Artificial Intelligence Techniques in Surveillance Video Anomaly Detection

Subjects: Computer Science, Artificial Intelligence | Computer Science, Cybernetics
Contributor: Erkan Şengönül, Refik Samet, Qasem Abu Al-Haija, Ali Alqahtani, Badraddin Alturki, Abdulaziz A. Alsulami

The Surveillance Video Anomaly Detection (SVAD) system is a sophisticated technology designed to detect unusual or suspicious behavior in video surveillance footage without human intervention. The system operates by analyzing the video frames and identifying deviations from normal patterns of movement or activity. This is achieved through advanced algorithms and machine learning techniques that can detect and analyze the position of pixels in the video frame at the time of an event.

Keywords: video surveillance ; abnormal events ; anomaly detection ; artificial intelligence

1. Introduction

The taxonomy of surveillance video anomaly detection (SVAD), consisting of two main groups, is described in **Table 1**.

Table 1. Taxonomy for anomaly detection in video surveillance.

Learning	Algorithms
Supervised learning	**Statistics-based algorithms**
Unsupervised learning	**Classification-based algorithms**
Semi-supervised learning	**Reconstruction-based algorithms**
	Prediction-based algorithms
	Other algorithms

2. Learning

Several Artificial Intelligence (AI) subsets are based on various applications and use cases. This text mainly focused on Machine Learning (ML) and Deep Learning (DL). DL is a subset of machine learning methods. ML is a powerful technology that can be applied for anomaly detection. The process varies considerably depending on the problem. The performance of an ML algorithm may vary depending on the features selected in the dataset or the weight assigned to each feature, even if the same model runs on two identical datasets [1]. A model may become overfit if it has fewer features that are only sometimes good. To better comprehend and construct a model using available ML techniques and data, reviewing and comparing the current solutions is worthwhile. Machine learning (ML) can be divided into three groups: Supervised Learning (SL), Unsupervised Learning (UL), and Semi-Supervised Learning (SSL).

2.1. Supervised Learning

SL acquires knowledge from pre-existing labeled datasets or “the training set”, then compares the predicted output to the known labels. A high-level training set is always required to build a model that works effectively, but more is needed to ensure that the final product will be satisfactory; the training procedure is also a crucial element in creating a reliable predictor. A classifier model is first developed in SL through training, and after that, it can forecast either discrete or continuous outputs. The ASL model’s performance, such as accuracy, is typically validated before prediction to demonstrate its dependability. Additionally, classification and regression techniques can be used to categorize SL tasks [2].

The training data are first divided into separate categories in the classification technique. It then calculates the probability of test samples falling into each category and chooses the category with the most votes [3]. This probability represents the likelihood that a sample is a class member. Credit scoring and medical imaging are examples of typical applications. The regression technique uses input factors such as temperature changes or variations in electricity demand to forecast continuous responses, often in quantity [4]. Forecasting power load and algorithmic trading are examples of typical applications. While the regression model can calculate the root-mean-squared error, the classification model can quantify the percentage of accurate predictions. Nevertheless, a discrepancy between the expected and actual values is acceptable since the output data are continuous.

Several works have been performed with SL. One of the suggestions in this area is presented by the study [5]. They proposed a unique way to identify fights or violent acts based on learning the temporal and spatial information from consecutive video frames that are evenly spaced. Using the proposed feature fusion approach, features with many levels for two sequential frames are retrieved from the first and last layers of the Convolutional Neural Network (CNN) and fused to consider the action knowledge. They also suggested a "Wide-Dense Residual Block" to learn the unified spatial data from the two input frames. These learned characteristics are subsequently consolidated and delivered to long-term memory components to store temporal dependencies. Using the domain adaptation strategy, the network may learn to efficiently merge features from the input frames, improving the results' accuracy. They evaluated their experiments by using four public datasets, namely HockeyFight, Movies, ViolentFlow, and BEHAVE, to show the performance of their model, which was compared with the existing models. There are several important learning techniques in SL, such the Hidden Markov Model (HMM) [6], Support Vector Machine (SVM) [7], Gaussian Regression (GR) [8], CNN [9], Multiple Instance Learning (MIL) [10], and Long Short-Term Memory (LSTM) [11]. It is clear that each technique has advantages and disadvantages in anomaly detection, and it is impossible to say that one technique can solve all problems efficiently.

2.2. Unsupervised Learning

UL groups data by identifying hidden patterns or intrinsic structures. Data input is necessary, but there are no predetermined output variables. There is neither labeled input data nor a training technique, in contrast to SL. As a result, it operates independently, and its performance could be more measurable. Although some researchers use the UL model's pre-existing labeled data to verify its results, this is only sometimes possible in practice. To conduct an external evaluation, specialists may need to analyze the results manually.

UL is mostly used for reducing dimensionality and clustering. UL is used in dimensionality reduction to find the dataset's linked features so that redundant data can be removed to reduce noise. Using clustering techniques, the clustering problem allows for the possibility of a sample belonging to more than one cluster or just one. Market research and object identification are common applications [12].

One proposed approach in UL is that of [13]. They provided a technique for detecting anomalies in surveillance missions, including UAV-acquired footage. They combined an unsupervised classification technique called One-Class Support Vector Machine (OCSVM) with a deep feature extraction technique utilizing a pre-trained CNN. Their quantitative findings demonstrated that their proposed strategy produces positive outcomes for the dataset studied. The authors in [14] extended their previous work by using mobile cameras to assist UAVs when acquiring videos. They added two feature extraction methods, the Histogram of Oriented Gradients (HOG) and HOG3D. They used the same UL method, which was OCSVM [15]. They obtained good results based on the used video-obtained datasets. There are many techniques under UL; PCA [16] and GANs [17] are examples of them.

2.3. Semi-Supervised Learning

SSL is a machine learning method that utilizes labeled and unlabeled data to create a classifier. This approach is particularly useful in situations with a limited amount of labeled data available. The SSL algorithm utilizes the training procedure described in Supervised Learning (SL) to create a predictor with a small amount of labeled data. The predictor then categorizes unlabeled samples and assigns each pseudo-labeled sample a confidence rating. This confidence rating informs the administrator of the prediction's certainty level. Once all data have been labeled, confident examples are added to the new training set to update the classifier.

Certain assumptions must be made before training unlabeled examples, such as smoothness and clustering. This is because unlabeled data are randomly labeled in the prediction process [18]. The anomaly detection (AE) model [19] is an important SSL model, as it utilizes labeled and unlabeled data to detect and identify anomalies in a given dataset. Overall, SSL is an effective method for creating a classifier with a limited amount of labeled data while leveraging the information present in unlabeled data to improve the accuracy of the classifier.

2.4. Supervised vs. Unsupervised vs. Semi-Supervised

Supervised learning techniques for SVAD offer several advantages, including the ability to accurately identify and classify anomalies using labeled data and the ability to identify specific types of anomalies. These techniques are also useful for detecting anomalies in surveillance and security applications. However, a significant amount of labeled data is required, and these techniques can be sensitive to environmental changes, affecting their accuracy.

Unsupervised learning techniques for SVAD offer advantages such as not requiring labeled data and the ability to detect anomalies in real-time. These techniques can also be used to identify patterns in the data that deviate from the norm and classify them as anomalies. However, unsupervised learning techniques are not able to identify specific types of anomalies and can also be sensitive to changes in the environment.

Semi-supervised learning techniques for SVAD can use labeled and unlabeled data, allowing for accurate identification and classification of anomalies. These techniques can also be used to identify specific types of anomalies and detect anomalies in real-time. However, semi-supervised learning techniques require significant labeled data and can also be sensitive to environmental changes.

In conclusion, supervised, unsupervised, and semi-supervised learning techniques each offer advantages and disadvantages when it comes to anomaly detection in SVAD. Each technique has its limitations, and the accuracy of the results can be affected by changes in the environment. Therefore, the choice of technique will depend on the specific needs of the application and the availability of labeled data.

3. Algorithms

3.1. Statistics-Based Algorithms

Two main algorithms are used in video anomaly detection: parametric and non-parametric [20].

Parametric algorithms assume the data follow a specific probability distribution, such as a Gaussian distribution. These algorithms estimate the parameters of the distribution using the data and then use these parameters to calculate the likelihood of new data points. One popular parametric algorithm for video anomaly detection is the Gaussian Mixture Model (GMM). The GMM is a probabilistic model representing a dataset as a mixture of multiple Gaussian distributions. The algorithm estimates the parameters of the Gaussian distributions using the data and then uses these parameters to calculate the likelihood of new data points. If the likelihood of a new data point is below a certain threshold, it is considered an anomaly.

Non-parametric algorithms do not make any assumptions about the distribution of the data. Instead, these algorithms rely on the empirical distribution of the data, which is estimated using Kernel Density Estimation (KDE) [21]. One popular non-parametric algorithm for video anomaly detection is the Local Outlier Factor (LOF) [22]. The LOF is a density-based algorithm that calculates the local density of a data point by measuring the distance to its k-nearest neighbors. The algorithm then compares a data point's local density to its neighbors' density. The data point is considered an anomaly if the ratio is below a certain threshold. Several studies have been conducted on statistical-based algorithms, some of which are listed below: Gaussian Mixture Model (GMM), selective histogram of optical flow, Histogram of Magnitude and Momentum (HoMM), Histogram of the oriented Swarm (HoS), Histogram of Gradients (HoG), Bayesian, Fully-Convolutional-Network (FCNs)-based models, and Structural Context Descriptor (SCD). Some statistics-based studies are presented in **Table 2**.

Table 2. Statistics-based methods.

Methods	Summary
HoMM [23]	**The histogram of magnitudes is used to record the motion of objects. The anomalous motion of an object is represented by its momentum about the foreground region's occupancy. Feature descriptors for typical situations are learned in an unsupervised manner using K-means clustering. By measuring the distance between cluster centers and the test frame's feature vector, frame-level anomalies are discovered. The region and anything leaving that region are regarded as anomalous.** **Datasets: UCSD, UMN; Techniques: background subtraction, optical flow, K-means clustering.**

Methods	Summary
Novel scheme based on SVDD [24]	**Statistical histograms are used to model normal motion distributions. It combines motion detection and appearance detection criteria to find anomalous objects. They created a method based on Support Vector Data Description (SVDD), which creates a sphere-shaped boundary around the regular items to keep out anomalous ones. They took into account a fixed-dimension region, and anything left in that region is regarded as anomalous.** **Datasets: UCSD, UMN, Subway; Techniques: histogram model, optical flow, support vector data description.**
Gaussian classifier [25]	**Deep autoencoder networks and single-class image-level classification are proposed to detect event anomalies in surveillance videos.** **Datasets: UCSD, Subway; Techniques: Gaussian classifier, CNN, sparse autoencoder.**
CoP [26]	**The method called Consistency Pursuit (CoP) is based on the idea that normal samples have a very high correlation with each other, can span low-dimensional subspaces, and therefore, have strong mutual consistency with a large number of data points.** **Datasets: Hopkins155; Techniques: robust PCA, saliency map.**

3.2. Classification-Based Algorithms

One of the most-widely used methods for SVAD is classification-based methods, which involve training a classifier to distinguish between normal and anomalous video frames or segments.

The first step in using classification-based methods for video anomaly detection is to extract features from the video frames. These features can include spatial and temporal information, such as color, texture, motion, and object shape. Several feature extraction techniques have been proposed in the literature, including hand-crafted features, such as the Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT), as well as in-depth learning-based features, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

Once the features have been extracted, the next step is to train a classifier to distinguish between normal and anomalous video frames or segments. Several classifiers have been proposed in the literature, including traditional machine learning classifiers, such as Support Vector Machines (SVMs), random forests, k-Nearest Neighbors (kNNs), and deep learning-based classifiers: CNNs and RNNs. The choice of the classifier will depend on the specific application and the type of features that have been extracted.

After the classifier has been trained, it can classify new video frames or segments as normal or anomalous. The classifier will output a score or probability for each frame or segment, indicating the likelihood that it is normal or anomalous. A threshold is usually set to make a final decision, and any frames or segments with a score below the threshold are considered anomalous.

One of the main advantages of classification-based methods for video anomaly detection is that they can be fine-tuned to a specific application by selecting appropriate features and classifiers. However, one of the main challenges is that these methods require a large amount of labeled training data to be effective. Additionally, they may be unable to detect anomalous events significantly different from the training data [20].

Several classification algorithms have been proposed in the literature on data science, which can be considered the most common in the field, and they were discussed in detail in [27]. Some commonly used algorithms are summarized as follows.

Support Vector Machine (SVM) is a widely used classification, regression, or other application method. An SVM generates a single hyperplane or a set of hyperplanes in a high or endless space. The goal is to separate the two classes using a hyperplane that reflects the greatest separation or margin. The larger the margin, the smaller the generalization error of the classifier is.

k-Nearest Neighbors (kNN) is a non-parametric supervised learning technique, also referred to as a “lazy learning” method. It maintains all occurrences that match the training set in an n-dimensional space, rather than focusing on building a large internal model. kNN uses data and employs similarity metrics to categorize new data points.

Decision Tree (DT) is a popular non-parametric SL approach. Both the classification and regression tasks are performed using DT learning techniques. The DT is a recursive operation; it starts with a single node and branches into a tree structure.

Some classification-based studies are shown in **Table 3**.

Table 3. Classification-based methods.

Methods	Summary
One-class classification [28]	**A histogram of optical flow orientation is integrated with a one-class SVM to identify abnormal events. Modeling high-density scenes may be performed quickly and precisely using optical flow techniques. Pattern identification is performed after feature extraction to discriminate between regular and irregular activities.** **Datasets: UMN; Techniques: SVM, optical flow, histogram of optical flow orientation.**
Asymptotic bounds [29]	**The crowd escape anomaly is detected using statistical and deep learning algorithms that directly evaluate the pixel coordinates.** **Datasets: UCSD, Avenue, ShanghaiTech; Techniques: YOLOv3, GAN-based frame predictor, kNN.**
Decision tree [30]	**To detect abnormalities from video surveillance while precisely estimating the start and end times of the anomalous event, a decision-tree-enabled solution leveraging deep learning was created.** **Datasets: ImageNet, COCO; Techniques: decision trees, YOLOv5.**
AE with kNN [31]	**A new approach combines an AE-based method with single-class deep feature classification. An AE is trained using normal images; then, anomaly maps are embedded using a pre-trained CNN feature extractor. A one-class classifier with kNN is trained to calculate the anomaly score.** **Datasets: MVTec; Techniques: convolutional autoencoder, high-density embedding, one-class classification.**
IGD [32]	**There is a high probability of overfitting as abnormal datasets are insufficient. The Interpolated Gaussian Descriptor (IGD) method, an OCC model that learns a one-class Gaussian anomaly classifier trained with inversely interpolated training samples, is proposed to solve this problem. The IGD is used to learn more meaningful data descriptions from typical normal samples. The crowd escape anomaly is detected using statistical and deep learning algorithms that directly evaluate the pixel coordinates.** **Datasets: MNIST, Fashion MNIST, CIFAR10, MVTec AD; Techniques: Gaussian classifier.**
Out of distribution [33]	**A classifier that is simultaneously trained to give the GAN samples less confidence is used in conjunction with a GAN. Samples from each test distribution of anomalies are used to arrange the classifier and GAN.** **Datasets: CIFAR, tree-enabled, LSUN; Techniques: DNN, GAN, Kullback–Leibler, Gaussian distribution.**

3.3. Reconstruction-Based Algorithms

Reconstruction-based methods operate under the presumption that normal data can be integrated into a lower-dimensional domain where normal samples and anomalies are represented in various ways [34].

An **Autoencoder (AE)** is a feed-forward neural network that includes an encoder and a decoder structure [35]. The objective is to train the network to capture the important parts of the input data and learn a lower-dimensional representation of the higher-dimensional data. The **Variational Autoencoder (VAE)** is a type of AE that includes an encoder network and a decoder network. The encoder network maps the input data to a low-dimensional latent space, while the decoder network maps the latent space back to the original data space. In this method, the VAE is trained on normal videos. The trained model is then used to reconstruct the input video, and the reconstruction error is calculated. Anomalies are detected by thresholding the reconstruction error. Any frame with a reconstruction error above a certain threshold is considered anomalous. The **Convolutional Autoencoder (CAE)** is also a type of AE consisting of convolution, deconvolution, pooling, and unpooling layers. The first two layer types may be found in the encoding step, whereas the others may be found in the decoding stage [36]. The **Variational Autoencoder (VAE)** is another type of AE that incorporates convolution, deconvolution, pooling, and unpooling layers. The first two layer types are used in the encoding step, while the others are used in the decoding stage [36].

Reconstruction-based methods are a variation of adversarial generative methods. **Generative-Adversarial-Network (GAN)**-based networks consist of two neural networks: a Generator (G) and a Discriminator (D) [35]. The generator network creates new examples in the target domain by mapping examples from the source domain to the target domain. The discriminator network then tries to distinguish between examples created by the generator and examples from the target domain. Through this process, the generator network learns to create examples indistinguishable from examples in the target domain.

In summary, reconstruction-based methods such as AEs and GANs have shown promising results in anomaly detection tasks by mapping normal data into a lower-dimensional domain and identifying anomalies based on the reconstruction error. Variants of AEs, such as Conv AEs and variational AEs, have also been utilized in this domain. These methods are part of a larger field of adversarial generative methods, including generative adversarial networks.

Some reconstruction-based studies are shown in **Table 4**.

Table 4. Reconstruction-based methods.

Methods	Summary
ST-AE [37]	**The spatiotemporal AE comprises one encoder and two decoders of 3D convolutional layers. It employs parallel training of decoders with monochrome frames, which is noteworthy compared to the distillation process.** **Datasets: Traffic, UCSD, Avenue; Techniques: CNN, autoencoder.**
AMDN [38]	**The appearance and motion DeepNet model employs AEs and a modified two-stream network with an additional third stream to improve detection performance. The two-stream method has two major drawbacks: the requirement for a pre-processing technique, such as optical flow, which may be costly for real-world applications, and multiple networks for inference.** **Datasets: Train, UCSD; Techniques: one-class SVM, optical flow.**
GMFC-VAE [39]	**The Gaussian mixture fully convolutional-variational AE uses the conventional two-stream network technique and uses a variational AE to enhance its feature extraction capability. This method estimates the appearance and motion anomaly score before combining the two clues to provide the final detection results.** **Datasets: Avenue, UCSD; Techniques: convolutional autoencoder, Gaussian mixture model.**
OF-ConvAE-LSTM [40]	**This method uses the convolutional AE and long short-term memory to detect anomalies. The framework produces the error function and reconstructed dense optical flow maps.** **Datasets: Avenue, UCSD; Techniques: convolutional autoencoder, LSTM, optical flow.**
Temporal cues [41]	**A conditional GAN is trained to learn two renderers that map pixel data to motion and vice versa. As a result, normal frames will have little reconstruction loss, while anomalous frames will have significant reconstruction loss.** **Datasets: Avenue, ShanghaiTech; Techniques: GAN, LSTM, optical flow.**
Ada-Net [42]	**An attention-based autoencoder using contentious learning is proposed to detect video anomalies.** **Datasets: UCSD, Avenue, ShanghaiTech; Techniques: GAN, autoencoder.**
Adversarial 3D CAE [43]	**A 3D CAE-based competitor anomalous event detection method is proposed to obtain the maximum accuracy by simultaneously learning motion and appearance features. It was developed to explore spatiotemporal features that help detect anomalous events in video frames.** **Datasets: UCSD, Avenue, Subway, ShanghaiTech; Techniques: convolutional autoencoder.**
Conv-AE + U-Net [44]	**A two-stream model is created that learns the connection between common item appearances and their related motions. A single encoder is paired with a U-net decoder to predict motion and a deconvolution decoder that reconstructs the input frame under the control of the** lp�� **reconstruction error loss terms using a single frame as the input.** **Datasets: UCSD, Avenue, Subway, Traffic; Techniques: convolutional autoencoder.**

3.4. Prediction-Based Algorithms

Prediction-based techniques can identify anomalies by assessing the difference between a feature descriptor's expected and actual spatiotemporal properties [34]. These models assume that normal activities are predictable, and any deviation from the prediction indicates an anomaly. They typically use a **Recurrent Neural Network (RNN)** to predict the next frame in the sequence, given the previous frames. During training, the model minimizes the difference between the predicted frame and the ground truth. Here are some commonly used algorithms:

Long Short-Term Memory (LSTM) is the most widely used neural array model, combining the principles of the forget gate, entry gate, and exit gate and successfully avoiding back-propagation errors caused by vanishing/exploding gradients.

The **convolutional LSTM** is an LSTM variation that addresses the precipitation nowcasting problem. In contrast to LSTM, convolution operations are employed to calculate the feature maps instead of matrix operations, resulting in a significant decrease in the count of the training parameters of the model [36].

Another prediction-based approach is the **Vision Transformer (ViT)** [45][46][47]. The ViT model combines CNNs and transformers to extract spatiotemporal features from video data and model the temporal relationships between these features. This approach effectively captures long-term dependencies in the video data and is especially useful for detecting anomalies.

In summary, RNN-based prediction techniques are effective at detecting anomalies by comparing the expected and actual spatiotemporal properties of a feature descriptor. LSTM is the most widely used and successful neural array model, while the convolutional LSTM and ViT are variations that address specific problems.

Some prediction-based studies are shown in **Table 5**.

Table 5. Prediction-based methods.

Methods	Summary
FFP [34]	Spatial and motion constraints are used to estimate the future frame for normal events in addition to density and gradient losses. Datasets: UCSD, ShanghaiTech, Avenue; Techniques: GAN, optical flow.
Deep BD-LSTM [48]	A model combining CNN and bidirectional LSTM is proposed to recognize the human movement in video sequences. Datasets: YouTube 11 Actions, UCF-101, HMDB51; Techniques: LSTM, CNN.
LSTM [49]	By using the effective gradient and quadratic programming-based training methods, the parameters of the LSTM architecture and the support vector data description algorithm are trained and optimized. Datasets: Avenue, Subway, ShanghaiTech, UCSD; Techniques: LSTM, one-class SVM.
SSPCAB [50]	A Self-Supervised Predictive Convolutional Attentive Block (SSPCAB) is proposed, which can be easily incorporated into various anomaly detection methods. The block acquires the ability to recreate the masked area utilizing contextual information for each site where the dilated convolutional filter is applied. Datasets: Avenue, MVTec AD, ShanghaiTech; Techniques: CNN, convolutional attentive block.
Spatiotemporal feature extraction [51]	A neural network built with transaction blocks, including dictionary learning, feature learning, and sparse representation, is proposed. A novel long short-term memory was also proposed and reformulated using an adaptive iterative hard-thresholding technique (LSTM). Datasets: UCSD, Avenue, UMN; Techniques: LSTM, RNN-based sparsity learning.
ISTL [52]	An Incremental Spatiotemporal Learner (ISTL) model is proposed to address anomaly detection and localization's difficulties and limitations to keep track of anomalies' changing character through active learning using fuzzy aggregation. Datasets: UCSD, Avenue; Techniques: convolutional LSTM, fuzzy aggregation.
Residual attention-based LSTM [53]	Using a lightweight CNN and an attention-based LSTM for anomaly detection reduces the time complexity with competitive accuracy. Datasets: Avenue, UCF-Crime, UMN; Techniques: residual attention-based LSTM.
CT-D2GAN [45]	A Conv-transformer is used to perform future frame prediction. Dual-discriminator adversarial training maintains local consistency and global coherence for future frame prediction. Datasets: UCSD Ped2, Avenue, ShanghaiTech; Techniques: GANs; transformer; CNN.
ViT-based framework [46]	Using a ViT model for anomaly detection involves processing a single frame as one patch. This approach yields good performance on the SVAD task while maintaining the advantages of the transformer architecture. Datasets: UCSD Ped2, Avenue, ShanghaiTech; Techniques: vision transformer.

3.5. Other Algorithms

Two clustering methods are available. Their argument is based on the idea that normal data are clustered, whereas anomalous data are not [54] connected to any cluster. The second type is predicated on the idea that, whereas anomalies belong to tiny clusters, typical data instances belong to massive or dense clusters. Fuzzy traffic density and flow are built using fuzzy theory to identify abnormalities in complicated traffic videos [55]. Heuristic techniques intuitively make decisions regarding anomalies based on feature values, geographical locations, and contextual data [56]. However, many real-world systems do not rely only on one technology. Using a lightweight CNN and an attention-based LSTM for anomaly detection reduces the time complexity with competitive accuracy.

3.6. Analysis of Algorithms

Statistics-based algorithms assume that normal behavior follows a certain statistical pattern, and any deviation from this pattern is considered an anomaly. They are simple and efficient and can detect real-time anomalies without requiring much training data. However, they may not be effective at detecting novel anomalies or anomalies that do not follow a statistical pattern.

Classification-based algorithms use machine learning techniques to classify behavior or events as normal or abnormal based on labeled training data. They can detect novel anomalies and adapt to changing environments with high accuracy. However, they require a large amount of training data, and the labeling process can be time-consuming and costly.

Reconstruction-based algorithms reconstruct normal behavior or events and compare them to the actual behavior or events to detect subtle anomalies. They do not require labeled training data, but they can be computationally expensive and unsuitable for real-time anomaly detection.

Prediction-based algorithms use machine learning techniques to predict future behavior or events based on past behavior or events. Any deviation from the predicted behavior or events is considered an anomaly. They can detect anomalies before they occur, which can be useful in preventing security threats or safety issues. However, they require a large amount of training data, and the accuracy of the predictions may decrease over time as the environment changes.

In conclusion, the selection of the algorithm depends on the specific application and requirements. Statistics-based algorithms are simple and efficient but may not detect novel anomalies. Classification-based algorithms have a high accuracy rate but require a large amount of training data. Reconstruction-based algorithms can detect subtle anomalies but can be computationally expensive. Prediction-based algorithms can detect anomalies before they occur but require a large amount of training data, and the accuracy of predictions may decrease over time. **Table 6** shows an overview of the algorithms.

Table 6. Overview of algorithms.

Algorithms	Strengths	Weaknesses
Statistics-based	Generally, they are suitable for real-time applications as they are simple and computationally efficient.	They cannot detect subtle or complex anomalies, such as changes in spatial or temporal relationships.
	Subtle or complex anomalies, such as those involving spatial or temporal relationship changes, cannot be detected.	False alarms may also occur when the data distribution deviates from a Gaussian distribution.
	High-dimensional datasets can be handled, and robustness to noise can be exhibited.	
Classification-based	Anomalies can be learned to be detected based on labeled training data, allowing them to adapt to changes in the data distribution over time.	Generally, they could be improved for real-time applications, being more computationally expensive than statistics-based algorithms.
	High-dimensional datasets can be handled, and global and local anomalies can be detected.	A large amount of labeled training data is also required, which can be difficult and time-consuming.
		Only the anomalies encountered in the training data can be detected, and new unseen anomalies cannot be detected.
Reconstruction-based	A compact representation of normal data can be learned, allowing subtle or complex anomalies to be detected.	They are generally less well-suited for real-time applications as they are more computationally expensive than statistics-based algorithms.
	High-dimensional datasets can be handled, and global and local anomalies can be detected.	A large amount of normal data for training is also required, which can be difficult to obtain in some scenarios.
		They are not robust to noise, and false alarms may occur when the data are noisy or corrupted.

Algorithms	Strengths	Weaknesses
Prediction-based	**Temporal dependencies in the data can be leveraged to detect anomalies, making them well-suited for time series data.** **High-dimensional datasets can be handled, and global and local anomalies can be detected.**	**They are generally less well-suited for real-time applications as they are more computationally expensive than statistics-based algorithms.** **A large amount of normal data for training is also required, which can be difficult to obtain in some scenarios.** **Anomalies involving spatial or temporal relationship changes may be difficult to detect.** **They do not possess robustness to noise, and false alarms can be produced when the data are noisy or corrupted.**

References

1. Alsulami, A.A.; Abu Al-Haija, Q.; Tayeb, A.; Alqahtani, A. An Intrusion Detection and Classification System for IoT Traffic with Improved Data Engineering. Appl. Sci. 2020, 12, 12336.
2. Nasteski, V. An overview of the supervised machine learning methods. Horizons B 2017, 4, 51–62.
3. Morente-Molinera, J.A.; Mezei, J.; Carlsson, C.; Herrera-Viedma, E. Improving supervised learning classification metho ds using multigranular linguistic modeling and fuzzy entropy. IEEE Trans. Fuzzy Syst. 2016, 25, 1078–1089.
4. Angarita-Zapata, J.S.; Masegosa, A.D.; Triguero, I. A taxonomy of traffic forecasting regression problems from a superv ised learning perspective. IEEE Access 2019, 7, 68185–68205.
5. Asad, M.; Yang, J.; He, J.; Shamsolmoali, P.; He, X. Multi-frame feature-fusion-based model for violence detection. Vis. Comput. 2021, 37, 1415–1431.
6. Wang, T.; Qiao, M.; Deng, Y.; Zhou, Y.; Wang, H.; Lyu, Q.; Snoussi, H. Abnormal event detection based on analysis of movement information of video sequence. Optik 2018, 152, 50–60.
7. Patil, N.; Biswas, P.K. Global abnormal events detection in surveillance video—A hierarchical approach. In Proceedings of the 2016 Sixth International Symposium on Embedded Computing and System Design (ISED), Patna, India, 15–17 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 217–222.
8. Kaltsa, V.; Briassouli, A.; Kompatsiaris, I.; Strintzis, M.G. Multiple Hierarchical Dirichlet Processes for anomaly detectio n in traffic. Comput. Vis. Image Underst. 2018, 169, 28–39.
9. Abu Al-Haija, Q.; Al Badawi, A. High-performance intrusion detection system for networked UAVs via deep learning. Ne ural Comput. Appl. 2022, 34, 10885–10900.
10. Sultani, W.; Chen, C.; Shah, M. Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE Confe rence on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6479–6488.
11. Alsulami, A.A.; Abu Al-Haija, Q.; Alqahtani, A.; Alsini, R. Symmetrical Simulation Scheme for Anomaly Detection in Auto nomous Vehicles Based on LSTM Model. Symmetry 2022, 14, 1450.
12. Huang, G.; Song, S.; Gupta, J.N.; Wu, C. Semi-supervised and unsupervised extreme learning machines. IEEE Trans. Cybern. 2014, 44, 2405–2417.
13. Chriki, A.; Touati, H.; Snoussi, H.; Kamoun, F. Uav-based surveillance system: An anomaly detection approach. In Proc eedings of the 2020 IEEE Symposium on Computers and Communications (ISCC), Rennes, France, 7–10 July 2020; I EEE: Piscataway, NJ, USA, 2020; pp. 1–6.
14. Chriki, A.; Touati, H.; Snoussi, H.; Kamoun, F. Deep learning and handcrafted features for one-class anomaly detection in UAV video. Multimed. Tools Appl. 2021, 80, 2599–2620.
15. Al-Qudah, M.; Ashi, Z.; Alnabhan, M.; Abu Al-Haija, Q. Effective One-Class Classifier Model for Memory Dump Malware Detection. J. Sens. Actuator Netw. 2022, 12, 5.
16. Wang, L.L.; Ngan, H.Y.; Yung, N.H. Automatic incident classification for large-scale traffic data by adaptive boosting SV M. Inf. Sci. 2018, 467, 59–73.
17. Ravanbakhsh, M.; Nabi, M.; Sangineto, E.; Marcenaro, L.; Regazzoni, C.; Sebe, N. Abnormal event detection in videos using generative adversarial nets. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICI P), Beijing, China, 17–20 September 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1577–1581.

18. Van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. Mach. Learn. 2020, 109, 373–440.

19. Bhakat, S.; Ramakrishnan, G. Anomaly detection in surveillance videos. In Proceedings of the ACM India Joint Internati onal Conference on Data Science and Management of Data, Kolkata, India, 3–5 January 2019; pp. 252–255.

20. Santhosh, K.K.; Dogra, D.P.; Roy, P.P. Anomaly detection in road traffic using visual surveillance: A survey. ACM Comp ut. Surv. (CSUR) 2020, 53, 1–26.

21. Hu, W.; Gao, J.; Li, B.; Wu, O.; Du, J.; Maybank, S. Anomaly detection using local kernel density estimation and context -based regression. IEEE Trans. Knowl. Data Eng. 2018, 32, 218–233.

22. Rüttgers, A.; Petrarolo, A. Local anomaly detection in hybrid rocket combustion tests. Exp. Fluids 2021, 62, 136.

23. Bansod, S.D.; Nandedkar, A.V. Crowd anomaly detection and localization using histogram of magnitude and momentu m. Vis. Comput. 2020, 36, 609–620.

24. Zhang, Y.; Lu, H.; Zhang, L.; Ruan, X. Combining motion and appearance cues for anomaly detection. Pattern Recogni t. 2016, 51, 443–452.

25. Sabokrou, M.; Fayyaz, M.; Fathy, M.; Moayed, Z.; Klette, R. Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. Comput. Vis. Image Underst. 2018, 172, 88–97.

26. Rahmani, M.; Atia, G.K. Coherence pursuit: Fast, simple, and robust principal component analysis. IEEE Trans. Signal Process. 2017, 65, 6260–6275.

27. Sarker, I.H. Machine learning: Algorithms, real-world applications and research directions. SN Comput. Sci. 2021, 2, 16 0.

28. Wang, T.; Snoussi, H. Detection of abnormal visual events via global optical flow orientation histogram. IEEE Trans. Inf. Forensics Secur. 2014, 9, 988–998.

29. Doshi, K.; Yilmaz, Y. Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate. Patter n Recognit. 2021, 114, 107865.

30. Aboah, A. A vision-based system for traffic anomaly detection using deep learning and decision trees. In Proceedings o f the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4 207–4212.

31. Saeedi, J.; Giusti, A. Anomaly Detection for Industrial Inspection using Convolutional Autoencoder and Deep Feature-b ased One-class Classification. In Proceedings of the VISIGRAPP (5: VISAPP), Online, 6–8 February 2022; pp. 85–96.

32. Chen, Y.; Tian, Y.; Pang, G.; Carneiro, G. Deep one-class classification via interpolated gaussian descriptor. In Proceed ings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2022; Volume 36, pp. 383–392.

33. Lee, K.; Lee, H.; Lee, K.; Shin, J. Training confidence-calibrated classifiers for detecting out-of-distribution samples. arX iv 2017, arXiv:1711.09325.

34. Liu, W.; Luo, W.; Lian, D.; Gao, S. Future frame prediction for anomaly detection–a new baseline. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6536–6 545.

35. Jiang, T.; Li, Y.; Xie, W.; Du, Q. Discriminative reconstruction constrained generative adversarial network for hyperspect ral anomaly detection. IEEE Trans. Geosci. Remote Sens. 2020, 58, 4666–4679.

36. Cheng, H.; Liu, X.; Wang, H.; Fang, Y.; Wang, M.; Zhao, X. SecureAD: A secure video anomaly detection framework on convolutional neural network in edge computing environment. IEEE Trans. Cloud Comput. 2020, 10, 1413–1427.

37. Zhao, Y.; Deng, B.; Shen, C.; Liu, Y.; Lu, H.; Hua, X.S. spatiotemporal autoencoder for video anomaly detection. In Proc eedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1 933–1941.

38. Xu, D.; Ricci, E.; Yan, Y.; Song, J.; Sebe, N. Learning deep representations of appearance and motion for anomalous e vent detection. arXiv 2015, arXiv:1510.01553.

39. Fan, Y.; Wen, G.; Li, D.; Qiu, S.; Levine, M.D.; Xiao, F. Video anomaly detection and localization via gaussian mixture fu lly convolutional variational autoencoder. Comput. Vis. Image Underst. 2020, 195, 102920.

40. Duman, E.; Erdem, O.A. Anomaly detection in videos using optical flow and convolutional autoencoder. IEEE Access 2 019, 7, 183914–183923.

41. Madan, N.; Farkhondeh, A.; Nasrollahi, K.; Escalera, S.; Moeslund, T.B. Temporal cues from socially unacceptable traje ctories for anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2150–2158.

42. Song, H.; Sun, C.; Wu, X.; Chen, M.; Jia, Y. Learning normal patterns via adversarial attention-based autoencoder for abnormal event detection in videos. IEEE Trans. Multimed. 2019, 22, 2138–2148.

43. Sun, C.; Jia, Y.; Song, H.; Wu, Y. Adversarial 3d convolutional auto-encoder for abnormal event detection in videos. IEEE Trans. Multimed. 2020, 23, 3292–3305.

44. Nguyen, T.N.; Meunier, J. Anomaly detection in video sequence with appearance-motion correspondence. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1273–1283.

45. Feng, X.; Song, D.; Chen, Y.; Chen, Z.; Ni, J.; Chen, H. Convolutional transformer based dual discriminator generative adversarial networks for video anomaly detection. In Proceedings of the 29th ACM International Conference on Multimedia, Nice, France, 21–25 October 2021; pp. 5546–5554.

46. Lee, J.; Nam, W.J.; Lee, S.W. Multi-Contextual Predictions with Vision Transformer for Video Anomaly Detection. In Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 21–25 August 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1012–1018.

47. Yuan, H.; Cai, Z.; Zhou, H.; Wang, Y.; Chen, X. Transanomaly: Video anomaly detection using video vision transformer. IEEE Access 2021, 9, 123977–123986.

48. Ullah, A.; Ahmad, J.; Muhammad, K.; Sajjad, M.; Baik, S.W. Action recognition in video sequences using deep bi-directional LSTM with CNN features. IEEE Access 2017, 6, 1155–1166.

49. Ergen, T.; Kozat, S.S. Unsupervised anomaly detection with LSTM neural networks. IEEE Trans. Neural Netw. Learn. Syst. 2019, 31, 3127–3141.

50. Ristea, N.C.; Madan, N.; Ionescu, R.T.; Nasrollahi, K.; Khan, F.S.; Moeslund, T.B.; Shah, M. Self-supervised predictive convolutional attentive block for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 13576–13586.

51. Zhou, J.T.; Du, J.; Zhu, H.; Peng, X.; Liu, Y.; Goh, R.S.M. Anomalynet: An anomaly detection network for video surveillance. IEEE Trans. Inf. Forensics Secur. 2019, 14, 2537–2550.

52. Nawaratne, R.; Alahakoon, D.; De Silva, D.; Yu, X. Spatiotemporal anomaly detection using deep learning for real-time video surveillance. IEEE Trans. Ind. Inform. 2019, 16, 393–402.

53. Ullah, W.; Ullah, A.; Hussain, T.; Khan, Z.A.; Baik, S.W. An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos. Sensors 2021, 21, 2811.

54. Ranjith, R.; Athanesious, J.J.; Vaidehi, V. Anomaly detection using DBSCAN clustering technique for traffic video surveillance. In Proceedings of the 2015 Seventh International Conference on Advanced Computing (ICoAC), Chennai, India, 15–17 December 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–6.

55. Li, Y.; Guo, T.; Xia, R.; Xie, W. Road traffic anomaly detection based on fuzzy theory. IEEE Access 2018, 6, 40281–40288.

56. Chang, M.C.; Wei, Y.; Song, N.; Lyu, S. Video analytics in smart transportation for the AIC'18 challenge. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 61–68.