing or corrupted data due to occlusions, camera malfunctions, or transmission errors. GANs can be employed to fill in the missing or corrupted parts of the images, thereby providing more complete and accurate visual information for analysis.
- Object detection and tracking [219]: GANs can be employed to generate realistic object samples to supplement training data for object detection and tracking models. By offering more diverse and representative samples, GANs can enhance the performance of object detection and tracking algorithms in intricate and dynamic settings.
- Crowd simulation and analysis [10]: GANs can create lifelike synthetic crowd scenes for studying crowd behavior and dynamics. This can be useful for developing and evaluating surveillance algorithms within crowded environments, such as public transportation hubs, shopping centers, or sporting venues.
- Privacy preservation [33]: GANs can be employed to anonymize sensitive information in surveillance footage, like faces or license plates, while maintaining the overall scene structure. This allows for the protection of individual privacy while still enabling effective monitoring and analysis of video data.

#### 2) DRAWBACKS

GANs have become popular in recent years due to their capability to create realistic synthetic data. However, they also exhibit some drawbacks and limitations, such as:

- Training Difficulty: GANs are often challenging to train due to the intricate balance required between the generator and the discriminator. Discrepancies between the two networks can lead to issues such as mode collapse, wherein the generator outputs a limited range of variants.
- Mode Collapse: This situation arises when the generator starts yielding a restricted set of outputs, neglecting the broader diversity of the data. Essentially, the generator collapses to generate only a few modes of the data.
- Lack of Explicit Control: GANs generally do not offer an explicit control mechanism over the types of outputs they produce. Despite certain possibilities to influence the generation process, there's no direct way to dictate the specifics of the desired network output.
- Evaluation Difficulty: Evaluating the performance of GANs is not straightforward. Conventional metrics used for other machine learning models often fail to apply to GANs, and the subjective nature of assessing generated content (like images or music) further complicates the evaluation.
- Resource-Intensive: GANs demand substantial computational resources for their training, and they typically need to process vast volumes of data to deliver high-quality outputs. This can render them less efficient for certain applications.

- Instability: The training procedure of GANs can exhibit instability due to the adversarial relationship between the two networks. Without careful supervision, this could lead to subpar performance and outcomes.

### D. CONVOLUTIONAL NEURAL NETWORK (CNN)

A Convolutional Neural Network (CNN or ConvNet) [127], [177] are a category of artificial neural networks specifically engineered to process grid-structured data, making them especially suitable for image data interpretation. CNNs excel in identifying hierarchical patterns of spatial features within images due to their automatic and adaptive learning capabilities. These networks are characterized by three distinctive layers: convolutional, pooling, and fully connected layers. The convolutional layer applies different filters to the input, the pooling layer decreases the data's spatial dimensions to simplify processing, and the fully connected layers use the previously extracted features for classification tasks. The inception and advancement of CNNs have significantly boosted the growth of deep learning algorithms, leading to substantial progress in fields like image and video recognition, recommendation systems, and natural language processing.

A defining characteristic of CNNs is their resistance to shifts or changes in the position of features in the input data, an aspect often referred to as Shift Invariant/Area Invariant Artificial Neural Networks (SIANNs). This attribute primarily stems from the use of shared-weight convolution kernels, or filters, which move across the input data, generating feature maps that are equivariant to translation. The application of CNNs spans a broad range of domains, encompassing image and video recognition systems, recommendation systems, classification tasks, image segmentation and analysis, natural language processing, and financial time series analysis. The specific CNN approaches discussed in this document are itemized in Table 10.

In their work, the authors referred to as Wu and Wang [205], proposed the utilization of CNN features to identify distinct visual objects and components for scene categorization. The methodology involved a region-focused technique to generate potentially significant patches containing objects of interest. They subsequently implemented unsupervised and semi-supervised learning methodologies on a pre-trained CNN. As a final step, they used a clustering approach and a localized CNN-oriented method to assemble alike objects into collections, termed as meta-groups. However, the authors overlooked the possibility that a CNN may not be the best tool for feature extraction. Furthermore, applying a fixed CNN to a consistent data scale could introduce data bias, given the dynamic nature of data in scale and features, while the feature extractor remains unaltered.

The authors referenced as Lan et al. [108] introduced a CNN-based procedure for visual surveillance, which is particularly effective in resource-constrained scenarios. The approach comprises three steps. Initially, shot segmentation takes place via deep feature memorability and entropy score prediction, paired with the creation of a summary. This process smartly splits the video into meaningful shots, a critical task in video summarization. Next, each frame's memorability within the shot is determined using a refined image prediction system that leverages entropy measures. Ultimately, the frame with the top memorability and entropy score in each shot is selected for the final determination. Despite the promise, a notable shortcoming of this proposed method is the selection of keyframes based on memorability score, which may result in a summary that inadequately encapsulates the entire video's content. Furthermore, memorability doesn't ensure a diverse range of keyframes. Yet, in spite of these restrictions, CNNs have demonstrated remarkable performance in various computer vision undertakings, rendering them apt for surveillance applications.

#### 1) APPLICATIONS

CNNs have a wide range of applications in enhancing security, safeguarding, and monitoring efficacy within surveillance systems. Given their proficiency in managing diverse computer vision responsibilities, they offer considerable potential in amplifying surveillance capabilities. The following are a few instances demonstrating the application of CNNs in surveillance systems specifically for detecting irregularities:

- Motion detection [102]: Motion detection using CNNs has become increasingly effective and prevalent in the field of surveillance. CNNs excel in detecting and identifying motion patterns by analyzing sequential frames within a video. By training on a multitude of diverse movement scenarios, these networks learn to distinguish between regular and irregular motion, allowing for more accurate anomaly detection. Furthermore, the spatial and temporal features extracted by CNNs contribute to a more comprehensive understanding of the scene dynamics, thus improving the system's capacity to detect motion and changes in real-time surveillance footage.
- Object recognition: CNNs have become a pivotal tool in the field of object recognition. By automatically learning hierarchical patterns of spatial features from input images, CNNs excel at recognizing a wide array of objects in varying contexts and viewpoints. They do this by processing an image through multiple layers of filters or 'convolutions', which identify different aspects of the object, such as edges, textures, colors, and shapes.
- Crowd analysis [217]: CNNs have demonstrated remarkable capabilities in understanding and analyzing crowd behavior. By processing visual data from scenes with dense crowds, CNNs can learn to identify specific patterns and trends that characterize various behaviors. They can be used to estimate crowd density, providing valuable data for crowd management in public events

**TABLE 10.** Popular of CNN learning methods surveyed.

| Method | Anomalies | Dataset | AUC |
|---|---|---|---|
| 3D-CNN [187] | Behavioral | Hockey Fight | 91% |
| 3D-CNN [187] | Temporal | UCF [184] | 87.01% |
| 3D-CNN [135] | Object-based | CAVIAR, UMN [155] | 97% |
| CNN [74] | Point-based | eBot-IoT dataset | 97.91–98.94% |
| CNN-LSTM [74] | Object-based | online sources | 82% |
| CNN [217] | Behavioral | UCF [184], HMDB51 | 96.3% |
| CNN-RNN [94] | Object-based | NAHFE | 89.5% |
| CNN-GAN [91] | Object-based | Avenue [128], ShanghaiTech [123] | 68.94–88.26% |
| CNN-ConvLSTM [197] | Motion-based | UCF [184] | 97.64% |
| CNN-YOLOv5 [5] | Behavioral | Data | 87.4–95.0% |
| CNN-YOLOv5[154] | Object-based | UCF [184] | 88.92–91.24% |

or places. Furthermore, by studying the flow and direction of crowd movement, CNNs can predict future behavior, which can be particularly useful for planning and control in urban environments.

- Anomaly Identification [74]: CNNs have the capability to spot irregular or dubious actions by studying and contrasting typical behavior patterns captured in surveillance videos. This can serve as an effective tool in averting potential security intrusions, acts of vandalism, or any unwanted activities.
- Object Tracking [74]: With the power to identify and monitor objects or individuals' movements in real-time, CNNs can be employed to maintain surveillance across different camera perspectives or over a period of time. This proves beneficial for maintaining security, aiding law enforcement, or for general monitoring objectives.
- Activity Recognition [94]: CNNs can pinpoint certain actions or activities executed by individuals in surveillance videos, such as sprinting, engaging in a conflict, or transporting objects. This lends additional insight for security surveillance and aids informed decision-making.

#### 2) DRAWBACKS

CNNs are a valuable technique for various tasks, but it is crucial to consider the following potential drawbacks when determining their suitability for anomaly detection in surveillance applications:

- Data reliance: CNNs need a substantial quantity of annotated data to discern valuable characteristics. With unsupervised learning, data is often unlabelled, which could reduce the efficiency of CNNs for these tasks compared to supervised learning assignments.
- Restricted understandability: While CNNs are capable of extracting intricate attributes from data, the comprehending of these features can be complex. This lack of transparency can hinder understanding of how the network reaches decisions or troubleshooting when performance falls short.
- High computational requirements: Training CNNs can be resource-intensive, particularly with large datasets. This can pose challenges when trying to upscale unsupervised learning tasks that involve vast datasets.
- Risk of overfitting: CNNs, like all machine learning models, can overfit if the model's complexity exceeds the scope of the training data. Overfitting can be a major issue in unsupervised learning, where the training data is usually limited or contains noise.
- Limited extrapolation: CNNs are adept at identifying specific patterns or features in data but might fall short in applying these patterns to novel or unseen data. This could limit their value in unsupervised learning assignments aiming to unearth general patterns or structure in the data.

### E. SUMMARY OF UNSUPERVISED LEARNING METHODS

The limitations and strengths of unsupervised learning methods are listed in Table 11. The table presents a summary of several unsupervised methods, detailing their applications and potential drawbacks. PCA focuses on reducing data dimensionality by projecting it onto a subspace that highlights the primary variations. Notably, it's utilized to detect anomalies based on reconstruction error. However, challenges with PCA include loss of interpretability, the necessity for data standardization, potential information loss, dependence on precise parameter tuning, and a limited capacity for interpretation in surveillance contexts. AEs stand out in detecting anomalies by training on typical, non-anomalous data to discern underlying structures. But they face hurdles, notably in reconstructing complex or noisy input signals, which could result in inaccuracies in anomaly detection. Furthermore, AEs demand vast training data and computational resources. GANs, though not detailed in the findings, are marred by several issues, such as training difficulties, the well-documented mode collapse, a lack of direct control, challenges in evaluation, their resource-intensive nature, and instability stemming from the adversarial interplay of their dual networks. CNNs, employed

**TABLE 11.** Summary of un-supervised methods surveyed.

| Method | Findings | Limitations |
|---|---|---|
| **PCA** | PCA is a technique that reduces the dimensionality of data by projecting it to a lower-dimensional subspace that captures the main variation. It is used to detect anomalies by measuring the reconstruction error | • Loss of Interpretability in independent variables<br>• Standardization of data<br>• Information Loss<br>• Dependence on parameter tuning<br>• Limited interpretability in surveillance contexts.<br>. |
| **AEs** | AEs is to identify unusual or abnormal patterns within data. In the context of anomaly detection, AEs are trained on normal, non-anomalous data to learn the underlying patterns and structures. | • Not able to reconstruct complex or noisy input signals accurately, leading to false positives or false negatives in anomaly detection.<br>• Requires a large amount of training data and computational resources to learn the features of the input signals |
| **GANS** | Logistic regression is useful for predicting binary outcomes based on input features. It finds applications in fraud detection, security surveillance, and traffic monitoring. | • Training Difficulty<br>• Mode Collapse Issue<br>• Lack of Explicit Control<br>• Evaluation Difficulty<br>• Resource-Intensive<br>• Instability due to the adversarial relationship between the two networks. |
| **CNN** | CNNs for anomaly detection in surveillance recognize complex patterns and features within images or video data making them well-suited for tasks like object recognition, image classification, and anomaly detection in surveillance scenarios. | • Data reliance.<br>• Restricted understandability.<br>• Risk of overfitting.<br>• Limited extrapolation. |

primarily for anomaly detection in surveillance, are adept at discerning intricate patterns in images or videos, making them apt for tasks like object recognition. Nevertheless, their efficiency is tempered by a heavy reliance on data, limited understandability, a propensity for overfitting, and constraints in extrapolation.

## V. CHALLENGES, FUTURE DIRECTIONS, AND CONCLUSION

This survey paper offers a comprehensive and critical analysis of various learning methods for anomaly detection, covering supervised, semi-supervised, and unsupervised approaches. We review a wide range of techniques, such as discriminative modeling, linear and logistic regression, SVM, RNN, LSTM, MIL, one-class learning, transfer learning, deep neural networks, CNN, GANs, AEs, CAE, CNN-AE, and VAE. We examine their strengths, weaknesses, and applicability in different domains and scenarios of anomaly detection. Finally, we identify the challenges and opportunities for future research in this field, such as improving the accuracy, robustness, and efficiency of anomaly detection methods, handling complex and dynamic situations, such as crowded scenes, occlusions, and varying illumination, and evaluating and comparing the performance of different methods on diverse datasets and domains.

### A. CHALLENGES

During our research, we observed following key points which makes detection of anomalies in surveillance videos challenging:

- Shortage of real-world data: There is a significant demand for collecting real-world data to develop efficient algorithms and create computer vision applications that excel in real-life scenarios.
- Lighting conditions: Handling varying lighting conditions is challenging, as extracting trained features from videos becomes difficult.
- Camera angles and perspective: The camera angles defining the surveillance area greatly influence the

performance of deep learning algorithms, as the appearance of objects or people may change based on their distance from the camera.
- Diverse objects: Learning the movements of various objects and entities within a scene can sometimes be problematic, leading to reduced application performance due to differences in appearance.
- Sparse versus Dense environments: The techniques employed to detect anomalies in sparse and dense settings differ. Some methods work well for event recognition in sparse contexts but may produce numerous false negatives in densely populated scenes, such as large crowds.
- Occlusions: Detecting and tracking occluded instances (people or objects) that are partially or fully obscured is a complex task, even though humans can handle it relatively easily.
- One significant limitation of linear regression in surveillance applications is the assumption that predictor variables and the outcome variable have a linear relationship. Secondly, high-dimensional data, characterized by a large number of features, presents challenges for linear regression models in surveillance. Problems include overfitting, where the model becomes too complex and fails to generalize to unseen data; multicollinearity, leading to unstable parameter estimates and poor predictive performance; the curse of dimensionality, causing data sparsity and reduced learning efficiency; and increased computational complexity. To mitigate these issues, various techniques can be utilized. Feature selection methods, like LASSO or Ridge regression, can reduce the number of features.
- A notable disadvantage of using logistic regression in surveillance applications is its design specifically for binary classification problems, meaning it can only predict the probability of an event belonging to one of two classes. In many surveillance scenarios, predicting multiple classes or events is necessary, making logistic regression less suitable.
- A major drawback of using SVM)in surveillance applications is their computational complexity, particularly for large datasets. Training an SVM involves solving a quadratic optimization problem, which can be computationally demanding and time-consuming, especially when dealing with high-dimensional data or numerous instances. This may limit the applicability of SVMs in real-time or near-real-time surveillance systems.
- One significant limitation of MIL is the assumption that if a bag is labeled as positive, there's at least one positive instance within it. In real-world scenarios, particularly in anomaly detection, this may not always hold true. For instance, an anomaly might be manifested only when considering the combination of multiple instances within the bag, rather than a single instance. Additionally, the performance of MIL can be significantly affected by the choice of bag representation and instance classifier, which may not always be straightforward to select. The model could also struggle with the concept drift, which is a common issue in anomaly detection where the nature of normal and anomalous instances changes over time. Lastly, the training process for MIL can be more complex and computationally intensive compared to standard supervised learning approaches, potentially limiting its applicability in scenarios where computational resources or time are constrained.
- AEs in surveillance face significant challenges, such as susceptibility to environmental changes leading to inconsistent performance. They also demand substantial computational resources and time, making them less suitable for real-time scenarios. The requirement for large datasets for training and their inherent design of reconstructing input data, potentially carrying forward irrelevant information, further hinder their effective utilization in surveillance contexts.
- Data dependencies in CNNs can lead to challenges such as lengthy training times, issues of vanishing and exploding gradients, increased memory usage, limited parallelization during training, and risk of overfitting. They may also necessitate high-performance computing infrastructure, potentially limiting accessibility for some users. However, these issues can be mitigated through various techniques, and ongoing research continues to improve CNN performance and scalability.
- A major limitation of using Principal Component Analysis (PCA) in surveillance applications is its linearity as a dimensionality reduction technique. This means PCA assumes the data's underlying structure is linear, which may not be accurate for many complex surveillance scenarios where relationships between variables are nonlinear. In such cases, PCA may not effectively capture the underlying structure and patterns, leading to decreased performance in subsequent analysis or classification tasks.
- It's important to note that while one-class classification can be a powerful tool in surveillance, it is not without its challenges. The success of this approach relies heavily on the representativeness and quality of the 'normal' training data. If this data doesn't adequately capture the range of normal behaviors, the classifier may either miss genuine anomalies (false negatives) or incorrectly flag normal behavior as anomalous (false positives). Therefore, careful data collection and model validation are essential steps in implementing a one-class classification system for surveillance.
- Surveillance data is often unbalanced, consisting of a large number of regular events and a small number of unusual events. This imbalance can make it challenging for machine learning algorithms to effectively detect anomalies.
- Supervised machine learning models require labeled data for training. However, annotating surveillance data can be labor-intensive and costly. This limitation can

restrict the availability of labeled data, making it difficult to efficiently train machine learning models.
- Surveillance data can be affected by noise, incompleteness, or corruption. These issues can impact the performance of machine learning algorithms, as they might learn from inaccurate data.
- Deep learning models can be difficult to interpret, and explaining their specific decisions can be challenging. This lack of interpretability can be problematic in surveillance systems, where understanding why a particular event was identified as anomalous may be necessary.
- Deep learning models can require considerable computational resources for training and deployment. This requirement can pose a challenge in real-time surveillance systems, where prompt detection and response are crucial.
- Surveillance data may evolve over time due to changes in the environment, individuals, and objects within the scene. These changes can cause the data distribution to shift over time, resulting in concept drift. Machine learning algorithms trained on historical data might struggle to adapt to such changes.
- Surveillance data can be extensive and complex, with high-dimensional features. This complexity can lead to increased computational costs for training and deploying machine learning models, particularly in real-time situations.
- Anomaly detection in surveillance raises ethical issues surrounding privacy, monitoring, and bias. Machine learning algorithms trained on biased or prejudiced data can reinforce and even magnify these biases, potentially resulting in unfair or discriminatory consequences.

### B. FUTURE DIRECTIONS

In future this survey can be enhanced by surveying critical analysis of hybrid methods that combine different learning techniques, such as supervised and unsupervised, or sem-supervised methods to leverage their advantages and overcome their drawbacks. Furthermore, it is worthwhile to explore the use of multimodal data, such as audio, video, and text, to capture more information and context for anomaly detection. For example, such methods could use audio data to detect abnormal sounds or events, video data to detect abnormal motions or behaviors, and text data to detect abnormal sentiments or topics. Finally, there is a pressing need of a detailed and in depth survey and analysis of available surveillance datasets highlighting their strengths and limitations for both commercial and academic research.

## REFERENCES

[1] S. A. Aa, S. Tiwari, V. Tiwari, and S. Yadav, "Voice assistant for physically challenged individuals: Enhancing accessibility and independence," *Int. Res. J. Modernization Eng. Technol. Sci.*, vol. 8, no. 1, pp. 27–31, May 2023.

[2] M. Aamir, N. M. Nawi, F. Wahid, and H. Mahdin, "A deep contractive autoencoder for solving multiclass classification problems," *Evol. Intell.*, vol. 14, no. 4, pp. 1619–1633, 2020.

[3] J. Abawajy, S. Huda, S. Sharmeen, M. M. Hassan, and A. Almogren, "Identifying cyber threats to mobile-IoT applications in edge computing paradigm," *Future Gener. Comput. Syst.*, vol. 89, pp. 525–538, Dec. 2018.

[4] K. Abbas, M. K. Hasan, A. Abbasi, U. A. Mokhtar, A. Khan, S. N. H. Abdullah, S. Dong, S. Islam, D. Alboaneen, and F. R. A. Ahmed, "Predicting the future popularity of academic publications using deep learning by considering it as temporal citation networks," *IEEE Access*, vol. 11, pp. 83052–83068, 2023.

[5] M. Abduljabbar Ali, A. Jaafar Hussain, and A. T. Sadiq, "Deep learning algorithms for human fighting action recognition," *Int. J. Online Biomed. Eng. (iJOE)*, vol. 18, no. 2, pp. 71–87, Feb. 2022.

[6] O. Abualghanam, H. Alazzam, E. Alhenawi, M. Qatawneh, and O. Adwan, "Fusion-based anomaly detection system using modified isolation forest for Internet of Things," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 1, pp. 131–145, Jan. 2023.

[7] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 555–560, Mar. 2008.

[8] aditianu1998. (Jan. 2021). *Understanding of LSTM Networks*. Accessed: Mar. 14, 2023. [Online]. Available: https://www.geeksforgeeks.org/understanding-of-lstmnetworks/

[9] A. B. Ahmad and T. Tsuji, "Traffic monitoring system based on deep learning and seismometer data," *Appl. Sci.*, vol. 11, no. 10, p. 4590, May 2021.

[10] T. Alafif, B. Alzahrani, Y. Cao, R. Alotaibi, A. Barnawi, and M. Chen, "Generative adversarial network based abnormal behavior detection in massive crowd videos: A Hajj case study," *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 8, pp. 4077–4088, Aug. 2022.

[11] A. Aldayri and W. Albattah, "Taxonomy of anomaly detection techniques in crowd scenes," *Sensors*, vol. 22, no. 16, p. 6080, Aug. 2022.

[12] F. Alrowais, S. S. Alotaibi, F. N. Al-Wesabi, N. Negm, R. Alabdan, R. Marzouk, A. S. Mehanna, and M. Al Duhayyim, "Deep transfer learning enabled intelligent object detection for crowd density analysis on video surveillance systems," *Appl. Sci.*, vol. 12, no. 13, p. 6665, Jun. 2022.

[13] S. Anoopa and A. Salim, "Survey on anomaly detection in surveillance videos," *Mater. Today, Proc.*, vol. 58, pp. 162–167, Jan. 2022.

[14] M. Aqqa and S. K. Shah, "CAR-DCGAN: A deep convolutional generative adversarial network for compression artifact removal in video surveillance systems," in *Proc. VISIGRAPP*, 2021, pp. 455–464.

[15] D. Arivudainambi, K. A. Varun, S. Chakkaravarthy, and P. Visu, "Malware traffic classification using principal component analysis and artificial neural network for extreme surveillance," *Comput. Commun.*, vol. 147, pp. 50–57, Nov. 2019.

[16] J. R. Arunkumar, S. Velmurugan, B. Chinnaiah, G. Charulatha, M. R. Prabhu, and A. P. Chakkaravarthy, "Logistic regression with elliptical curve cryptography to establish secure IoT," *Comput. Syst. Sci. Eng.*, vol. 45, no. 3, pp. 2635–2645, 2023.

[17] INGA Astawa and Gusti Caturbawa, "Detection of license plate using sliding window, histogram of oriented gradient, and support vector machines method," *J. Phys., Conf.*, vol. 953, Jan. 2018, Art. no. 012062.

[18] D. Avola, I. Cannistraci, M. Cascio, L. Cinque, A. Diko, A. Fagioli, G. L. Foresti, R. Lanzino, M. Mancini, A. Mecca, and D. Pannone, "A novel GAN-based anomaly detection and localization method for aerial video surveillance at low altitude," *Remote Sens.*, vol. 14, no. 16, p. 4110, Aug. 2022.

[19] D. Avola, L. Cinque, G. L. Foresti, N. Martinel, D. Pannone, and C. Piciarelli, "A UAV video dataset for mosaicking and change detection from low-altitude flights," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 6, pp. 2139–2149, Jun. 2020.

[20] D. Avola, L. Cinque, A. Di Mambro, A. Diko, A. Fagioli, G. L. Foresti, M. R. Marini, A. Mecca, and D. Pannone, "Low-altitude aerial video surveillance via one-class SVM anomaly detection from textural features in UAV images," *Information*, vol. 13, no. 1, p. 2, Dec. 2021.

[21] M. C. Bakkay, H. A. Rashwan, H. Salmane, L. Khoudour, D. Puig, and Y. Ruichek, "BSCGAN: Deep background subtraction with conditional generative adversarial networks," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 4018–4022.

[22] S. Bansod and A. Nandedkar, "Transfer learning for video anomaly detection," *J. Intell. Fuzzy Syst.*, vol. 36, no. 3, pp. 1967–1975, Mar. 2019.

[23] A. M. Bartkowiak, "Anomaly, novelty, one-class classification: A comprehensive introduction," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 3, no. 1, pp. 61–71, 2011.

[24] M. Behniafar, A. Nowroozi, and H. R. Shahriari, "A survey of anomaly detection approaches in Internet of Things," *ISeCure*, vol. 10, no. 2, pp. 79–92, 2018.

[25] M. Bendali-Braham, J. Weber, G. Forestier, L. Idoumghar, and P.-A. Müller, "Transfer learning for the classification of video-recorded crowd movements," in *Proc. 11th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2019, pp. 271–276.

[26] P. Bhandari. (Jul. 2021). *Correlation vs. Causation, Difference, Designs and Examples*. Accessed: Mar. 14, 2023. [Online]. Available: https://www.scribbr.com/methodology/correlation-vs-causation/

[27] A. Biswal. (Mar. 2023). *Principal Component Analysis in Machine Learning: Complete Guide*. Accessed: Mar. 15, 2023. [Online]. Available: https://www.simplilearn.com/tutorials/machine-learning-tutorial/principal-component-analysis

[28] A. Biswal. (Jan. 2023). *Recurrent Neural Networks (RNN) Tutorial*. Accessed: Mar. 17, 2023. [Online]. Available: https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn

[29] S. Blunsden and R. Fisher, "The BEHAVE video dataset: Ground truthed video for multi-person behavior classification," *Ann. BMVA*, vol. 4, p. 4, May 2010.

[30] C. Brax, "Anomaly detection in the surveillance domain," Ph.D. thesis, Orebro Univ., Orebro, Sweden, 2011.

[31] J. Brownlee. (Feb. 2020). *One-Class Classification Algorithms for Imbalanced Datasets*. Accessed: Mar. 27, 2023. [Online]. Available: https://machinelearningmastery.com/one-class-classification-algorithms/

[32] A. Butt, S. Narejo, M. R. Anjum, M. U. Yonus, M. Memon, and A. A. Samejo, "Fall detection using LSTM and transfer learning," *Wireless Pers. Commun.*, vol. 126, no. 2, pp. 1733–1750, Sep. 2022.

[33] Z. Cai and Z. Xiong, "Generative adversarial networks," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–38, Jul. 2021.

[34] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognit.*, vol. 77, pp. 329–353, May 2018.

[35] D. Cha and D. Kim, "DAM-GAN: Image inpainting using dynamic attention map based on fake texture detection," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process. (ICASSP)*, May 2022, pp. 4883–4887.

[36] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," 2019, *arXiv:1901.03407*.

[37] H. Chandel and S. Vatta, "Occlusion detection and handling: A review," *Int. J. Comput. Appl.*, vol. 120, no. 10, pp. 33–38, Jun. 2015.

[38] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, pp. 1–58, Jul. 2009.

[39] Y. Chang and Z. Tu, "Clustering driven deep autoencoder for video anomaly detection," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 329–345.

[40] M. F. C. Aminudin and S. A. Suandi, "Video surveillance image enhancement via a convolutional neural network and stacked denoising autoencoder," *Neural Comput. Appl.*, vol. 34, no. 4, pp. 3079–3095, Oct. 2021.

[41] A. Chen, Y. Fu, X. Zheng, and G. Lu, "An efficient network behavior anomaly detection using a hybrid DBN-LSTM network," *Comput. Secur.*, vol. 114, Mar. 2022, Art. no. 102600.

[42] D. Chen, P. Wang, L. Yue, Y. Zhang, and T. Jia, "Anomaly detection in surveillance video based on bidirectional prediction," *Image Vis. Comput.*, vol. 98, Jun. 2020, Art. no. 103915.

[43] D. Chen, L. Yue, X. Chang, M. Xu, and T. Jia, "NM-GAN: Noise-modulated generative adversarial network for video anomaly detection," *Pattern Recognit.*, vol. 116, Aug. 2021, Art. no. 107969.

[44] Z. Chen, B. Gao, H. Zhang, Z. Zhao, H. Liu, and D. Cai, "User personalized satisfaction prediction via multiple instance deep learning," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 907–915.

[45] Y. Cheng and G. Yue, "Application of convolutional neural network technology in vehicle parking management," in *Proc. 5th Int. Conf. Comput. Sci. Appl. Eng.*, Oct. 2021, pp. 1–6.

[46] Y. Ming, H. Qian, and L. Guangyuan, "CNN-LSTM facial expression recognition method fused with two-layer attention mechanism," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, Oct. 2022.

[47] F. Colangelo, F. Battisti, M. Carli, A. Neri, and F. Calabró, "Enhancing audio surveillance with hierarchical recurrent neural networks," in *Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2017, pp. 1–6.

[48] E. Cruz-Esquivel and Z. J. Guzman-Zavaleta, "An examination on autoencoder designs for anomaly detection in video surveillance," *IEEE Access*, vol. 10, pp. 6208–6217, 2022.

[49] K. Datchanamoorthy and B. Padmavathi, "Anomaly and activity recognition in a video surveillance using masked autoencoder," in *Proc. Int. Conf. Innov. Comput., Intell. Commun. Smart Electr. Syst. (ICSES)*, Jul. 2022, pp. 1–7.

[50] S. Deep, X. Zheng, C. Karmakar, D. Yu, L. G. C. Hamey, and J. Jin, "A survey on anomalous behavior detection for elderly care using dense-sensing networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 352–370, 1st Quart., 2020.

[51] D. Rajaram and K. Sivakumar, "Moving objects detection, classification and tracking of video streaming by improved feature extraction approach using k-SVM," *DYNA*, vol. 97, no. 3, pp. 274–280, May 2022.

[52] Cem Dilmegani. (Jan. 2023). *Transfer Learning in 2023: What it is; How it Works*. [Online]. Available: https://research.aimultiple.com/transfer-learning/

[53] K. Doshi and Y. Yilmaz, "Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate," *Pattern Recognit.*, vol. 114, Jun. 2021, Art. no. 107865.

[54] E. Duman and O. A. Erdem, "Anomaly detection in videos using optical flow and convolutional autoencoder," *IEEE Access*, vol. 7, pp. 183914–183923, 2019.

[55] S. Dutta. (Jan. 2021). *Predictive Maintainance Using LSTM - Application Based on IIoT*. Accessed: Mar. 29, 2023. [Online]. Available: https://medium.com/ai-techsystems/

[56] A. El Sayed, M. Ruiz, H. Harb, and L. Velasco, "Deep learning-based adaptive compression and anomaly detection for smart B5G use cases operation," *Sensors*, vol. 23, no. 2, p. 1043, Jan. 2023.

[57] Y. Fan, G. Wen, D. Li, S. Qiu, M. D. Levine, and F. Xiao, "Video anomaly detection and localization via Gaussian mixture fully convolutional variational autoencoder," *Comput. Vis. Image Understand.*, vol. 195, Jun. 2020, Art. no. 102920.

[58] P. Filzmoser and K. Nordhausen, "Robust linear regression for high-dimensional data: An overview," *WIREs Comput. Statist.*, vol. 13, no. 4, p. e1524, Jul. 2021.

[59] R. Foorthuis, "On the nature and types of anomalies: A review of deviations in data," *Int. J. Data Sci. Analytics*, vol. 12, no. 4, pp. 297–331, Aug. 2021.

[60] H. Frihia and H. Bahi, "One-class training for intrusion detection," in *Proc. 1st Int. Conf. Intell. Syst. Pattern Recognit.*, Oct. 2020, pp. 12–16.

[61] M. Fritz and P. D. Berger, "Chapter 11—Will anybody buy? Logistic regression," in *Improving the User Experience Through Practical Data Analytics*, M. F. Paul and D. Berger, Eds. Boston, MA, USA: Morgan Kaufmann, 2015, pp. 271–304.

[62] T. Ganokratanaa, S. Aramvith, and N. Sebe, "Anomaly event detection using generative adversarial network for surveillance videos," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Nov. 2019, pp. 1395–1399.

[63] C. Gautam, P. K. Mishra, A. Tiwari, B. Richhariya, H. M. Pandey, S. Wang, and M. Tanveer, "Minimum variance-embedded deep kernel regularized least squares method for one-class classification and its applications to biomedical data," *Neural Netw.*, vol. 123, pp. 191–216, Mar. 2020.

[64] S. Gautam, A. Henry, M. Zuhair, M. Rashid, A. R. Javed, and P. K. R. Maddikunta, "A composite approach of intrusion detection systems: Hybrid RNN and correlation-based feature optimization," *Electronics*, vol. 11, no. 21, p. 3529, Oct. 2022.

[65] M. George, B. R. Jose, and J. Mathew, "Abnormal activity detection using shear transformed spatio-temporal regions at the surveillance network edge," *Multimedia Tools Appl.*, vol. 79, nos. 37–38, pp. 27511–27532, Oct. 2020.

[66] R. Gholami and N. Fakhari, "Support vector machine: Principles, parameters, and applications," in *Handbook of Neural Computation*, P. Samui, S. Sekhar, and V. E. Balas, Eds. Cambridg, MA, USA: Academic Press, 2017, pp. 515–535.

[67] A. Girma, X. Yan, and A. Homaifar, "Driver identification based on vehicle telematics data using lstm-recurrent neural network," 2019, *arXiv:1911.08030*.

[68] A. Goel. (Jan. 2023). *Support Vector Machine in Machine Learning*. Accessed: Jan. 19, 2023. [Online]. Available: https://www.geeksforgeeks.org/support-vector-machine-in-machine-learning/
[69] D. Grahn, "mil-benchmarks: Standardized evaluation of deep multiple-instance learning techniques," 2021, *arXiv:2105.01443*.
[70] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," 2020, *arXiv:2001.06937*.
[71] H. Guo, J. Liu, Z. Xiao, and L. Xiao, "Deep CNN-based hyperspectral image classification using discriminative multiple spatial–spectral feature fusion," *Remote Sens. Lett.*, vol. 11, no. 9, pp. 827–836, Sep. 2020.
[72] C. Guyon, T. Bouwmans, and E.-H. Zahzah, "Foreground detection via robust low rank matrix decomposition including spatio-temporal constraint," in *Proc. Int. Conf. Pattern Recognit.*, vol. 7728, 2012, pp. 315–320.
[73] A. Habibian, T. V. Rozendaal, J. Tomczak, and T. Cohen, "Video compression with rate-distortion autoencoders," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7032–7041.
[74] B. I. Hairab, M. S. Elsayed, A. D. Jurcut, and M. A. Azer, "Anomaly detection based on CNN and regularization techniques against zero-day attacks in IoT networks," *IEEE Access*, vol. 10, pp. 98427–98440, 2022.
[75] N. L. Hakim, T. K. Shih, S. P. K. Arachchi, W. Aditya, Y.-C. Chen, and C.-Y. Lin, "Dynamic hand gesture recognition using 3DCNN and LSTM with FSM context-aware model," *Sensors*, vol. 19, no. 24, p. 5429, Dec. 2019.
[76] R. Hammouche, A. Attia, S. Akhrouf, and Z. Akhtar, "Gabor filter bank with deep autoencoder based face recognition system," *Exp. Syst. Appl.*, vol. 197, Jul. 2022, Art. no. 116743.
[77] J. Hanlon. (Mar. 2017). *How to Solve the Memory Challenges of Deep Neural Networks*. Accessed: Mar. 29, 2023. [Online]. Available: https://www.topbots.com/
[78] P. Hemanth, S. Behera, R. J. Ramudu, P. J. Krishna, N. V. S. Shankar, D. K. S. Naik, and G. Bhiogade, "Fabrication of quadcopter with face recognition for surveillance," in *Proc. Int. Conf. Ind. Eng. Operations Manag.*, Aug. 2022, pp. 2726–2733.
[79] B. Hernandez and P. Herrero-Vinas, "Resistance trend estimation using regression analysis to enhance antimicrobial surveillance: A multi-centre study in London 2009–2016. Antibiotics (Basel)," in *Proc. Int. Congr. Image Signal Process., BioMed. Eng. Inform. (CISP-BMEI)*, vol. 10, 2021, p. 1267.
[80] F. Herrera and S. Ventura, *Multiple Instance Learning*. Berlin, Germany: Springer, Nov. 2016.
[81] M. Hosseinzadeh, A. M. Rahmani, B. Vo, M. Bidaki, M. Masdari, and M. Zangakani, "Improving security using SVM-based anomaly detection: Issues and challenges," *Soft Comput.*, vol. 25, no. 4, pp. 3195–3223, Oct. 2020.
[82] L. Hu, Y. Liu, and W. Qiu, "A deep spiking neural network anomaly detection method," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–13, Sep. 2022.
[83] W. Hu, Y. Liao, and R. Vemuri, "Robust anomaly detection using support vector machines," *Proc. Int. Conf. Mach. Learn.*, Jun. 2003, pp. 282–289.
[84] X. Huang, K. Xu, C. Huang, C. Wang, and K. Qin, "Multiple instance learning convolutional neural networks for fine-grained aircraft recognition," *Remote Sens.*, vol. 13, no. 24, p. 5132, Dec. 2021.
[85] T. Hussain, K. Muhammad, A. Ullah, Z. Cao, S. W. Baik, and V. H. C. de Albuquerque, "Cloud-assisted multiview video summarization using CNN and bidirectional LSTM," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 77–86, Jan. 2020.
[86] i2tutorials. (Oct. 2019). *What Are the Pros and Cons of the PCA?* Accessed: Mar. 27, 2023. [Online]. Available: https://www.i2tutorials.com/what-are-the-pros-and-cons-of-the-pca/
[87] R. T. Ionescu, F. S. Khan, M.-I. Georgescu, and L. Shao, "Object-centric auto-encoders and dummy anomalies for abnormal event detection in video," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7834–7843.
[88] M. Islam, A. S. Dukyil, S. Alyahya, and S. Habib, "An IoT enable anomaly detection system for smart city surveillance," *Sensors*, vol. 23, no. 4, p. 2358, Feb. 2023.
[89] R. Itano, T. Nohara, and T. Koita, "Crowd-aided anomaly detection in surveillance videos," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2022, pp. 3992–3994.
[90] A. R. Javed, Z. Jalil, W. Zehra, T. R. Gadekallu, D. Y. Suh, and M. J. Piran, "A comprehensive survey on digital video forensics: Taxonomy, challenges, and future directions," *Eng. Appl. Artif. Intell.*, vol. 106, Nov. 2021, Art. no. 104456.
[91] S. A. Jebur, K. A. Hussein, H. K. Hoomod, L. Alzubaidi, and J. Santamaría, "Review on deep learning approaches for anomaly event detection in video surveillance," *Electronics*, vol. 12, no. 1, p. 29, Dec. 2022.
[92] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," in *Proc. 25th ACM Int. Conf. Multimedia*, Oct. 2017, pp. 795–816.
[93] N. Joshi. (Feb. 2020). *Exploring the Limits of Transfer Learning*. [Online]. Available: https://www.allerin.com/blog/exploring-the-limits-of-transfer-learning
[94] M. M. Kabir, F. B. Safir, S. Shahen, J. Maua, I. A. B. Awlad, and M. F. Mridha, "Human abnormality classification using combined CNN-RNN approach," in *Proc. IEEE 17th Int. Conf. Smart Communities, Improving Quality Life Using ICT*, Dec. 2020, pp. 204–208.
[95] R. Kabir, Y. Watanobe, M. R. Islam, K. Naruse, and M. M. Rahman, "Unknown object detection using a one-class support vector machine for a cloud–robot system," *Sensors*, vol. 22, no. 4, p. 1352, Feb. 2022.
[96] A. M. Kamoona, A. K. Gostar, A. Bab-Hadiashar, and R. Hoseinnezhad, "Multiple instance-based video anomaly detection using deep temporal encoding–decoding," *Exp. Syst. Appl.*, vol. 214, Mar. 2023, Art. no. 119079.
[97] P. Kamranfar, D. Lattanzi, A. Shehu, and D. Barbará, "Multiple instance learning for detecting anomalies over sequential real-world datasets," 2022, *arXiv:2210.01707*.
[98] A. Khan, J. P. Li, N. Ahmad, S. Sethi, A. U. Haq, S. H. Patel, and S. Rahim, "Predicting emerging trends on social media by modeling it as temporal bipartite networks," *IEEE Access*, vol. 8, pp. 39635–39646, 2020.
[99] A. Khan, J. P. Li, and M. A. Husain, "Power grid stability analysis using pipeline machine," *Multimedia Tools Appl.*, vol. 82, no. 17, pp. 25651–25675, Feb. 2023.
[100] S. W. Khan, Q. Hafeez, M. I. Khalid, R. Alroobaea, S. Hussain, J. Iqbal, J. Almotiri, and S. S. Ullah, "Anomaly detection in traffic surveillance videos using deep learning," *Sensors*, vol. 22, no. 17, p. 6563, Aug. 2022.
[101] B. Kiran, D. Thomas, and R. Parakkal, "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos," *J. Imag.*, vol. 4, no. 2, p. 36, Feb. 2018.
[102] P. Kolaskar and A. Maitre, "Anomaly motion detection and tracking for real-time security system," in *Computer Networks and Inventive Communication Technologies*. 2021, pp. 707–717.
[103] A. Krieger, M. Pollak, and B. Yakir, "Surveillance of a simple linear regression," *J. Amer. Stat. Assoc.*, vol. 98, pp. 456–469, Feb. 2003.
[104] A. Kumar. (Apr. 2023). *Linear Regression Explained With Real Life Example*. Accessed: Mar. 20, 2023. [Online]. Available: https://vitalflux.com/linear-regression-real-life-example
[105] N. Kumar. (Mar. 2019). *The Professionals Point: Advantages and Disadvantages of Principal Component Analysis in Machine Learning*. Accessed: Mar. 24, 2023. [Online]. Available: http://theprofessionalspoint.blogspot.com/2019/03/advantages-and-disadvantages-of_4.html
[106] M. Kwet. (May 2023). *The Rise of Smart Camera Networks, and Why We Should Ban Them*. [Online]. Available: https://theintercept.com/2020/01/27/surveillance-cctv-smart-camera-networks/
[107] D. T. Lan and S. Yoon, "Trajectory clustering-based anomaly detection in indoor human movement," *Sensors*, vol. 23, no. 6, p. 3318, Mar. 2023.
[108] X. Lan, M. Ye, S. Zhang, H. Zhou, and P. C. Yuen, "Modality-correlation-aware sparse representation for RGB-infrared object tracking," *Pattern Recognit. Lett.*, vol. 130, pp. 12–20, Feb. 2020.
[109] J. Lani. (Jan. 2023). *Assumptions of Logistic Regression*. Accessed: Mar. 27, 2023. [Online]. Available: https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/assumptions-of-logistic-regression/
[110] V.-T. Le and Y.-G. Kim, "Attention-based residual autoencoder for video anomaly detection," *Int. J. Speech Technol.*, vol. 53, no. 3, pp. 3240–3254, Feb. 2023.
[111] Q. Li, R. Yang, F. Xiao, B. Bhanu, and F. Zhang, "Attention-based anomaly detection in multi-view surveillance videos," *Knowl.-Based Syst.*, vol. 252, Sep. 2022, Art. no. 109348.
[112] 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, Jun. 2021.
[113] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 18–32, Jan. 2014.

[114] Y. Li, Y. Xu, Y. Cao, J. Hou, C. Wang, W. Guo, X. Li, Y. Xin, Z. Liu, and L. Cui, "One-class LSTM network for anomalous network traffic detection," *Appl. Sci.*, vol. 12, no. 10, p. 5051, May 2022.

[115] Y. Liang, S. Li, C. Yan, M. Li, and C. Jiang, "Explaining the black-box model: A survey of local interpretation methods for deep neural networks," *Neurocomputing*, vol. 419, pp. 168–182, Jan. 2021.

[116] S. H. Lim, N. B. Erichson, L. Hodgkinson, and M. W. Mahoney, "Noisy recurrent neural networks, *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 5124–5137.

[117] S. Lin, H. Yang, X. Tang, T. Shi, and L. Chen, "Social MIL: Interaction-aware for crowd anomaly detection," in *Proc. 16th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Sep. 2019, pp. 1–8.

[118] S. Lin, S. Roberts, N. Trigoni, and R. Clark, "Balancing reconstruction quality and regularisation in ELBO for VAEs," *arXiv:1909.03765*.

[119] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[120] R. Lippmann and D. Fried, "Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation," in *Proc. DARPA Inf. Survivability Conf. Expo.*, vol. 2, 2000, pp. 12–26.

[121] A.-A. Liu, Z. Shao, Y. Wong, J. Li, Y.-T. Su, and M. Kankanhalli, "LSTM-based multi-label video event detection," *Multimedia Tools Appl.*, vol. 78, no. 1, pp. 677–695, Jan. 2019.

[122] K. Liu, M. Zhu, H. Fu, H. Ma, and T.-S. Chua, "Enhancing anomaly detection in surveillance videos with transfer learning from action recognition," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 4664–4668.

[123] W. Liu, W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly detection—A new baseline," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6536–6545.

[124] X. Liu, J. Xu, M. Li, and J. Peng, "Sensitivity analysis based SVM application on automatic incident detection of rural road in China," *Math. Problems Eng.*, vol. 2018, pp. 1–9, 2018.

[125] Y. Liu, X. Feng, and Z. Zhou, "Multimodal video classification with stacked contractive autoencoders," *Signal Process.*, vol. 120, pp. 761–766, Mar. 2016.

[126] Y. Liu, Z. Li, C. Zhou, Y. Jiang, J. Sun, M. Wang, and X. He, "Generative adversarial active learning for unsupervised outlier detection," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1517–1528, Sep. 2020.

[127] A. López-Cifuentes, M. Escudero-Viñolo, J. Bescós, and Á. García-Martín, "Semantic-aware scene recognition," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107256.

[128] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2720–2727.

[129] E. Luna, J. San Miguel, D. Ortego, and J. Martínez, "Abandoned object detection in video-surveillance: Survey and comparison," *Sensors*, vol. 18, no. 12, p. 4290, Dec. 2018.

[130] Q. Ma, C. Sun, B. Cui, and X. Jin, "A novel model for anomaly detection in network traffic based on kernel support vector machine," *Comput. Secur.*, vol. 104, May 2021, Art. no. 102215.

[131] M. Magfirawaty, F. Budi Setiawan, M. Yusuf, R. Kurniandi, R. F. Nafis, and N. Hayati, "Principal component analysis and data encryption model for face recognition system," in *Proc. 2nd Int. Conf. Electron. Electr. Eng. Intell. Syst.*, Nov. 2022, pp. 381–386.

[132] M. S. Mahmood. (Jul. 2021). *Outlier Detection in Regression Analysis*. Accessed: Mar. 20, 2023. [Online]. Available: https://towardsdatascience.com/

[133] M. S. Mahmud, J. Z. Huang, and X. Fu, "Variational autoencoder-based dimensionality reduction for high-dimensional small-sample data classification," *Int. J. Comput. Intell. Appl.*, vol. 19, no. 1, Mar. 2020, Art. no. 2050002.

[134] J. Masci, U. Meier, D. Cirean, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2011, pp. 52–59.

[135] A. Mehmood, "Abnormal behavior detection in uncrowded videos with two-stream 3D convolutional neural networks," *Appl. Sci.*, vol. 11, no. 8, p. 3523, Apr. 2021.

[136] Y. Miao and J. Song, "Abnormal event detection based on SVM in video surveillance," in *Proc. IEEE Workshop Adv. Res. Technol. Ind. Appl. (WARTIA)*, Sep. 2014, pp. 1379–1383.

[137] A. A. Mohamed, F. Alqahtani, A. Shalaby, and A. Tolba, "Texture classification-based feature processing for violence-based anomaly detection in crowded environments," *Image Vis. Comput.*, vol. 124, Aug. 2022, Art. no. 104488.

[138] N. H. Mohammed and S. C. R. Maram, "Fraud detection of credit card using logistic regression," *SSRN Electron. J.*, p. 4135514, Mar. 2022.

[139] M. A. Mondal and Z. Rehena, "Road traffic outlier detection technique based on linear regression," *Proc. Comput. Sci.*, vol. 171, pp. 2547–2555, Jan. 2020.

[140] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 44, pp. 22071–22080, Oct. 2019.

[141] M. Murugesan and S. Thilagamani, "Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network," *Microprocess. Microsyst.*, vol. 79, Nov. 2020, Art. no. 103303.

[142] P. Novello, G. Poëtte, D. Lugato, and P. M. Congedo, "Goal-oriented sensitivity analysis of hyperparameters in deep learning," *J. Sci. Comput.*, vol. 94, no. 3, p. 45, 2022.

[143] Y. Pan, S.-H. Tsang, Y.-L. Chan, and D. P. K. Lun, "Blur detection for surveillance camera system," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Nov. 2022, pp. 1879–1884.

[144] C. Parera, Q. Liao, I. Malanchini, D. Wellington, A. E. C. Redondi, and M. Cesana, "Transfer learning for multi-step resource utilization prediction," in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Aug. 2020, pp. 1–6.

[145] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1310–1318.

[146] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Comput. Netw.*, vol. 51, no. 12, pp. 3448–3470, Aug. 2007.

[147] K. Pawar and V. Attar, "Deep learning-based intelligent surveillance model for detection of anomalous activities from videos," *Int. J. Comput. Vis. Robot.*, vol. 10, no. 4, p. 289, 2020.

[148] K. Pawar and V. Attar, "Application of deep learning for crowd anomaly detection from surveillance videos," in *Proc. 11th Int. Conf. Cloud Comput., Data Sci. Eng.*, Jan. 2021, pp. 506–511.

[149] P. Pedamkar. (Jun. 2023). *Recurrent Neural Networks (RNN)*. Accessed: Mar. 29, 2023. [Online]. Available: https://www.educba.com/recurrent-neural-networks-rnn/

[150] D. Pesic, D. Pesic, A. Trifunovic, and S. Cicevic, "Application of logistic regression model to assess the impact of smartwatch on improving road traffic safety: A driving simulator study," *Mathematics*, vol. 10, no. 9, p. 1403, Apr. 2022.

[151] ProjectPro. (Feb. 2023). *The Ultimate Guide to Building Your Own LSTM Models*. Accessed: Mar. 14, 2023. [Online]. Available: https://www.projectpro.io/article/lstmmodel/832

[152] J. Prosise. (Jan. 2023). *PCA-Based Anomaly Detection*. Accessed: Mar. 27, 2023. [Online]. Available: https://www.atmosera.com/blog/pca-based-anomaly-detection

[153] P. Purkait, C. Zach, and I. Reid, "SG-VAE: Scene grammar variational autoencoder to generate new indoor scenes," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 155–171.

[154] M. Qasim and E. Verdu, "Video anomaly detection system using deep convolutional and recurrent models," *Results Eng.*, vol. 18, Jun. 2023, Art. no. 101026.

[155] M. Raamin, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 935–942.

[156] A. Rahman, Y. Chang, and J. Rubin, "Interpretable additive recurrent neural networks for multivariate clinical time series," 2021, *arXiv:2109.07602*.

[157] M. H. Rahman, M. Shahjalal, M. K. Hasan, M. O. Ali, and Y. M. Jang, "Design of an SVM classifier assisted intelligent receiver for reliable optical camera communication," *Sensors*, vol. 21, no. 13, p. 4283, Jun. 2021.

[158] R. Raja, P. C. Sharma, M. R. Mahmood, and D. K. Saini, "Analysis of anomaly detection in surveillance video: Recent trends and future vision," *Multimedia Tools Appl.*, vol. 82, no. 8, pp. 12635–12651, Mar. 2023.

[159] B. Ramachandra and M. J. Jones, "Street scene: A new dataset and evaluation protocol for video anomaly detection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 2558–2567.

[160] M. Ravanbakhsh, E. Sangineto, M. Nabi, and N. Sebe, "Training adversarial discriminators for cross-channel abnormal event detection in crowds," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1896–1904.

[161] K. Rezaee and S. Rezakhani, "A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance," *Pers. Ubiquitous Comput.*, vol. 25, no. 3, pp. 1–17, 2021.

[162] S. Roka, M. Diwakar, P. Singh, and P. Singh, "Anomaly behavior detection analysis in video surveillance: A critical review," *J. Electron. Imag.*, vol. 32, no. 4, Mar. 2023, Art. no. 042106.

[163] A. R. Rout. (Jan. 2023). *Advantages and Disadvantages of Logistic Regression*. Accessed: Mar. 24, 2023. [Online]. Available: https://www.geeksforgeeks.org/advantages-and-disadvantages-of-logistic-regression/

[164] M. Rwibasira, "ADOBSVM: Anomaly detection on block chain using support vector machine," *Meas., Sensors*, vol. 24, Dec. 2022, Art. no. 100503.

[165] M. Said Elsayed, N.-A. Le-Khac, S. Dev, and A. D. Jurcut, "Network anomaly detection using LSTM based autoencoder," in *Proc. 16th ACM Symp. QoS Secur. Wireless Mobile Netw.*, Nov. 2020, pp. 37–45.

[166] V. K. Saini. (Jan. 2022). *Understanding Support Vector Machines (SVMs) in Depth*. Accessed: Mar. 19, 2023. [Online]. Available: https://iq.opengenus.org/understand-support-vector-machine-in-depth/

[167] S. Samal and Y.-D. Zhang, "ASYv3: Attention-enabled pooling embedded Swin transformer-based YOLOv3 for obscenity detection," *Expert Syst.*, vol. 40, no. 8, May 2023.

[168] F. P. dos Santos, C. Zor, J. Kittler, and M. A. Ponti, "Learning image features with fewer labels using a semi-supervised deep convolutional network," *Neural Netw.*, vol. 132, pp. 131–143, Dec. 2020.

[169] K. S. Sujith and G. Sasikala, "Optimal support vector machine and hybrid tracking model for behaviour recognition in highly dense crowd videos," *Data Technol. Appl.*, vol. 55, no. 1, pp. 19–40, Nov. 2020.

[170] B. Sauvalle and A. de La Fortelle, "Autoencoder-based background reconstruction and foreground segmentation with background noise estimation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2022, pp. 3244–3255.

[171] S. Schmidl, P. Wenig, and T. Papenbrock, "Anomaly detection in time series: A comprehensive evaluation," *Proc. VLDB Endowment*, vol. 15, no. 9, pp. 1779–1797, May 2022.

[172] N. Seliya, A. A. Zadeh, and T. M. Khoshgoftaar, "A literature review on one-class classification and its potential applications in big data," *J. Big Data*, vol. 8, no. 1, p. 122, Sep. 2021.

[173] M. Shen, T. Sun, X. Jiang, and K. Xu, "Crowd counting estimation in video surveillance based on linear regression function," in *Proc. 9th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2016, pp. 60–65.

[174] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.

[175] W. Shin and S. B. Cho, "CCTV image sequence generation and modeling method for video anomaly detection using generative adversarial network," in *Intelligent Data Engineering and Automated Learning—IDEAL* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2018, pp. 457–467.

[176] W. Shin, S.-J. Bu, and S.-B. Cho, "3D-convolutional neural network with generative adversarial network and autoencoder for robust anomaly detection in video surveillance," *Int. J. Neural Syst.*, vol. 30, no. 6, Jun. 2020, Art. no. 2050034.

[177] O. Siméoni, Y. Avrithis, and O. Chum, "Local features and visual words emerge in activations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11643–11652.

[178] K. Sreekala, C. P. D. Cyril, S. Neelakandan, S. Chandrasekaran, R. Walia, and E. O. Martinson, "Capsule network-based deep transfer learning model for face recognition," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–12, Jul. 2022.

[179] G. Sreenu and M. A. S. Durai, "Intelligent video surveillance: A review through deep learning techniques for crowd analysis," *J. Big Data*, vol. 6, no. 1, pp. 1–27, Dec. 2019.

[180] H. Steck, "Autoencoders that don't overfit towards the identity," in *Proc. Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2020, pp. 19598–19608.

[181] F. Strub, R. Gaudel, and J. Mary, "Hybrid recommender system based on autoencoders," in *Proc. 1st Workshop Deep Learn. Recommender Syst.*, Sep. 2016, pp. 11–16.

[182] J. J. P. Suarez and P. C. Naval, "A survey on deep learning techniques for video anomaly detection," 2020, *arXiv:2009.14146*.

[183] L. Sullivan. (Jan. 2022). *Correlation and Linear Regression*. Accessed: Mar. 29, 2023. [Online]. Available: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Correlation-Regression/

[184] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6479–6488.

[185] Indeed Editorial Team. (Jan. 2022). *Linear vs. Logistic Regression (Differences and Limitations)*. Accessed: Mar. 29, 2023. [Online]. Available: https://ca.indeed.com/career-advice/career-development/linear-vs-logistic-regression

[186] C.-W. Tien, T.-Y. Huang, P.-C. Chen, and J.-H. Wang, "Using autoencoders for anomaly detection and transfer learning in IoT," *Computers*, vol. 10, no. 7, p. 88, Jul. 2021.

[187] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4489–4497.

[188] R. K. Tripathi, A. S. Jalal, and S. C. Agrawal, "Suspicious human activity recognition: A review," *Artif. Intell. Rev.*, vol. 50, no. 2, pp. 283–339, Aug. 2018.

[189] A. Ullah, K. Muhammad, I. U. Haq, and S. W. Baik, "Action recognition using optimized deep autoencoder and CNN for surveillance data streams of non-stationary environments," *Future Gener. Comput. Syst.*, vol. 96, pp. 386–397, Jul. 2019.

[190] 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," *Multimedia Tools Appl.*, vol. 80, no. 11, pp. 16979–16995, May 2021.

[191] W. Ullah, A. Ullah, T. Hussain, Z. A. Khan, and S. W. Baik, "An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos," *Sensors*, vol. 21, no. 8, p. 2811, Apr. 2021.

[192] U. A. Umoh, E. Udo, and E. E. Nyoho, "Support vector machine-based fire outbreak detection system," 2019, *arXiv:1906.05655*.

[193] N. Vallez, A. Velasco-Mata, and O. Deniz, "Deep autoencoder for false positive reduction in handgun detection," *Neural Comput. Appl.*, vol. 33, no. 11, pp. 5885–5895, 2021.

[194] K. K. Verma, B. M. Singh, and A. Dixit, "A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system," *Int. J. Inf. Technol.*, vol. 14, no. 1, pp. 397–410, Sep. 2019.

[195] A. Vijayan, B. Meenaskshi, A. Pandey, A. Patel, and A. Jain, "Video anomaly detection in surveillance cameras," in *Proc. Int. Conf. Advancement Technol. (ICONAT)*, Jan. 2022, pp. 1–4.

[196] J. von Schleinitz, M. Graf, W. Trutschnig, and A. Schröder, "VASP: An autoencoder-based approach for multivariate anomaly detection and robust time series prediction with application in motorsport," *Eng. Appl. Artif. Intell.*, vol. 104, Sep. 2021, Art. no. 104354.

[197] K. Vos, Z. Peng, C. Jenkins, M. R. Shahriar, P. Borghesani, and W. Wang, "Vibration-based anomaly detection using LSTM/SVM approaches," *Mech. Syst. Signal Process.*, vol. 169, Apr. 2022, Art. no. 108752.

[198] T.-H. Vu, J. Boonaert, S. Ambellouis, and A. Taleb-Ahmed, "Multi-channel generative framework and supervised learning for anomaly detection in surveillance videos," *Sensors*, vol. 21, no. 9, p. 3179, May 2021.

[199] A. Harjoko, A. Dharmawan, F. D. Adhinata, G. Kosala, and K.-H. Jo, "Loitering detection using spatial–temporal information for intelligent surveillance systems on a vision sensor," *J. Sensor Actuator Netw.*, vol. 12, no. 1, p. 9, Jan. 2023.

[200] J. Wang, A. Cherian, F. Porikli, and S. Gould, "Video representation learning using discriminative pooling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1149–1158.

[201] W. Wang, X. Du, D. Shan, R. Qin, and N. Wang, "Cloud intrusion detection method based on stacked contractive auto-encoder and support vector machine," *IEEE Trans. Cloud Comput.*, vol. 10, no. 3, pp. 1634–1646, Jul. 2022.

[202] Z. Wang. (Mar. 2018). *Analyzing Optimistic Concurrency Control Anomalies and Solutions*. Accessed: Mar. 27, 2023. [Online]. Available: https://wangziqi2013.github.io/article/2018/03/21/Analyzing-OCC-Anomalies-and-Solutions.html

[203] F. Waseem, R. P. Martinez, and C. Wu, "Visual anomaly detection in video by variational autoencoder," 2022, *arXiv:2203.03872*.

[204] P. Wu and J. Liu, "Not only look, but also listen: Learning multimodal violence detection under weak supervision," in *Computer Vision—ECCV 2020*. 2020, pp. 322–339.
[205] R. Wu and B. Wang, "Harvesting discriminative meta objects with deep CNN features for scene classification," 2015, *arXiv:1510.01440*.
[206] W. Wu, "Multi-source selection transfer learning with privacy-preserving," *Neural Process. Lett.*, vol. 54, no. 6, pp. 4921–4950, May 2022.
[207] X. Xia, X. Pan, N. Li, X. He, L. Ma, X. Zhang, and N. Ding, "GAN-based anomaly detection: A review," *Neurocomputing*, vol. 493, pp. 497–535, Jul. 2022.
[208] W. Xiao, X. Huang, J. Silva, S. Emrani, and A. Chaudhuri, "Online robust principal component analysis with change point detection," 2017, *arXiv:1702.05698*.
[209] M. Xu, X. Yu, D. Chen, C. Wu, and Y. Jiang, "An efficient anomaly detection system for crowded scenes using variational autoencoders," *Appl. Sci.*, vol. 9, no. 16, p. 3337, Aug. 2019.
[210] W. Hao, R. Zhang, S. Li, J. Li, F. Li, S. Zhao, and W. Zhang, "Anomaly event detection in security surveillance using two-stream based model," *Secur. Commun. Netw.*, vol. 2020, pp. 1–15, Aug. 2020.
[211] K. Yang, S. Kpotufe, and N. Feamster, "An efficient one-class SVM for anomaly detection in the Internet of Things," 2021, *arXiv:2104.11146*.
[212] Y. Yang, J. Yu, C. Wang, and J. Wen, "Risk assessment of crowd-gathering in urban open public spaces supported by spatio-temporal big data," *Sustainability*, vol. 14, no. 10, p. 6175, May 2022.
[213] R. Yao, C. Liu, L. Zhang, and P. Peng, "Unsupervised anomaly detection using variational auto-encoder based feature extraction," in *Proc. IEEE Int. Conf. Prognostics Health Manag. (ICPHM)*, Jun. 2019, pp. 1–7.
[214] L. Ye, S. Yan, J. Zhen, T. Han, H. Ferdinando, T. Seppänen, and E. Alasaarela, "Physical violence detection based on distributed surveillance cameras," *Mobile Netw. Appl.*, vol. 27, no. 4, pp. 1688–1699, Feb. 2022.
[215] Z. Yin and P. Barucca, "Stochastic recurrent neural network for multistep time series forecasting," 2021, *arXiv:2104.12311*.
[216] J. Yu, J. Liu, Y. Cheng, R. Feng, and Y. Zhang, "Modality-aware contrastive instance learning with self-distillation for weakly-supervised audio-visual violence detection," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 6278–6287.
[217] J. Yuan, X. Wu, and S. Yuan, "A rapid recognition method for pedestrian abnormal behavior," in *Proc. Int. Conf. Comput. Vis., Image Deep Learn. (CVIDL)*, Jul. 2020, pp. 241–245.
[218] U. Zahoora, M. Rajarajan, Z. Pan, and A. Khan, "Zero-day ransomware attack detection using deep contractive autoencoder and voting based ensemble classifier," *Int. J. Speech Technol.*, vol. 52, no. 12, pp. 13941–13960, Sep. 2022.
[219] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, "Efficient GAN-based anomaly detection," 2018, *arXiv:1802.06222*.
[220] F. Zhang, T.-Y. Wu, J.-S. Pan, G. Ding, and Z. Li, "Human motion recognition based on SVM in VR art media interaction environment," *Hum.-Centric Comput. Inf. Sci.*, vol. 9, no. 1, p. 40, Dec. 2019.
[221] H.-B. Zhang, Y.-X. Zhang, B. Zhong, Q. Lei, L. Yang, J.-X. Du, and D.-S. Chen, "A comprehensive survey of vision-based human action recognition methods," *Sensors*, vol. 19, no. 5, p. 1005, Feb. 2019.
[222] J. Zhang, C. Meng, C. Xu, J. Ma, and W. Su, "Deep transfer learning method based on automatic domain alignment and moment matching," *Mathematics*, vol. 10, no. 14, p. 2531, Jul. 2022.
[223] T. Zhang, W. Aftab, L. Mihaylova, C. Langran-Wheeler, S. Rigby, D. Fletcher, S. Maddock, and G. Bosworth, "Recent advances in video analytics for rail network surveillance for security, trespass and suicide prevention—A survey," *Sensors*, vol. 22, no. 12, p. 4324, Jun. 2022.
[224] W. Zhang, P. He, S. Wang, L. An, and F. Yang, "A dynamic convolutional generative adversarial network for video anomaly detection," *Arabian J. Sci. Eng.*, vol. 48, no. 2, pp. 2075–2085, Jul. 2022.
[225] W. Zhao, Y. Xue, and X. Liu, "Monitoring parameter change in linear regression model based on the efficient score vector," *Phys. A, Stat. Mech. Appl.*, vol. 527, Aug. 2019, Art. no. 121135.
[226] Y. Zhao, H. Zhu, R. Liang, Q. Shen, S. Zhang, and K. Chen, "Seeing isn't believing: Towards more robust adversarial attack against real world object detectors," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2019, pp. 1989–2004.
[227] W. Zheng, K. Wang, and F. Wang, "Background subtraction algorithm based on Bayesian generative adversarial networks," *Acta Autom. Sinica*, vol. 44, no. 5, pp. 878–890, 2018.
[228] X. Zheng and X. Yin, "A privacy-preserved variational-autoencoder for DGA identification in the education industry and distance learning," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–8, Mar. 2022.
[229] F. Zhou, L. Wang, Z. Li, W. Zuo, and H. Tan, "Unsupervised learning approach for abnormal event detection in surveillance video by hybrid autoencoder," *Neural Process. Lett.*, vol. 52, no. 2, pp. 961–975, Oct. 2020.
[230] Y. Zhou, X. Song, Y. Zhang, F. Liu, C. Zhu, and L. Liu, "Feature encoding with autoencoders for weakly supervised anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2454–2465, Jun. 2022.
[231] S. Zhu, C. Chen, and W. Sultani, "Video anomaly detection for smart surveillance," 2020, *arXiv:2004.00222*.
[232] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023.

**NOMICA CHOUDHRY** is currently pursuing the Ph.D. degree with Deakin University, Australia. Additionally, she holds the position of a Lecturer with the National University of Modern Languages, Pakistan. Her ongoing research interest includes anomaly detection through the application of deep learning techniques.

**JEMAL ABAWAJY** is currently a Full Professor with Deakin University, Australia. He is also the Director of the Parallel and Distributing Computing Laboratory. He is also a respected international scholar with numerous publications to his name; including several books, conference volumes, and more than 500 refereed papers in conferences and journals. He has been an integral part of the organizing committees for more than 300 international conferences, serving in varied roles, including those of the chair and the general co-chair. Furthermore, he has mentored numerous Ph.D. students during his distinguished career.

**SHAMSUL HUDA** received the Ph.D. degree in computer science from Federation University, in 2009. He is currently a Senior Lecturer with Deakin University. His research interests include communication and network security, the IoT security, malware detection, and reverse engineering for endpoint security.

**IMRAN RAO** received the Ph.D. degree from The University of Melbourne, Australia. He is currently the Director of IT with Blue Brackets Technologies, Australia and Pakistan. He is also a senior IT professional with vast academic experience in teaching and research at prestigious universities. His current research interests include deep learning and quantum computing.