A Systematic Review of Rare Events Detection Across Modalities using Machine Learning and Deep Learning

Yahaya Idris Abubakar[1], Alice Othmani[2], Patrick Siarry[2], and Aznul Qalid Md. Sabri[2]

[1]Université Paris-Est Créteil
[2]Affiliation not available

April 08, 2024

Abstract

Rare event detection (RED) involves the identification and detection of events characterized by low frequency of occurrences, but of high importance or impact. This paper presents a Systematic Review (SR) of rare event detection across various modalities using Machine Learning (ML) and Deep Learning (DL) techniques. This review comprehensively outlines techniques and methods best suited for rare event detection across various modalities, while also highlighting future research prospects. To the extent of our knowledge, this paper is a pioneering SR dedicated to exploring this specific research domain. This SR identifies the employed methods and techniques, the datasets utilized, and the effectiveness of these methods in detecting rare events. Four modalities concerning RED are reviewed in this SR: video, sound, image, and time series. The corresponding performances for the different ML and DL techniques for RED are discussed comprehensively, together with the associated RED challenges and limitations as well as the directions for future research are highlighted. This SR aims to offer a comprehensive overview of the existing methods in RED, serving as a valuable resource for researchers and practitioners working in the respective field.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 February 2024, accepted 20 March 2024, date of publication 27 March 2024, date of current version 4 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3382140

SURVEY

A Systematic Review of Rare Events Detection Across Modalities Using Machine Learning and Deep Learning

YAHAYA IDRIS ABUBAKAR[1], ALICE OTHMANI[1], PATRICK SIARRY[1], AND AZNUL QALID MD SABRI[2]

[1]Laboratoire Images, Signaux et Systémes Intelligents (LiSSi)–EA 3956, Université Paris-Est Créteil (UPEC), 94010 Créteil Cedex, France
[2]Faculty of Computer Science & Information Technology, Universiti Malaya, Kuala Lumpur 50603, Malaysia

Corresponding author: Alice Othmani (alice.othmani@u-pec.fr)

This work was supported by the Petroleum Technology Development Fund (PTDF) Abuja, Nigeria, Grant 2059/22.

ABSTRACT Rare event detection (RED) involves the identification and detection of events characterized by low frequency of occurrences, but of high importance or impact. This paper presents a Systematic Review (SR) of rare event detection across various modalities using Machine Learning (ML) and Deep Learning (DL) techniques. This review comprehensively outlines techniques and methods best suited for rare event detection across various modalities, while also highlighting future research prospects. To the extent of our knowledge, this paper is a pioneering SR dedicated to exploring this specific research domain. This SR identifies the employed methods and techniques, the datasets utilized, and the effectiveness of these methods in detecting rare events. Four modalities concerning RED are reviewed in this SR: video, sound, image, and time series. The corresponding performances for the different ML and DL techniques for RED are discussed comprehensively, together with the associated RED challenges and limitations as well as the directions for future research are highlighted. This SR aims to offer a comprehensive overview of the existing methods in RED, serving as a valuable resource for researchers and practitioners working in the respective field.

INDEX TERMS Artificial intelligence, deep learning, detection, machine learning, rare event detection.

I. INTRODUCTION

Rare Event Detection (RED) refers to the task of identifying and detecting events that have a low frequency of occurrences but of high importance or impact [1]. Due to its infrequency, rare events pose significant challenges for detection, prediction, and classification tasks [2].

In the context of deep learning, rare event detection involves developing techniques that can effectively identify and classify these rare events using deep learning models [2]. The goal is to achieve high accuracy in detecting these rare events, even when there is limited labeled data available for training the models [3].

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

Rare events are events that are exceptionally infrequent and have significant consequences. They are difficult to predict, but they are expected to occur eventually [4]. Some examples of rare events as highlighted by Ivanov et al. [5] include a person walking in a forbidden zone, a vehicle driving on the wrong side of the road, and a person running in an area where people are expected to walk [5]. Narayanan et al. [6] provided a similar definition, defining rare events as events having low probabilities of occurrences with high impact.

Rare events can also be defined as events that occur infrequently, displaying transient characteristics with unpredictable occurrences, and no prior indication of their completion once they have transpired [7].

Recently, it has been widely acknowledged that rare event detection is a prominent area of research [8] within the

domain of event detection. This area of research primarily focuses on the environment, health, and various industrial applications dealing with digital signals.

However, it is worth noting that the majority of research works concentrate on the realm of anomaly and outlier detection [8]. According to [9], the identification of different types of anomalous events can be challenging due to diverse patterns and typical incidents within distinct scenes. Anomalies are defined as events that deviate from normal patterns, which may not always result in accuracy or effectiveness [10]. Anomaly detection involves the process of identifying data patterns that exhibit behaviors significantly different from the expected norm [11]. In review [12], Zideh et al. examine the use of Physics-Informed Machine Learning (PIML) in power systems for detecting, classifying, localizing, and mitigating anomalies. They address challenges in smart grid data usage, the role of ML in supporting control room decision-making, and the integration of system physics into ML models.

Detecting rare events is the process of identifying patterns that occur much less frequently. It involves analyzing events that occur infrequently or irregularly in a given dataset or system [1]. If an observation occurs infrequently and deviates significantly from other observations, it may indicate an irregularity or anomaly in the given set of observations [8], [13], [14].

Furthermore, when determining a rare event, it is based on a combination of factors including the rarity of the event, its unique characteristics, and the use of both positive and negative evidence for detection.

RED is incredibly significant in various applications. Some examples include RED as part of face detection systems for security purposes [15], RED based fraud detection systems used in financial transactions, RED in video and image retrieval, and using RED within the context of epidemiology for disease outbreak detection [1]. Developing efficient RED systems can positively impact the accuracy and effectiveness of these respective applications [1], [14].

RED can be used to create surveillance systems that detect and alert users or authorities of potential disturbances or threats in real-world environments [8]. According to Sokolova et al. [4], these systems have the ability to identify and predict valuable events that occur with extreme rarity, making them difficult to predict. RED systems have the potential to quickly detect critical events close to their source, allowing necessary actions to be taken in a timely manner [13].

Failure to detect rare events can result in serious consequences such as system failures, security breaches, or medical errors [14]. Automatic detection of rare events is challenging due to pattern variations in different scenes. In fact, rare events vary based on the types of applications and scenarios.

A. MAIN CONTRIBUTION OF THE SYSTEMATIC REVIEW

The contributions of this review can be summarized below.

- A comprehensive and systematic review of RED using both machine and deep learning techniques is presented to provide a state-of-the-art review of RED across various modalities. These modalities include time series, videos, sounds, and images. As presented in Table 1, this review addresses the scarcity of comprehensive systematic reviews and surveys in the field of RED across multiple modalities.
- A discussion of the different processes involved in RED is presented. The different processes in RED are elaborated in detail for each type of modality. This provides a clear overview of the required approach to achieve RED for the different modalities.
- The review systematically identifies challenges, gaps, and limitations within the field of RED, with a clear indication of areas in RED that require further research and development, allowing an enhanced understanding of the existing limitations within the field.

B. REVIEW ORGANISATION

This review is structured as follows: Section II outlines the search strategy and eligibility criteria for rare event detection, detailing the sources as well as the inclusion and exclusion criteria. Sections III, IV, V, and VI present separately the processes of preprocessing, low-level feature extraction, high-level feature extraction, classification and the available datasets for video, sound, image, and time series modalities, respectively.

Section VII provides the comparative analysis of the performance of different approaches across different modalities. Section VIII discusses and highlights the review's findings.

Finally, Section IX concludes the review and provides recommendations for future research.

II. SEARCH STRATEGY AND ELIGIBILITY CRITERIA

In this review, we analyzed articles published since 2000 that primarily focus on detecting rare events using ML and DL techniques as shown in Figure 9, with several relevant articles related to anomaly and outlier detection. We conducted a comprehensive literature search across several databases within AI, ML, DL, and computer science domains.

Other sources, including reference lists, archives, bibliographies, and other materials that met the systematic review criteria, were considered. Our search spanned articles, journals, conference proceedings, and theses/dissertations from reputable scientific publishers such as IEEE Xplore, Elsevier, Scopus, JMLR, PubMed, Springer, ACM Digital Library, BASE, Google Scholar, ResearchGate, ArXiv, and Litmaps. The search keywords used in this study included ''rare event detection'', ''artificial intelligence'', ''machine learning'', ''deep learning'', ''video rare events'', ''sound rare events'', ''image rare events'', ''1D signal rare events'', and ''time series rare events''.

The inclusion criteria, as illustrated in Figure 2, for selecting pertinent studies were: (i) publication in peer-reviewed journals or conference proceedings; (ii) a focus on the application of artificial intelligence for rare events; and (iii) the use of machine learning or deep learning techniques for

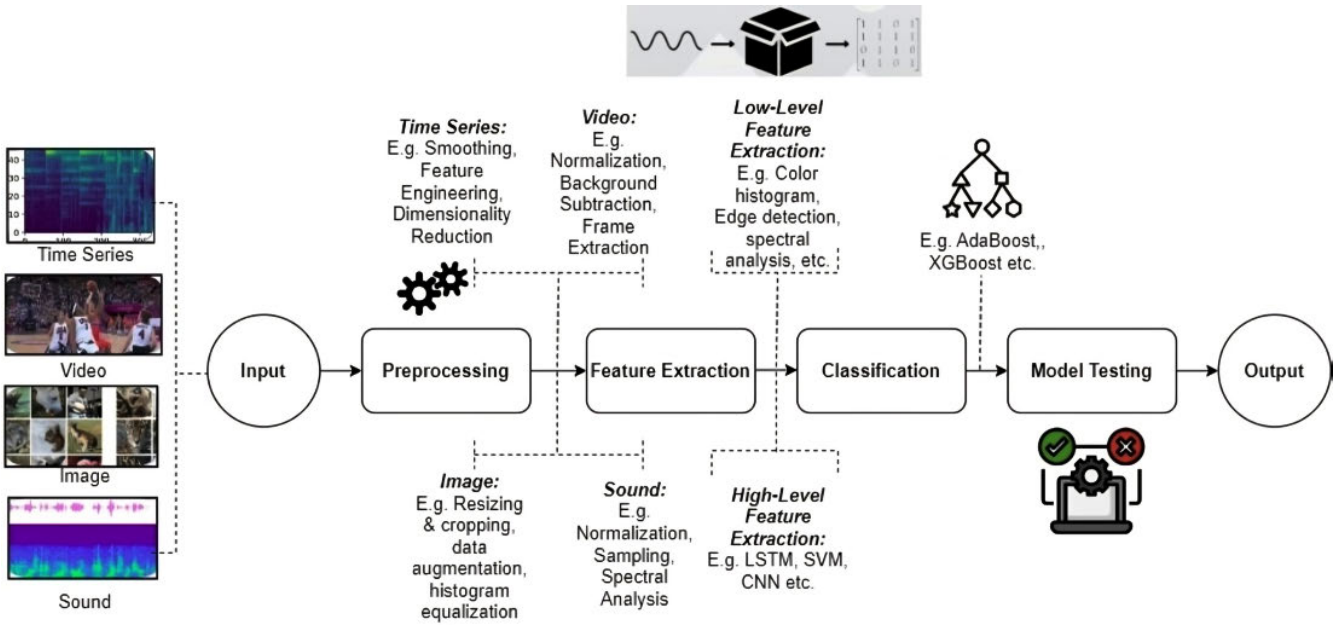

FIGURE 1. The figure shows a typical ML pipeline for RED, which includes preprocessing, high & low levels feature extraction, classification and model testing.

TABLE 1. Overview of key reviews related to rare event detection, detailing the reference, publication year, modality or area covered, and specific remarks. This table provides a comparative snapshot, highlighting the scope and coverage of each study. This review focuses on multiple modalities and comprises of 28 articles.

Reference	Year	Modality/Area Covered	Remark
Maalouf et al. [16]	2015	Imbalanced dataset	General imbalanced dataset
Liu et al. [17]	2019	Meta-analysis on rare events	List of articles were not provided
Zhou et al. [18]	2020	Existing protocols on SR	1,004 articles
Jia et al. [19]	2021	Meta-analysis on RE on Cochrane database of SR	4,177 articles
Radulescu et al. [20]	2021	Case reports on RE	74 articles
Current Review	2023	Various Modalities (Video, Sound, Image & Time series)	28 articles

rare events in video, sound, image, and time series modalities. Studies without abstracts, inaccessible full papers, non-English publications, or those published before 2000 were excluded.

A total of 28 research articles meeting our inclusion criteria were identified and subsequently analyzed. Figure 3 outlines the PRISMA steps followed in article selection. The identified articles were further evaluated based on the inclusion and exclusion criteria detailed in Figure 2. This survey follows closely the PRISMA guideline [22] for quality assessment, performance evaluation of results, and reporting. The break-down of the percentage of the articles reviewed in this article can be found from Figure 4.

The ML pipeline adopted in this review is summarized in Figure 1. It encompasses preprocessing, high-level feature extraction, low-level feature extraction, and classification for all the discussed approaches. Each of these blocks will be presented in detail in the following sections.

III. VIDEO BASED APPROACHES FOR RARE EVENT DETECTION

Video-based rare event detection involves the identification of uncommon or abnormal events within video sequences. This issue poses a formidable obstacle in the realm of computer vision and video processing. With the increasing popularity of video recording and sharing, there is a need for solutions that can analyze video content in a robust and scalable manner [23].

A. PREPROCESSING

Preprocessing plays a pivotal role in video rare event detection. Some of the preprocessing approaches include dimensional reductions of captured videos, selection of distinctive features from the video frames, and performing background subtraction on the captured videos.

In the study by Sharma et al., specific features are chosen from each frame to delineate pixels, and background subtraction is executed using the Gaussian mixture procedure [24].

The incorporation of binary values to mitigate the influence of noise and other sources of uncertainty in data is introduced in [25]. Spatial relations are encoded within a binarized feature vector representation, and temporal constraints in events are articulated using the Hidden Markov Model (HMM) framework. This approach facilitates the modeling of semantic primitives' dynamics, enabling the detection and recognition of rare events in video frames.

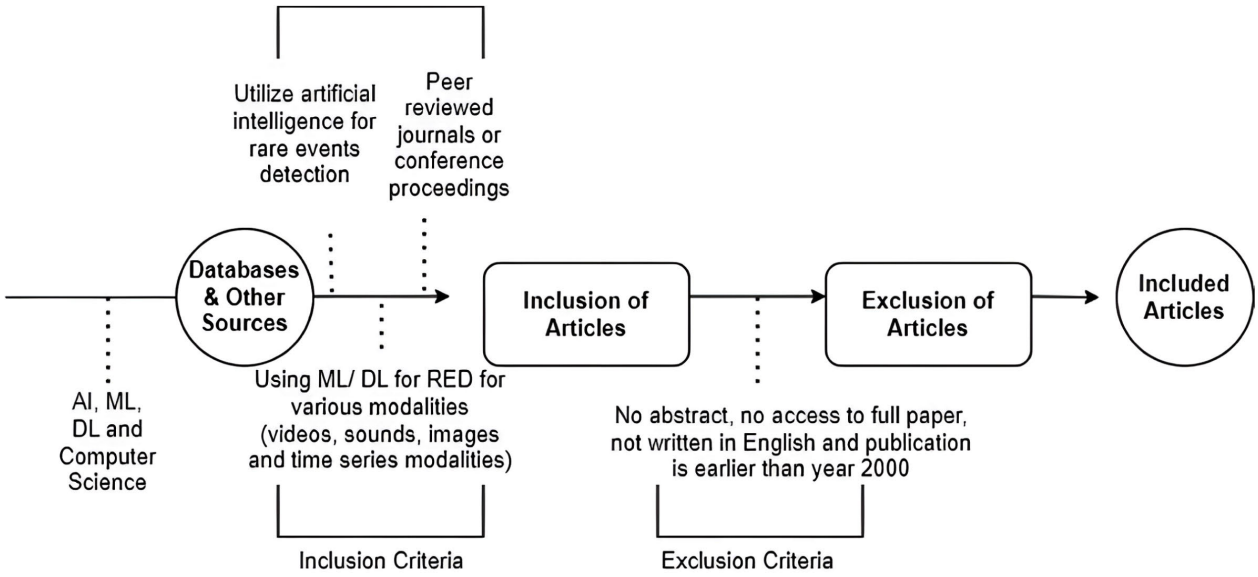

FIGURE 2. Illustrates the inclusion criteria for selecting pertinent studies, emphasizing publications in peer-reviewed platforms, the use of artificial intelligence in rare events, and the application of machine or deep learning techniques across video, sound, image, and time series modalities. Exclusions were made for studies not meeting these criteria, those lacking abstracts, inaccessible full papers, non-English publications, and those published before 2000.

In the work of Ma et al. [26], as depicted in Figure 5, a set of preprocessing techniques is used to improve the rate-distortion performance of a video. The goal of their work is to preserve the quality of the video while at the same time reducing the bitrate usage of the video.

In reference to the preprocessing step for videos, with regards to RED, most of the works are focused on making use of the spatial correlation between the video frames within the videos (e.g., background subtraction). This is done to achieve, among others, a reduction in size or removal of noises with the goal of enhancing the final accuracy and efficiency of the detection process [9], [24].

B. LOW-LEVEL FEATURES EXTRACTION

Low-level feature extraction involves generating basic features and cross-scale path embedding to enhance fine-grained details in video frame interpolation [27]. These encompass color features (such as histograms, color moments, and color spaces like RGB, HSV, & LAB), temporal features (such as texture and STFT), shape features (such as object trajectory and silhouette), and motion features (such as vectors, frame differencing, and optical flow) [28].

Aljaloud et al. [9] and Ullah et al. [29] highlighted hand-crafted features such as trajectory, flow, and vision modeling. These features encode spatiotemporal information based on color, texture, optical flow, and bag-of-words features. Such handcrafted features are fundamental components in existing models for detecting unusual events.

In their study, Sharma et al. employed trajectory-based trackers to identify movements. Shape-based motion detection was determined by extracting relevant features. They utilized a spatial-temporal method to extract features from images, with a focus on critical elements [24].

In the work by Kwon et al. [30], a video is graphically represented using nodes to denote segmented events. Node edges describe related events, with edge weights reflecting the degree of the relationship. The graph is optimized using the Data-Driven Markov Chain Monte Carlo technique. This optimization reduces energy consumption by merging subgraphs or pruning edges. The energy model incorporates parameters representing events' causality, frequency, and significance. Specific models were designed for event summarization and rare-event detection.

Extracted low-level features discussed in [23], such as Scale Invariant Features Transform (SIFT) [31], Space-Time Interest Points (STIP) [32], and OpponentSIFT [33], are inadequate for comprehending semantic nuances in complex situations.

Binary features vectors derived from video training enable HMMs to recognize events using a single exemplar. Thresholds are established based on object sizes and calibration data. For visual features extraction in a 3D scene, a color mask serves as a features list filter [25].

It is observed that the majority of the approaches that utilized low-level features extraction for RED, focus on utilizing spatio-temporal information from the video frames in order to potentially use it to characterize the unusual events within the videos.

C. HIGH-LEVEL FEATURES EXTRACTION

High-level features extraction in video RED involves abstracted representations, that are used to represent infrequent or unusual events. This process captures complex

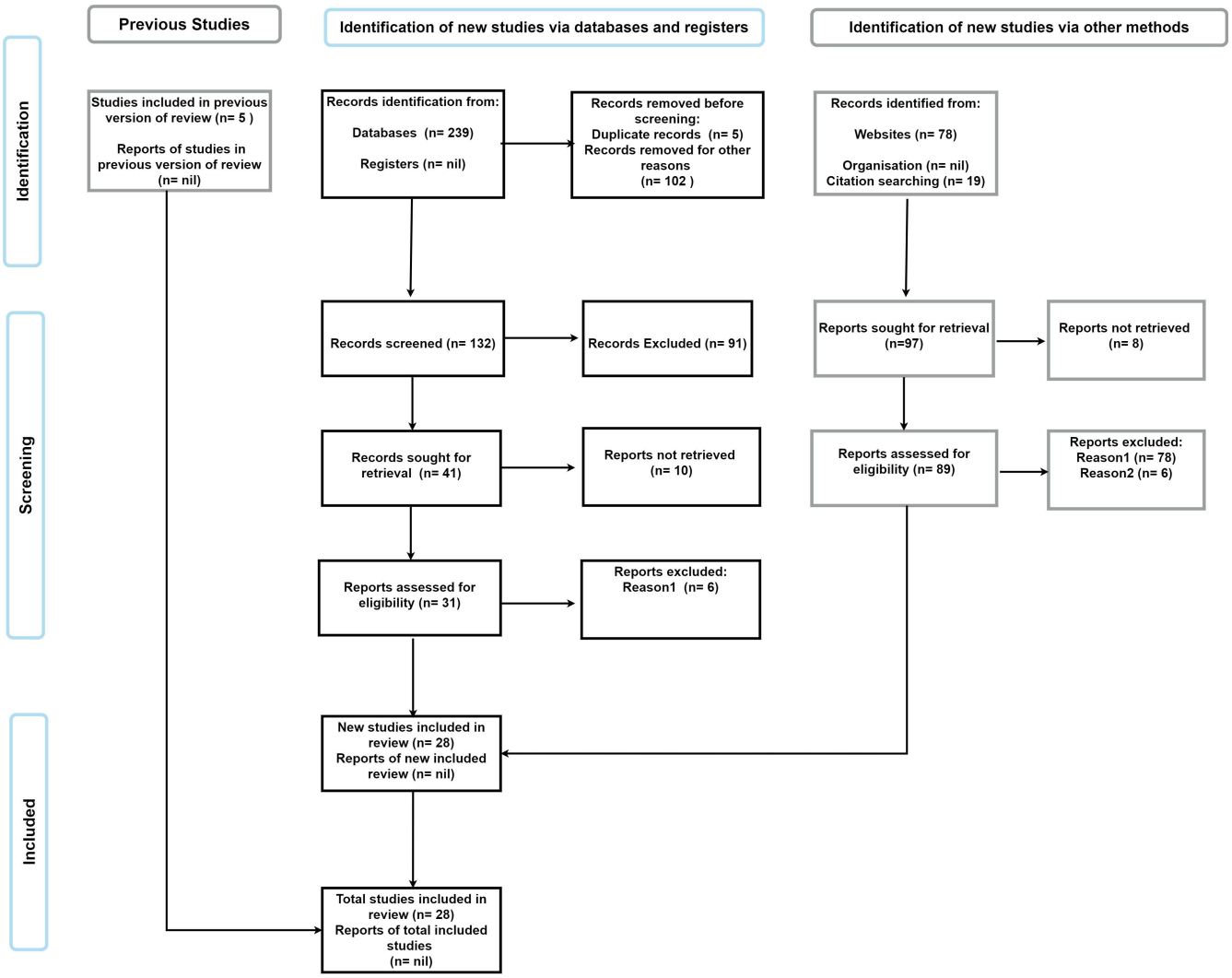

FIGURE 3. Flowchart adopted from [21] illustrating the PRISMA-based selection process for the 28 research articles that met our inclusion criteria, in alignment with the PRISMA guidelines [22].

patterns, behaviors, and contextual information. It relies on ML and DL models that are capable of analyzing vast amounts of data to recognize patterns [11].

The use of Bayesian Deep learning by [30] employs the Data-Driven Markov Chain Monte Carlo (DDMCMC) approach for event summarization by reducing the energy model. Multiple graph samples were generated, from which the best was selected. This was then used for event synthesis or rare event recognition.

Convolutional Neural Network (CNN) have demonstrated its effectiveness and yielded good results when applied to event detection tasks [34]. CNNs are also used for video analysis tasks such as for action recognition and event detection. CNNs excel at extracting high-level features from video data [35]. This features extraction method captures meaningful data with semantic information. In another approach, Luisier et al. [34] used deep convolutional neural network features that were trained on the ImageNet dataset. These features contain both mid-level and semantic information.

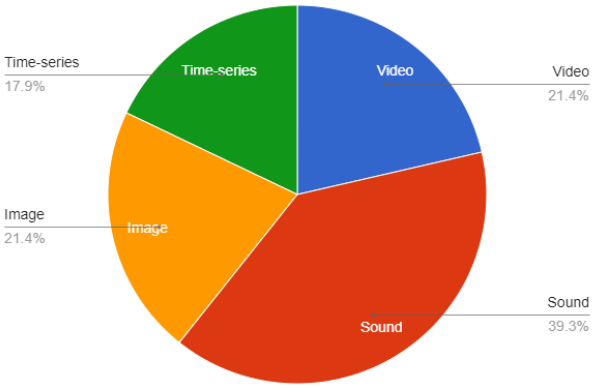

FIGURE 4. Distribution of research articles by modality, showing the proportion of articles focused on video, sound, image, and time Series out of a total of 28 articles.

CNNs were proposed by [35] for detecting events in videos. Bansod et al. [36] utilized 3D CNNs to learn deep representations of appearance and motion for anomalous event detection. They computed 3D gradient features in the

FIGURE 5. Preprocessing in video showing the deployment workflow of the rate-perception optimized preprocessor (RPP) and a comparison of frame segments of H.265 and RPP + H.265 at the same MS-SSIM, [26].

horizontal and vertical directions to represent appearance and motion features. These features were then characterized along the temporal direction to represent video events. The computed 3D gradient features were obtained by convolving the video frames using 3D Sobel filters.

It is evident that for videos, RED approaches that utilize high-level features extraction depended on ML or DL approaches to capture a higher magnitude of information that is beyond spatial and temporal. Instead, semantic information was successfully captured that proved to be useful for RED.

D. CLASSIFICATION

Classification in video-based rare event detection involves categorizing video segments/frames into different classes, with one class representing the rare event. This process entails training a model to accurately classify rare events in the video data [35].

The methodologies used for image classification are often modified and expanded to effectively work on video datasets [37].

In an example of a work [5] focusing on video classification for RED, trajectories of moving objects are categorized as normal or abnormal based on their high-level features.

In [24], both CNN and Support Vector Machine (SVM) were used for classification to perform video surveillance. This approach involves extracting image features, reducing time complexity, and utilizing a fused classification framework to accurately detect abnormal events in video frames.

Ye et al. [23] discusses various classification techniques that can be applied to video-based event detection. These techniques include one-versus-all classification based on various features representations using a two-class SVM, as well as graphical models like the transition Hidden Markov Model (HMM) and Conditional Random Fields (CRFs) for analyzing sequential video frames.

Bansod et al. [36] employs trackers and Restricted Boltzmann Machines (RBM) for features representation. They then used Support Vector Machines (SVM) for classification.

Madan et al. [38] presented an interesting anomaly detection approach that is achieved by utilizing the magnitude of the reconstruction error as an indicator for the abnormality level. This indicator is then used to classify and detect anomalies in videos. Their approach is depicted in Figure 6.

In [25], the authors trained a Hidden Markov Model (HMM) using binary features extracted from semantic

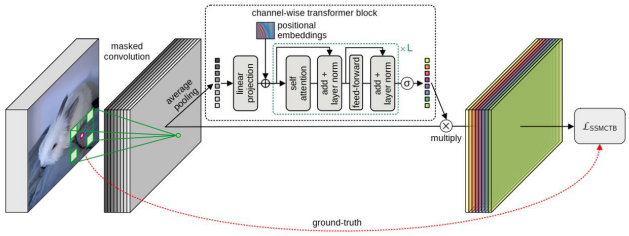

FIGURE 6. An overview of self-supervised masked convolutional transformer block (SSMCTB). At every location where the masked filters are applied, the proposed block has to rely on the visible regions (sub-kernels) to reconstruct the masked region (center area). A transformer module performs channel-wise self-attention to selectively promote or suppress reconstruction maps via a set of weights returned by a sigmoid layer. The block is self-supervised via the mean squared error loss (LSSMCTB) between masked and returned activation maps [38].

primitives. Their approach incorporates multiple objects and spatio-temporal dynamics. They differentiate true events from incidental activities by comparing the likelihood scores to a threshold. This approach outperforms direct continuous observable approaches and is effective at detecting rare events even with limited training data.

In this section, it is observed that the classification of RED in videos makes use of a variety of techniques and is not dependent on direct classification approaches. Instead, several interesting classification methods such as in [25] and [38], made use of the magnitude of reconstruction errors of videos, as well as usage of likelihood scores of true events against incidental activities. It is rather clear that RED in videos is a complex task that requires significant effort in designing the classification approach to achieve optimal performance.

E. DATASETS FOR VIDEO BASED APPROACHES FOR RARE EVENT DETECTION

Table 2 contains information on various video datasets and the associated performance metrics for RED. Various RED-specific video datasets are available, with various anomalous or rare events types of scenarios. There are different datasets focusing on various scenarios that are anomalous. This includes datasets recording a scenario where a person is approaching a train in motion or another dataset that records videos containing violence outbreaks. This proves that there is significant research attention with regard to RED in videos.

It is also important to note that the performance metrics recorded in Table 2 are the metrics achieved by the different

works reviewed in this systematic review. For some of the datasets, to the extent of our knowledge, there were no performance analysis recorded.

Researchers can make use of these datasets for performance metrics benchmarking and also for developing rare event detection algorithms across different scenarios. Among the datasets with available performance metrics is the **UCSDPed2 Dataset** [39]. For this dataset, the highest accuracy recorded is **99.44%** [36]. This dataset contains anomalies associated with running, walking, and biking.

Although some datasets lack performance metrics, it is observed that the UCF-101 [37] dataset is rather challenging since the accuracy achieved is slightly over 80%. Another challenging dataset is the Sport-1M dataset [23] which contains more than a million videos with 487 labels.

IV. SOUND BASED APPROACHES FOR RARE EVENT DETECTION

Sounds are auditory sensations created by vibrations that travel through a medium, such as air, and are perceived by our ears. The automatic detection of environmental sound events has recently gained attention [53]. Unlike speech and music, environmental sounds lack stationary patterns. Chen et al. [54] described sound event detection (SED) as a technology enabling devices to comprehend their surroundings by identifying multiple target sound events that can occur concurrently.

There are many similarities between sound and 1D signals when it comes to detecting rare events. These similarities include how data is represented, how features are extracted, and how machine learning techniques are used. By recognizing these similarities, we can apply proven methods and algorithms from one domain to the other. This approach leads to more effective and efficient rare event detection in both sound and 1D signal data.

The process of detecting rare events using 1D signal-centric techniques involves carefully examining 1D signals. This approach has applications in various fields, including healthcare (for ECG and EEG) [55], [56], finance (for stock analysis) [57], telecommunications (for network monitoring) [58], and environmental monitoring (for seismic detection) [59]. Specialized algorithms are employed to accurately identify and classify these rare events within the signal.

A. PREPROCESSING

Preprocessing sound for rare event detection involves enhancing sound quality and is necessary to reduce variability in the acoustic characteristics of rare events. It is also needed in order to enhance the performance of ML models for RED [8].

Surampudi et al. used a low-pass Butterworth filter in their extraction approaches to preprocess audio signals for detecting sound events. Audio signals are filtered with a 1500 Hz low-pass Butterworth filter before features extraction. They demonstrated the addition of features from different signals improves the performance of the learning algorithms [8].

In 1D signal preprocessing, raw signal data is prepared and enhanced for detection algorithms, ensuring improved data quality by removing noise and artifacts while identifying infrequent occurrences. The Time Window Slicing (TWS) function trims time series data and isolates anomaly incidents, increasing the sample substance before widening the augmentation possibilities across domains. Once the TWS process is complete, anomaly incident seeds are transformed using upsampling-downsampling, fast Fourier transform, and time series decomposition [60].

In their submission, Torres et al. [61] referred to preprocessing in signals as the techniques used to remove noise and artifacts from signals.

It can be observed that for RED in sound, most approaches employ sophisticated digital signal processing approaches to enhance the sound quality in order to improve the overall performance of their approaches.

B. LOW-LEVEL FEATURES EXTRACTION

Low-level feature extraction is crucial for sound-based rare event detection, as it involves extracting basic acoustic characteristics from sound data. This process captures the acoustic properties of rare events and provides discriminative information for ML models [8].

Spectrogram computation and features extraction methods (e.g., MFCCs or log-mel spectrograms) are commonly used in sound event detection [53]. In [62], Mel log energy (MLE) features are derived using the fast Fourier transform (FFT) to discern distinct frequencies in an audio signal. Recognizing these unique frequency bands is important in the detection of sound events, as each sound is characterized by its particular frequency spectrum.

Short-Time Fourier Transform (STFT) coefficients serve as low-level features for identifying infrequent sound occurrences [54]. STFT coefficients contain important information from the initial audio and enhance detection accuracy. This process involves extracting basic signal-level features from an audio waveform that represents its acoustic properties [63]. These features capture sound signal properties for analysis and classification.

In [3], Hyungui et al. extracted log-amplitude mel-spectrograms as the input acoustic features. These spectrograms are 2D time-frequency representations of sound signals used in signal processing. They represent energy distribution across different frequency bands, with higher resolution in the lower frequency range. The logarithmic compression enhances the perceptual relevance of features.

Wang et al. [64] presented a model that integrates both utterance-level and frame-level losses to categorize event instances and pinpoint their time boundaries. The utterance-level loss determines the presence of the event in the sound, while the frame-level loss pinpoints the specific frames related to the event. Both types of losses employ

TABLE 2. Summary of datasets used in video-based approach, detailing dataset references, benchmark performance, sample sizes, events detected, and repository locations.

Ref.	Name of Dataset	Benchmark Performance	No. of Sample	Event Detected	Repository / Location
[9]	Violent-Flows Dataset [40]	ACC = 92.5%	246 real-world videos of crowd violence (123 violence x 123 non-violence)	Violence outbreak detection	https://paperswithcode.com/dataset/violent-flows
[9]	UMN Dataset [41]	ACC = 91.5%	11 sequences x 3849 training x 3872 testing frames	Detect anomalous events, such as a person running in the wrong direction, vehicle driving on the sidewalk in videos captured by surveillance cameras	http://mha.cs.umn.edu/Movies/
[9]	Web Dataset [42]	For performance evaluation	12 normal crowd x 8 abnormal crowd	Escape panics, protesters clashing & crowd fighting	https://github.com/webdataset/webdataset
[23]	Sport-1M [43]	No performance provided	1,133,158 videos x 487 labels	Human actions recognition in videos such as basketball, soccer, tennis, and gymnastics	https://www.kaggle.com/datasets/sabahesaraki/sports-1m-dataset
[23]	CCV dataset [44]	No performance provided	9, 317 youtube videos x 20 semantics	human activities	https://www.ee.columbia.edu/ln/dvmm/CCV/
[23]	FCVID [45]	No performance provided	91, 223 web videos x 239 categories	social events,objects	https://fvl.fudan.edu.cn/dataset/fcvid/list.htm
[23]	Columbia EventNet Dataset [46]	No performance provided	95, 321 videos x 4490 events	NA	https://www.ee.columbia.edu/ln/dvmm/newDownloads.htm
[30]	BOSS [47]	No specific numerical values or detailed performance analysis provided	10 video sequences x 1 story-line x multiple events	Event summarization and rare event detection based on approaching train in motion	http://velastin.dynu.com/videodatasets/BOSSdata/
[30]	London Traffic [48]	No specific numerical values or detailed performance analysis provided	827 nodes x 2353 edges	Approaching train in motion	https://data.london.gov.uk/dataset/traffic-flows-borough
[30]	Subway Platform Sequence [49]	No specific numerical values or detailed performance analysis provided	1846 rows (subway entrance) x 32 columns (attribute of the entrances).	Wrong direction, no payment, and loitering	https://data.beta.nyc/en/dataset/subway-station-entrances
[35]	Swimming data [50]	F1-Score = 0.967	15,000 labeled strokes x 650,000 frames at 50fps	Continuous video to simple signals for swimming stroke detection with Convolutional Neural Networks	https://data.world/datasets/swimming
[35]	Tennis Data [51]	F1-Score = 0.977	1,300 labeled strokes x 270,000 frames at 30fps	Non-anomaly of strokes in the wild, freestyle and breast stroke	https://datahub.io/sports-data/atp-world-tour-tennis-data
[36]	UCSDPed1 [39]	ACC = 96.75%	34 training x 36 testing sequences (200 video frames each) x 6700 training and 7200 testing frames & (158x238) frame dim.	cycle, skater, truck, car, wheelchair & baby cart	http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm
[36]	UCSDPed2 [39]	ACC = 99.44%	16 training x 12 testing sequences (120-180 frame size) x 2550 training x 2010 testing frames & (240x320) frame dim.	running walking & biking	http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm
[36]	UMN Dataset [41]	ACC = 98.07%	11 sequences x 3849 training x 3872 testing frames	Abnormal crowd activity, sudden running of people	http://mha.cs.umn.edu/Movies/
[37]	UCF-101 [52]	ACC = 81.5%	13320 videos x 101 Categories	Human actions recognition in videos such as household activities, and animal interactions	https://www.crcv.ucf.edu/data/UCF101.php

a common vector representation and are interconnected through an attention mechanism, highlighting the model's efficiency in detecting rare sound events.

Venkatesh et al. [65] cautioned against the common use of spectrogram-based features in sound event detection tasks. These features involve transforming the audio signal into a time-frequency representation using techniques like short-time Fourier transform (STFT) or Mel-frequency cepstral coefficients (MFCCs). Mesaros et al. [63] illustrated how MFCCs are calculated using 40 ms frames with a Hamming window, 50

Log Filter Bank Energies (LFBEs) is a features extraction technique in sound and speech processing that captures spectral characteristics of sound signals [64]. These features describe key traits of sound signals in artificial intelligence (AI) analysis. Extracting features from 44.1 kHz mono audio signals, the authors extracted 64-dimensional LFBEs from 46 ms frames every 23 ms for 30-second audio clips.

The utilization of Finite Impulse Response (FIR) filters on input seismograms to extract effective signals for discriminating events was emphasized in [59]. These filters extracted low-frequency components with cut-off frequencies typically around 7-9 Hz. The distinction in power between high-frequency and low-frequency components proved crucial for accurate signal classification.

It can be observed that for low-level features extraction of sounds in RED, most approaches are focused on analyzing the sound frequencies via techniques such as the Fourier Transform, or the usage of the spectrum of frequencies with spectrogram-based features.

C. HIGH-LEVEL FEATURES EXTRACTION

High-level features extraction from sound data for rare event detection captures abstract and semantic representations. This process emphasizes the identification of advanced characteristics, including temporal and spectral patterns, to offer a richer understanding of rare events [5]. The 1D ConvNet effectively extracts high-level features and captures temporal dependencies, which are crucial for detecting rare sound events [3]. The 1D ConvNet converts spectral features in log-amplitude mel-spectrograms using 128 filters and batch normalization. It produces 97 elements with batch normalization, ReLU activation, max-pooling, and dropout, ensuring a consistent output features size. These features capture the unique traits required for the detection and classification of rare target events. High-level sound features originate from the basic features. These features offer semantic sound descriptions linked to human interpretations. To enhance the detection accuracy of their method, CNNs and FNNs were used, and a novel data augmentation (DA) technique was introduced. This technique utilizes dynamic time warping to mitigate the issue of data imbalance.

Recurrent Neural Networks (RNNs) were used by [64] for multi-resolution features extraction to handle time axis variations. This architecture sub-samples time at a rate of two, averaging the outputs of neighboring RNN cell frames. The resulting sequence is half the input length and is used as input into the next recurrent layer. Higher layers view the original utterances at coarser resolutions and extract information from a larger context.

The introduction of an adaptive few-shot learning algorithm for rare sound event detection aimed to improve few-shot learning in sound-event recognition [53]. This algorithm identifies rare auditory events with limited information, which is a common issue in practical situations.

Surampudi et al. classified the detection of rare events in sounds into two domains: audio processing and audio event detection [8]. Hyungui et al. introduced a rare sound event detection system using a combination of a 1D ConvNet and an RNN with LSTM [3].

In [53], metric-based few-shot learning with a task-adaptive module was used to detect rare sound events by identifying class uniqueness and support set commonality. The module improved the performance of the two datasets, particularly for the transductive propagation network.

High-level features play a crucial role in enhancing the accuracy and robustness of signal detection. Employing features extraction techniques like principal component analysis (PCA) and independent component analysis (ICA) offers the capability to effectively reduce the dimension of the data while extracting features [60].

In evaluating multi-class anomalous classifications, 1D-CNN is used. The model was trained on generated data and tested on original data. The 1D-CNN learns both high-level and low-level features [60].

In [66], DNNs were used to implement GANs for generating rare events in wireless communication data. The GANs utilized two networks - a generator and a discriminator - in an adversarial learning process. The generator creates a sample and the discriminator compares it to a real sample.

In [59], 1D CNN and 2D CNN were compared in discriminating events in a dataset. The study found that 1D CNN successfully learned the necessary features to effectively discriminate between signal classes.

High-level features extraction of sound for RED is dominated by the usage of DL approaches. DL approaches are effective in capturing the semantic information contained within the sound since DL has the capability of effectively extracting sound features at multi-resolution (through its various hidden layers).

D. CLASSIFICATION

Classification in sound-based RED, categorizes sounds into classes, with one class representing a rare event. The goal is to train a model to identify rare events in sound data. Classification in RED also involves assigning a label to a sound segment based on its acoustic characteristics. The labels correspond to the presence or absence of rare events in the sound segment [8]. Additionally, classifying each frame of a sound signal based on its acoustic contents is another classification method [65]. As an example, figure 7 presents the architecture of the You Only Hear Once (YOHO) algorithm [65], which is based on a similar architecture that is used for object detection.

Sound event detection involves identifying specific sound events within an audio recording, such as a dog barking, a car passing by, or a person speaking [63].

Detecting rare events based on sound requires accurate classification. Features derived from audio data can be used to train ML models to differentiate between normal and rare events. In [8], it is emphasized that the efficacy of rare event detection in sound can be enhanced by utilizing digital signal processing techniques for features extraction and supervised ML methods for classification. Five classification methods: the Boosted Tree Classifier (BTC), Random Forest Classifier (RFC), k-Nearest Neighbour (k-NN), SVM, and Artificial Neural Network (ANN), were employed in the model classification process.

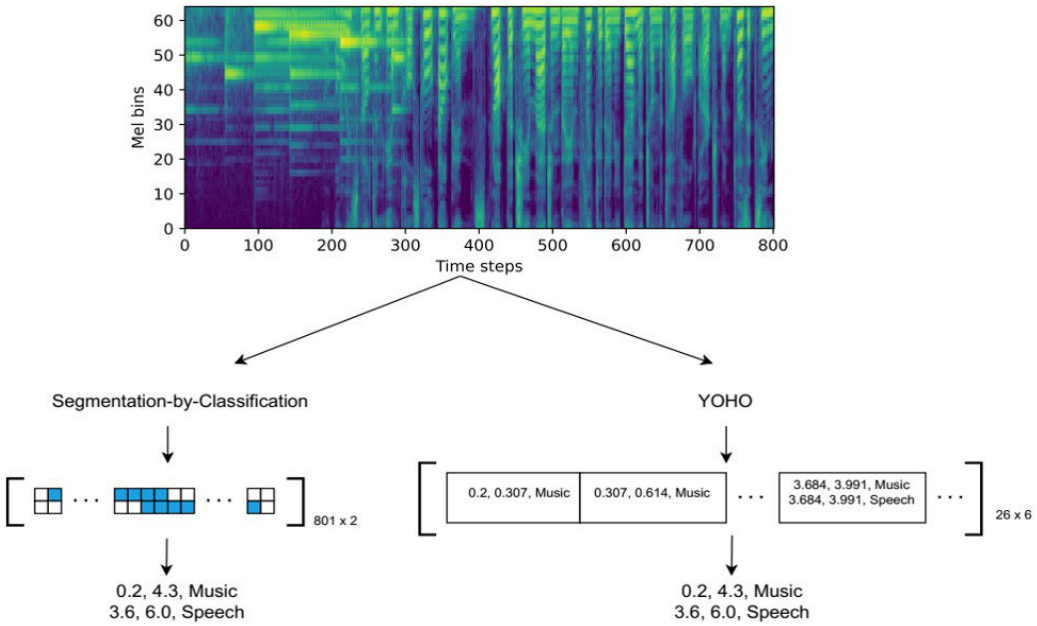

FIGURE 7. Architecture of the YOHO algorithm, which consists of a deep neural network that performs frame-level classification of mel spectrogram features. The network is based on the You Only Look Once (YOLO) object detection architecture and is modified to predict the presence or absence of sound events in each frame of the input audio signal, [65].

Hyungui et al. used the temporal dependency of the extracted features to incorporate the RNN-LSTM model. The RNN-LSTM is a well-known DL model that prevents the vanishing gradient problem and captures the temporal dependencies between the features over time [3].

In their proposed method, [54] used an SVM classifier with an RBF kernel for model fusion in sound event detection. The use of a hybrid approach combining CNN and RF for sound event detection (SED) in a natural forest environment in [62] achieved a remarkable performance, showing improvement with a 0.82 F1 score and a minimum false alarm rate of 10% in SED.

In signal-based methods, rare occurrences are identified by categorizing input signals according to their unique characteristics and properties. The aim is to differentiate normal patterns from abnormal ones. Classification algorithms use labeled data to learn how to categorize the signals. Classification is a technique for identifying patterns in the data that correspond to different states or conditions [61].

In a recent study [60], a 1D-CNN model was used as a classifier to evaluate the effectiveness of a newly proposed data generation framework. The model was trained on the generated data and then tested on the original datasets.

$$Accuracy = \left(\frac{TP + TN}{TP + TN + FP + FN}\right) \times 100 \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1{-}score = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall}\right) \quad (4)$$

The model's performance was assessed in terms of accuracy, precision, recall, and F1-score, as defined by equations (1), (2), (3), and (4), respectively.

The importance of carefully selecting features for classification and assessing performance was stressed in [61]. Techniques like LDA, SVM, and ANNs were discussed for classification in signals.

For features extraction and classification, a deep learning approach with CNN was employed [59]. CNNs are known for their ability to automatically learn high-level features from raw input data by utilizing multiple layers of convolution and pooling operations. The CNN's parameters were trained using a cross-entropy loss function with the Adam optimization algorithm in mini-batches of 64 waveforms.

It can be observed that for RED in sounds, the majority of the approaches generalize the approaches that are typically used for image or video types of modalities. It can also be observed often ML techniques are used in combination with DL approaches for classification.

E. DATASETS FOR SOUND BASED APPROACHES FOR RARE EVENT DETECTION

Table 3 provides information about various sound datasets used for rare sound event detection, along with corresponding performance metrics achieved by relevant works using the datasets, where available.

The first column in Table 3 denotes the works that evaluated their approaches against the specific datasets in a particular row.

Among the datasets with available performance metrics is the **TUT Acoustic Scenes2016**. Mesaros et al. [63]

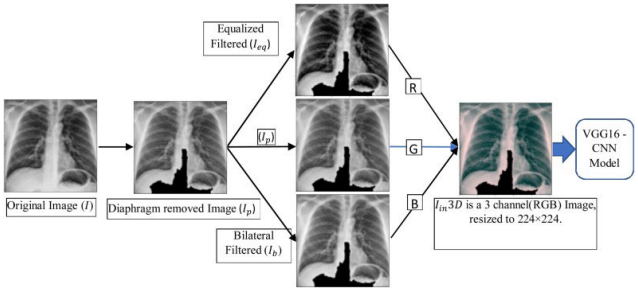

FIGURE 8. A flow diagram to illustrate image pre-processing steps to generate input of a CNN model, where (I) is the original Image in the dataset. (I_p) is the diaphragm removed image. (I_{eq}) is an image after applying histogram equalization on (I_p), and (I_b) is an image after applying bilateral filtering on (I_p). Three images (I_p), (I_b), and (I_{eq}) are fed into three channels of the CNN model to simulate the RGB image [77].

achieved an F1-Score of **96.26%** [3] and error rate of **0.07** [63].

In summary, the table provides insights into various sound datasets used for rare sound event detection. Some datasets achieve high performance, while others address specific challenges associated with detecting rare sound events. It can be observed that RED for sounds based on the various performances against the different datasets have reached a good level of RED performance. Most of the datasets are paired with corresponding works that achieved more than 90% in terms of classification performance.

V. IMAGE BASED APPROACHES FOR RARE EVENT DETECTION

Image-based methods utilize computer vision to identify rare events in visual data by extracting relevant features and classifying visual data.

The application of an image-based approach in RED often involves traditional handcrafted methods. However, these existing methods require domain-specific knowledge and manual tuning, making them time-consuming and difficult to scale for large datasets. In recent years, approaches such as deep learning have emerged, focusing on learning complex representations of input data automatically. These methods have shown promising results [10].

A. PREPROCESSING

Preprocessing in image-based rare event detection involves preparing visual data for analysis. This includes steps to improve image quality, reduce noise, and extract important data for rare event detection. The preprocessing stage aims to enhance raw image data [76].

Figure 8 depicts the flow diagram from the works of Heidari et al. [77], which contains image pre-processing steps to generate input for a CNN model.

Resampling plays a pivotal role in detecting rare events within industrial datasets. In the study by [78], imbalanced datasets were resampled to address the infrequent occurrences of rare events compared to others. By adjusting the dataset to balance these rare instances, the classifier's efficacy in recognizing and interpreting events is enhanced.

Dimokranitou et al. preprocessed frames by resizing them to 192 × 192 in order to reduce computational complexity and ensure uniform input images. They normalized the images by subtracting mean pixel values and dividing them by the standard deviation, which improved learning performance [10].

It can be observed that in the preprocessing of images for RED, most approaches are focused on reducing the size and dimension of the images to reduce the computational complexity via techniques such as images resampling and resizing.

B. LOW-LEVEL FEATURES EXTRACTION

Low-level features extraction in image rare event detection focuses on obtaining fundamental attributes without deeper semantic or contextual interpretations. Common techniques include edge detection, textures, color, shape, brightness, contrast, histograms, and spatial frequency. These features provide a foundational understanding of the image's content, especially when anomalies or rare events are subtle or embedded within the image [10].

In [10], handcrafted features such as optical flow, histogram of oriented gradients (HOG), and scale-invariant feature transform (SIFT) are used to extract important information from images.

Here, it is observed that low-level features extraction in images for RED is mostly based on handcrafted techniques that contains mainly visual information, however, lacking the semantic information.

C. HIGH-LEVEL FEATURES EXTRACTION

To detect rare events, image-based approaches extract high-level features by capturing abstract and semantic information from the images. These features are relevant to the detection of rare events and aid in advanced image analysis and classification.

High-level features extraction involves extracting more abstract and complex features from the raw pixel values of an image. These features are typically global and semantic, encompassing characteristics such as object categories, scene types, and human actions [10].

Hamaguchi et al. [76] employed a pre-trained CNN to extract features from input images. This CNN consists of multiple convolutional layers designed to discern features at various granularities, ranging from fundamental ones like edges and corners to more intricate attributes such as object components and textures. The features thus extracted serve as the foundation for training a change detection model. During the fine-tuning process, negative samples are purposefully under-sampled to align with the quantity of positive samples.

In their research work, [10] noted that deep learning (DL) methods like the proposed Adversarial Autoencoder (AAE) model can be valuable for high-level features extraction. DL methods are capable of learning complex and abstract representations of input data through multiple layers of nonlinear transformations. This enables the model to capture

TABLE 3. Summary of datasets used in sound-based approach, detailing dataset references, benchmark performance, sample sizes, events detected, and repository locations.

Ref.	Name of Dataset	Benchmark Performance	No. of Sample	Event Detected	Repository/Location
[3] [54] [64]	DCASE 2017 Challenge Task 2 Dataset [67]	ER=0.13,F1-Score=93.1 EER=15.33% F1-Score=94.2%	330-second audio segments from 15 acoustic scenes	Multiple target sound	https://dcase.community/challenge2017/download
[3] [63] [65]	TUT Acoustic Scenes 2016 [68]	ACC = 72.9% ER= 0.07,F1-Score= 96.26 F1-Score=0.63	1500 Mixtures	baby cry, gunshot, glass break	https://zenodo.org/record/401395#.ZDv9U3ZBw2w
[8]	Real-world Dataset [69]	F1-Score = 0.98	1500 Mixtures	baby cry, gunshot, glass break	https://scikit-learn.org/stable/datasets/real_world.html
[8]	Synthetic Dataset	F1-Score = 0.99	1500 Mixtures	baby cry, gunshot, glass break	Personally generated
[53]	ESC-50 & noiseESC-50 datasets [70]	ACC = 80.5%	2000audios x 50 classes	challenges of rare sound event detection, which include data deficit and cold start	https://github.com/karolpiczak/ESC-50
[54]	DCASE 2016, DCASE 2019 datasets [63]	EER=15.90% EER=19.58%	30-sec audio segments x 15 acoustic scenes	glass breaking, smoke alarm	https://dcase.community/challenge2016/download
[59]	Sakurajima Dataset [71]	BACC = 0.943	Ashfall deposition data	explosion earthquake (ER), non-explosion earthquake (NER) & tectonic tremor (TT)	https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-2848
[59]	NKL Dataset [72]	BACC = 0.965	5,000 images of various scenes	explosion earthquake (ER), non-explosion earthquake (NER) & tectonic tremor (TT)	https://kaizhao.net/nkl
[60]	ECG5000 Datasets [73]	F1-Score= 15.50%	5,000 ECG recordings x 1,500 datapoints	Anomaly detection in 1D signals	https://www.timeseriesclassification.com/description.php?Dataset=ECG5000
[62]	Personal data collection from natural forest	-	data was divided into in-sample & out-of-sample subsets	sound events of tree cutting, chain saw, vehicle activities	Not public
[65]	(MIREX) competition dataset 2018 [74]	F1-Score=90.20%	27 hrs audio x 8TV programs	Audio Segmentation and Sound Event Detection to detect the presence of an audio class and predict its start and end points	https://www.music-ir.org/mirex/wiki/2018:Music_and/or_Speech_Detection
[66]	Channel Estimate (CE) dataset [75]	No performance provided	10,000,000 data points	used in experiment 1 to generate rare events. The experiment compares the performance of a conventionally trained GAN with one trained using incremental learning for generating the CE data	https://github.com/topics/channel-estimation
[66]	SINR dataset	No performance provided	15,000,000 data points	used to quantify the rate of information that can reliably be transferred in wireless communication systems	Not public

high-level patterns and relationships in the data that may be difficult to capture using handcrafted features or low-level features alone.

In the study conducted by [10], a pre-trained VGG-16 network was utilized to extract features from input images through a transfer learning approach. The output of the VGG-16 network subsequently served as the input for the AAE model.

In this section, it can be concluded that high-level features extraction in images for RED is often achieved using DL techniques that make use of multiple layers, that are either convolutional or utilize nonlinear transformations. These multiple layers allow for semantic information within the images to be captured and utilized for image based RED.

D. CLASSIFICATION

Image-based rare event detection involves categorizing each image or segment as normal or rare based on the extracted features. This process is carried out by training a model using a set of labeled training data, which consists of a collection of images for which you know whether they correspond to a rare event.

Almost all classified models use a fine-tuned event detector that is trained on pairs of observations with class-imbalanced datasets. In these datasets, one observation contains a rare event while the other does not [76].

The AdaBoost-based features selection algorithm for rare event detection, proposed by [1], highlights different boosting algorithms, such as FloatBoost, Gentle AdaBoost, and CART

decision trees. For difficult classification tasks requiring real-time or online learning, a faster algorithm is preferred.

The use of a combination of algorithms to detect rare events in industrial datasets for classification was discussed in [78]. Methods such as feed-forward neural networks (FFNNs) were used to identify metal sheet surface defects by analyzing the chemical composition and cleaning process parameters. DTs were used for detecting surface defects on metal sheets, while RF enhances accuracy, robustness, and detects clogging in wastewater plants within the classification model. SVM and KNN are employed to detect faults in manufacturing plants.

The CNN's VGG-16 network, commonly used for image classification tasks, was employed to extract features from the input images in [10]. They adopted the Transfer Learning (TL) approach, using the last convolutional layer of the VGG-16 network as input for the AAE model. The primary advantage of using DL methods, such as VGG-16, is that they can automatically learn important features from any situation without the need for manual feature engineering.

In image-based RED classification, the task involves categorizing images as normal or rare based on the extracted features. Training models using labeled training data is crucial for successful classification. AdaBoost-based feature selection algorithms like FloatBoost, Gentle AdaBoost, and CART decision trees are commonly used in RED. The combination of algorithms such as FFNNs, DTs, RF, SVM, and KNN can effectively detect rare events in industrial datasets. CNNs, such as VGG-16, are utilized for image classification tasks, often adopting the Transfer Learning (TL) approach to automatically learn important features.

E. DATASETS FOR IMAGE BASED APPROACHES FOR RARE EVENT DETECTION

Table 4 provides information about various image datasets used for rare event detection, along with their performance metrics. The one with the highest performance is the **ABCD Dataset** [79] with an **accuracy (ACC) of 89.70%** achieved by the work of Hamaguchi et al. [76]. It focuses on detecting building changes from aerial images taken before and after the tsunami disaster.

It can be observed from the list of performances achieved against the various datasets that RED for images is a rather difficult task. Based on the results reported in Table 4, the performances range between 75% to almost 90% in terms of classification accuracy. This can be attributed to the fact that images are more complex, and RED in images is often a not straigh-forward task.

VI. TIME SERIES BASED APPROACHES FOR RARE EVENT DETECTION

Analyzing time series data involves detecting unusual patterns that occur over time. This technique is particularly valuable in industries like finance, healthcare, and environmental monitoring, where tracking sequential data helps to identify significant deviations from the norm. Such deviations are referred to as rare events, as they represent anomalies that deviate from the typical operation of a system [84]

A. PREPROCESSING

Effective preprocessing plays a crucial role in time series rare event detection, significantly influencing result quality and analysis performance. A range of preprocessing techniques are essential to optimize algorithm effectiveness. These techniques encompass noise handling, data normalization, addressing missing values, and transforming time series data. Preprocessing refers to a set of operations performed on raw time series data prior to its utilization in analysis or modeling [55]. Preprocessing steps are essential to ensure that the data is clean, relevant, and in the right format for building a classification model [85].

Filtering is also essential for preprocessing, particularly to remove trends in time series when detecting financial bubbles. Reference [57] applied Hodrick-Prescott (HP) and Kalman filters to preprocess time series data, and found that the HP filter was the best preprocessing method for their proposed approach.

$$W_t = \{x_{t-1+1}, \ldots, x_{t-1}, x_t\} \tag{5}$$

In Pillai et al. [86] study, several preprocessing steps were performed to optimize training and inference. These included imputing missing data and unevenly sampled time series using the mean, forward-filling missing rare event and workplace performance labels, applying within-subject feature normalization, and transforming each participant's time series into windows of length l=10 using equation (5).

It can be observed that the preprocessing techniques used for time-series data often rely on the nature of the data.

B. LOW-LEVEL FEATURES EXTRACTION

Low-level features extraction in multivariate time series data involves extracting features directly from the raw data without any prior knowledge of the data [85].

Time series-based methods extract basic features from raw data for rare event detection. These features capture key data properties, such as statistics, distribution, or temporal patterns [87]. Low-level features aid in identifying anomalies or rare events.

Coffinet et al. [57] suggested a machine learning toolkit to detect rare events in banking crises in time series data, utilizing low-level features extraction methods such as credit and GDP lag versions and inflation rates. Lagged credit and GDP series also enhance crisis detection.

C. HIGH-LEVEL FEATURES EXTRACTION

Extracting high-level features from time series data for rare event detection involves complex or abstract features extraction. These traits detect rare occurrences based on advanced patterns, relations, or context. High-level features enable advanced analysis for rare event detection.

A method for detecting rare life events using mobile sensing data proposed in [86], used a multi-task framework

TABLE 4. Summary of datasets used in image-based approach, detailing dataset references, benchmark performance, sample sizes, events detected, and repository locations.

Ref.	Name of Dataset	Benchmark Performance	No. of Sample	Event Detected	Repository/Location
[10]	UCSD Peds1 dataset [39]	AUC = 0.93	34 training x 36 testing videos x 200 frames	Passage of non-pedestrian access, such as vehicles & bicycles from the pedestrian path	http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm
[10]	UCSD Peds2 dataset [39]	AUC = 0.91	16 training x 12 testing videos x 12 abnormal events	Passage of non-pedestrian access, such as vehicles & bicycles from the pedestrian path	http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm
[34]	ImageNet dataset [80]	AP = 0.2797, R=0, FP = 0.5373, AUC = 0.9611	14,197,122 annotated images	Used to train a deep learning model and extract features that were used in the experiments	https://www.image-net.org/update-mar-11-2021.php
[76]	Augmented MNIST [81]	ACC = 81.54%	60,000 training set x 10,000 test set	baby cry, gunshot, glass break	https://www.kaggle.com/datasets/hojjatk/mnist-dataset
[76]	ABCD Dataset [79]	ACC = 89.70%	ABCD cohort (N = 11,880)	This dataset is used for detecting changes in buildings from a pair of aerial images taken before and after a tsunami disaster	https://nda.nih.gov/abcd/
[76]	PCD dataset [82]	ACC = 78.20%	45 classes of images	airport, beach, bridge, & farmland	https://paperswithcode.com/dataset/pcd
[76]	WDC Dataset [83]	ACC = 75.70%	14 rare events	landslide, flood & fire	http://webdatacommons.org/largescaleproductcorpus/
[78]	Metal Sheets Datasets	R = 20%	7,000 observations x 13 features	Machine faults, defective products	Not public

with an unsupervised autoencoder to capture irregular behaviour and an auxiliary sequence predictor to identify transitions in workplace performance to contextualize events.

The use of long short-term memory units (LSTM)-layer to take multivariate dependencies into account was highlighted in [84]. LSTMs are RNNs that can capture temporal dependencies in sequential multivariate time series data. Meng et al. [14] highlighted the use of an Extensible Markov Model (EMM). The EMM method models the spatiotemporal environment using a graph structure with nodes representing states and edges representing transitions. The EMM algorithm learns transition probabilities between states and constructs a graph representing the system behaviour. When a new event is observed, the EMM algorithm calculates the probability using transition probabilities in the graph. The event is rare and flagged as an anomaly if the probability is too low. The EMM approach works for supervised and unsupervised rare-event detection. In their study, Pillai et al. [86] utilized an unsupervised AE to extract low-level features from raw sensor data. These features, obtained through the AE, were used as input data for the sequence predictor. To learn the normal patterns of the multivariate time series data, the researchers employed AEs and counterfactual explanations. The counterfactual explanations were generated by perturbing the input data to the autoencoder and observing changes in the output. The aim of these perturbations was to highlight the features that are most relevant to the anomaly [84].

D. CLASSIFICATION

Classifying time-series approaches identify normal and rare events based on the given time series' features. Classification trains a model with labelled examples to predict unseen data. Rare event detection models differentiate between normal and rare patterns in time series data.

However, binary classification tasks can be used in time series data to discriminate between events with the goal to detect anomalies in the data. A method that combines strong simulation and multilevel splitting to estimate rare event probabilities in Markov processes was proposed in [88] with strong simulation ideas to avoid bias but there is a need for improvement in the scalability of the method. The utilization of the XGBoost and AdaBoost models has been employed in [85] for training the predictive models. Initially, the models were trained to utilize the provided predictor variables. Coffinet et al. [57] proposed the use of data science models to detect rare events like banking crises, including RF methods. They used Breiman's Random Forest (BRF) with 500 trees and replacement sampling.

ML methods are used to detect and classify unusual patterns or anomalies in time-series data, particularly in power systems, to help operators understand what is happening and make decisions quickly. This involves using advanced algorithms like Generative Adversarial Networks (GANs) and Neural Ordinary Differential Equations (ODEs) to create synthetic data that mimics real Phasor Measurement Unit (PMU) data [12].

TABLE 5. Summary of datasets used in time-series based approach, detailing dataset references, benchmark performance, sample sizes, events detected, and repository locations.

Ref.	Name of Dataset	Benchmark Performance	No. of Sample	Event Detected	Repository/Location
[14]	MnDot traffic data [90]	No performance provided		No information provided	https://www.dot.state.mn.us/traffic/data/tma.html
[14]	VoIP Traffic Data	No performance provided		No information provided	Not public
[58]	Multivariate time-series	DR = 80%, FDR = 9%	67% Training, 33% test	Contamination event in water	Not Public
[84]	New York City Taxi [91]	No performance provided	13yrs trips x 19 features	trip count,avg. trip duration, avg. no of pers. per trip	https://data.cityofnewyork.us/Transportation/2018-Yellow-Taxi-Trip-Data/t29m-gskq
[86]	Tesserae study dataset [89]	F1-Score = 0.29	10106days x 198 rare events	Time durations (walking, sedentary, running, distance, phone unlock), No. of locations visited & unique locations visited	https://tesserae.nd.edu/
[92]	MODIS [83]	No performance provided	spatial resolution of 250m x 250m	Time & amplitude	https://modis.gsfc.nasa.gov/data/dataprod/

E. DATASETS FOR TIME SERIES BASED APPROACHES FOR RARE EVENT DETECTION

Table 5 provides information about various time series datasets used for rare event detection, along with their performance metrics. It's important to note that several datasets do not have any associated performance metrics, which limits our ability to directly compare the different datasets.

For the **Tesserae Study Dataset** [89], Pillai et al. achieved an **F1-Score of 0.29** [86].

It can be observed that although there are a number of datasets focusing on time series data, there is a lack of work focusing on RED for time series data.

VII. COMPARATIVE ANALYSIS OF THE PERFORMANCES OF THE PROPOSED APPROACHES

The performances of different approaches across different modalities are shown in Table 6. The table contains a compilation of research studies for different modalities, including video, sound, image, and time series. Each row corresponds to a study, providing details about the year of publication, modality, ML/DL methods/ models used, techniques/ algorithms applied, associated datasets (referenced with their corresponding dataset tables), and the reported performance metrics.

In the video modality, multiple studies incorporate CNNs, AE, and SVMs. The VGG 16 architecture is utilized in one study for rare event detection on the UCSDPed1 dataset, achieving a Detection Rate (DR) of 92.5% and a False Alarm Rate (FAR) of 0.0001. Other studies involve methods like Bayesian DL and HMM.

However, in sound, various ML/DL methods are explored including RNNs, Few-Shot Learning (FSL), and CNNs. Different datasets are employed, and performance metrics include Accuracy (ACC), F1-Score, and Equal Error Rate (EER). The image modality studies utilized CNNs, Adaboost, FFNN, and DT. The ResNet18 architecture is used in one study with AUC-ROC metrics for class performance measurement.

Additionally, in time series modality, ML/DL methods such as AdaBoost, XGBoost, AE, Bayesian NN, LSTM, and RNN were used. Reported performance metrics consist of F1-Score, Precision, Recall, AUC-ROC, and more.

To summarize, we have conducted an analysis based on the information presented in the table, and these are our findings: The **video** modality exhibits the highest performance with an accuracy (ACC) of up to **98.5%**, F1-Score of up to **0.91**, and detection rate (DR) of **92.5%**. The **sound** modality has the most significant representation in the table, indicating its popularity in rare event detection research. The analysis reveals that the choice of modality significantly affects the performance of rare event detection methods. The video modality demonstrates the highest performance, while the sound modality is the most popular among researchers. The analysis also reveals that CNN is the most popular method used in video, sound, and image modalities and AE is the most commonly used in the time series modality.

These methods are widely adopted for rare event detection tasks across different modalities. Researchers should consider the specific modality, method, available datasets, and performance metrics when selecting the appropriate approach for their rare event detection tasks.

VIII. DISCUSSION

There are numerous challenges that the RED encounters across the modalities. In video modality, the lack of labels and the complexity of relations between events can make it

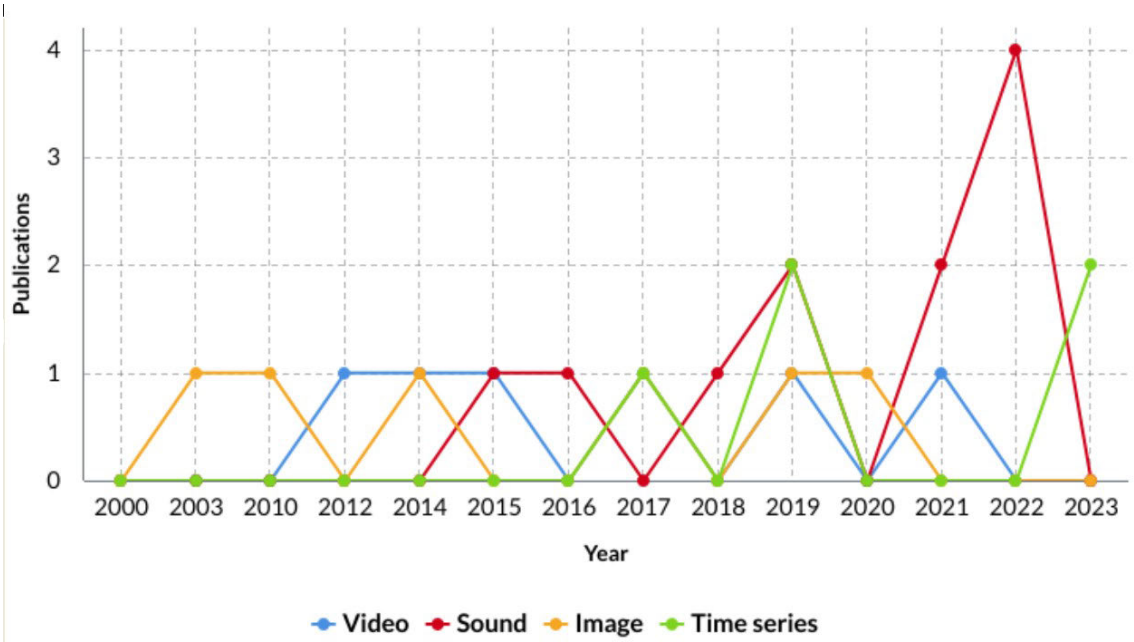

FIGURE 9. An overview of the number of publications per year across various modalities from 2000 to 2023. The figure showcases trends in publications related to Video (6 total), Sound (11 total), image (6 total), and time series (5 total). Notably, there's been a surge in sound-related publications in recent years, while time series publications saw a peak in 2019 and 2023.

TABLE 6. Table of performance: Overview of research works categorized by modality, showcasing the year of publication, utilized machine learning or deep learning models, applied techniques or algorithms, referenced dataset tables, and respective performance metrics.

Ref.	Year	Modality	ML/DL Method/Model	Technique/Algorithm	Dataset (Table)	Performance Metric
[9]	2021	Video	CNNs, SVM	IA-SSLM	Table [2]	ACC-98.5%,P-98.2%,R-98.4%,F1-Score-98.3%
[36]	2019	Video	CNN, AE, SVM	VGG 16	Table [2]	Ped1: DR-92.5%, FAR-0.0001
[35]	2017	Video	CNNs	-	Table [2]	F-Score - 0.91 & F-Score - 0.85
[23]	2015	Video	SVM	HMM & CRF	Table [2]	-
[37]	2014	Video	CNNs	ConvNets	Table [2]	ACC-88.0% & ACC-59.4%
[30]	2012	Video	Bayesian DL	DDMCMC	Table [2]	-
[53]	2022	Sound	Few-SHot Learning	AFSL	Table [3]	ACC- 80.5%
[65]	2022	Sound	CNNs	YOHO	Table [3]	F1-Score - 0.89
[62]	2022	Sound	CNN & RF	FFT	Table [3]	F1-score-0.82, FAR - 10%
[59]	2022	Sound	CNN	1D CNN	Table [3]	BACC-0.943 & BACC-0.986
[60]	2021	Sound	1DCNN	Data augmentation	Table [3]	ACC-84.5%
[66]	2021	Sound	GANs	Balance Replay GAN	Table [3]	(AUC-ROC):0.998 & PR-AUC:0.997
[8]	2019	Sound	KNN, SVM & ANN	Butterworth filter	Table [3]	ACC-69.13%
[54]	2019	Sound	CNN & FNN	-	Table [3]	EER- 16.70%, 23.15%, 18.76%
[64]	2018	Sound	RNNs	GRU	Table [3]	ER-011, F1-Score - 0.57
[63]	2016	Sound	MFCC & GMM	SASCS	Table [3]	ACC-72.9%, F1-Score - 0.44
[3]	2015	Sound	1D CNN, RNN & LSTM	HMM, NMF	Table [3]	EER-13.46
[93]	2020	Image	SVM, LR, NB & KNN	PMQ-L	-	MPCD - 0.9879
[76]	2019	Image	CNN	ResNet18	Table [4]	(AUC-ROC): 0.5(rand. Guess) to 1.0 (perf. class)
[10]	2017	Image	AAE, CNN & SVM	VGG16	Table [4]	(AUC-ROC): Ped1-0.98 & Ped2-0.96
[34]	2014	Image	DCNN	Linear SVM	Table [2]	-
[78]	2010	Image	FFNN & DT	BR	-	-
[1]	2003	Image	AdaBoost	FFS	-	-
[86]	2023	Time Series	AE	-	Table [5]	(AUC-ROC) - 0.7 to 0.9
[84]	2023	Time Series	LSTM,AE & RNN	-	Table [5]	(AUC-ROC) - 0 to 1
[85]	2019	Time Series	AdaBoost & XGBoost	-	-	F1-Score - 0.114, Prec -0.071, FPR-0.026
[57]	2019	Time Series	RF & FFANN	-	-	(AUC-ROC) - 0.7 to 0.9
[58]	2017	Time Series	Bayesian NN	BPNN Model	Table [5]	RD-40%, FAR-45%

difficult to learn the storyline and detect rare events. In sound, the low signal-to-noise ratio (SNR) and lack of labeled data make it difficult to train ML models accurately. Likewise, generating reliable data in 1D signals is also challenging due to the unavailability and inaccessibility of datasets for extremely rare cases signals.

In image, class imbalance and computational efficiency are challenges, as rare events are significantly outnumbered by non-event images. The imbalanced distribution of data can lead to the overfitting of common events and the underfitting of rare events. In time series, the rare and infrequent nature of events in time series data can lead to complexity in detection

and analysis. The demand for large amounts of data by deep learning approaches and the dimensionality of data make it difficult to identify relevant features and patterns.

During our review, we discovered a pressing need for research into real-time detection of rare events using video surveillance and audio. Furthermore, there is potential for improvement in the use of recently advanced deep learning architectures and techniques for rare event detection especially in the transformers and diffusion-based models, utilizing unlabeled data for training across multiple modalities.

This research delved into various studies regarding RED and evaluated their relevance to the research objective. Table [6] showcases the methods that are most suitable for detecting rare events across all four modalities and their corresponding performance. CNN and SVM have proven to be effective in detecting events in video, sound, and image-based approaches due to their dynamic nature. When combined with LSTM, they can accurately identify complex features like objects and scenes, which are formed from simpler features like edges and textures. On the other hand, AE and RNN are prominent in the remaining time series approaches, as noted in the studied articles. This study focuses solely on rare-event detection research and avoids any bias in the analysis by evaluating the performance of each method based on the performance metric used in the referenced article.

IX. CONCLUSION AND FURTHER WORK

Our objective is to study and evaluate advanced ML and DL techniques and frameworks that are suitable for rear event detection in video, sound, image, and time series modalities. This area has not been extensively studied in existing literature despite its significance. Our research discovered that specific methods are more effective and reliable in dealing with rare events in these four modalities. Although our study was limited by a lack of research, it provides valuable initial insights into the significant challenges in RED.

This research highlights the pressing need for more extensive studies in this area. By thoroughly examining techniques used in various articles and utilizing features extraction methods, it was discovered that CNN, SVM and AEs outperformed other ML/DL methods.

However, this study focused solely on identifying the most effective method for detecting rare events. We reviewed 217 research articles and found 28 that met our inclusion criteria. This study serves as a foundation for future studies in this field. Traditional signature-based detection methods cannot be effective for detecting rare events in a number of domains such as meteorology, environment, and finance due to their challenging nature, hence the need for the use of ML and DL methods.

To enhance the performance of Rare Event Detection (RED), future research should focus on exploring various types of AEs, evaluation metrics, and data preprocessing techniques. Furthermore, researchers could investigate the utilization of DL techniques and architectures to enhance detection in crowded scenes. Additionally, advanced approaches based on recent DL advancements, such as Transformers, LLM, and diffusion models in real-time, could be explored to further improve the detection of rare events.

REFERENCES

[1] J. Wu, J. M. Rehg, and M. D. Mullin, "Learning a rare event detection cascade by direct feature selection," Interact. Media Technol. Center, Atlanta, GA, USA, Tech. Rep. GIT-GVU-03-16, 2003.

[2] G. Pang, L. Cao, and C. Aggarwal, "Deep learning for anomaly detection: Challenges, methods, and opportunities," in *Proc. 14th ACM Int. Conf. Web Search Data Mining*, Mar. 2021, pp. 1127–1130.

[3] L. Hyungui, P. Jeongsoo, L. Kyogu, and H. Yoonchang, "Rare sound event detection using 1D convolutional recurrent neural networks," *IEEE Trans. Multimedia*, vol. 17, pp. 1733–1746, 2015.

[4] M. Sokolova, K. El Emam, S. Chowdhury, E. Neri, S. Rose, and E. Jonker, "Evaluation of rare event detection," in *Advances in Artificial Intelligence* (Lecture Notes in Computer Science), A. Farzindar and V. Kešelj, Eds. Berlin, Germany: Springer, 2010, pp. 379–383.

[5] I. Ivanov, F. Dufaux, T. M. Ha, and T. Ebrahimi, "Towards generic detection of unusual events in video surveillance," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2009, pp. 61–66.

[6] S. Narayanan, C. Maple, and M. Hooper, "A point process model for rare event detection," 2022, *arXiv:2209.04792*.

[7] D. C. Harrison, W. K. G. Seah, and R. Rayudu, "Rare event detection and propagation in wireless sensor networks," *ACM Comput. Surveys*, vol. 48, no. 4, pp. 1–22, Mar. 2016.

[8] N. Surampudi, M. Srirangan, and J. Christopher, "Enhanced feature extraction approaches for detection of sound events," in *Proc. IEEE 9th Int. Conf. Adv. Comput. (IACC)*, Tiruchirappalli, India, Dec. 2019, pp. 223–229.

[9] A. S. Aljaloud and H. Ullah, "IA-SSLM: Irregularity-aware semi-supervised deep learning model for analyzing unusual events in crowds," *IEEE Access*, vol. 9, pp. 73327–73334, 2021.

[10] A. Dimokranitou, "Adversarial autoencoders for anomalous event detection in images," A thesis, Dept. Comput. Inf. Sci., Indiana Purdue Univ., Indianapolis, 2017.

[11] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining, Volume 72 of Intelligent Systems Reference Library*, vol. 72. Cham, Switzerland: Springer, 2015.

[12] M. J. Zideh, P. Chatterjee, and A. K. Srivastava, "Physics-informed machine learning for data anomaly detection, classification, localization, and mitigation: A review, challenges, and path forward," *IEEE Access*, vol. 12, pp. 4597–4617, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10375385?denied=

[13] Z. H. Janjua, M. Vecchio, M. Antonini, and F. Antonelli, "IRESE: An intelligent rare-event detection system using unsupervised learning on the IoT edge," *Eng. Appl. Artif. Intell.*, vol. 84, pp. 41–50, Sep. 2019.

[14] Y. Meng, M. H. Dunham, F. M. Marchetti, and J. Huang, "Rare event detection in a spatiotemporal environment," in *Proc. IEEE Int. Conf. Granular Comput.*, May 2006, pp. 629–634.

[15] E. Şengönül, R. Samet, Q. A. Al-Haija, A. Alqahtani, B. Alturki, and A. A. Alsulami, "An analysis of artificial intelligence techniques in surveillance video anomaly detection: A comprehensive survey," *Appl. Sci.*, vol. 13, no. 8, p. 4956, Apr. 2023.

[16] M. Maalouf, *Rare Events and Imbalanced Datasets: An Overview*. Geneva, Switzerland: Inderscience Enterprises Ltd, Oct. 2015.

[17] D. Liu, "Meta-analysis of rare events," Univ. Cincinnati, Cincinnati, OH, USA, Tech. Rep., Aug. 2019, pp. 1–7.

[18] Y. Zhou, B. Zhu, L. Lin, J. Kwong, and C. Xu, "Protocols for meta-analysis of intervention safety seldom specified methods to deal with rare events," *J. Clin. Epidemiology*, vol. 128, pp. 109–117, Oct. 2020.

[19] P. Jia, L. Lin, J. S. W. Kwong, and C. Xu, "Many meta-analyses of rare events in the cochrane database of systematic reviews were underpowered," *J. Clin. Epidemiology*, vol. 131, pp. 113–122, Mar. 2021.

[20] L. Radulescu, D. Crisan, C. Grapa, and D. Radulescu, "Digestive toxicities secondary to immune checkpoint inhibition therapy–reports of rare events. A systematic review," *J. Gastrointestinal Liver Diseases*, vol. 30, no. 4, pp. 506–516, Dec. 2021.

[21] N. R. Haddaway, M. J. Page, C. C. Pritchard, and L. A. McGuinness, "PRISMA2020: An r package and shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis," *Campbell Systematic Rev.*, vol. 18, no. 2, p. e1230, Jun. 2022.
[22] M. J. Page, J. E. McKenzie, and P. M. Bossuyt, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *PLOS Med.*, vol. 18, Mar. 2021, Art. no. e1003583.
[23] G. Ye, "Large-scale video event detection," Ph.D. thesis, Graduate School Arts Sci., Columbia Univ., New York, NY, USA, 2015.
[24] D. R. Sharma and D. A. Sungheetha, "An efficient dimension reduction based fusion of CNN and SVM model for detection of abnormal incident in video surveillance," *J. Soft Comput. Paradigm*, vol. 3, pp. 55–69, May 2021.
[25] M. Chan, A. Hoogs, J. Schmiederer, and M. Petersen, "Detecting rare events using semantic primitives With HMM," *IEEE J.*, Sep. 2004, doi: 10.1109/ICPR.2004.1333726.
[26] C. Ma, Z. Wu, C. Cai, P. Zhang, Y. Wang, L. Zheng, C. Chen, and Q. Zhou, "Rate-perception optimized preprocessing for video coding," 2023, *arXiv:2301.10455*.
[27] G. Zhang, Y. Zhu, H. Wang, Y. Chen, G. Wu, and L. Wang, "Extracting motion and appearance via inter-frame attention for efficient video frame interpolation," 2023, *arXiv:2303.00440*.
[28] Z. Ye, X. Wang, H. Liu, Y. Qian, R. Tao, L. Yan, and K. Ouchi, "Sound event detection transformer: An event-based end-to-end model for sound event detection," 2021, *arXiv:2110.02011*.
[29] 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.
[30] J. Kwon and K. M. Lee, "A unified framework for event summarization and rare event detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1266–1273.
[31] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[32] Laptev and Lindeberg, "Space-time interest points," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, vol. 1, Nice, France, 2003, pp. 432–439.
[33] B. Mazin, J. Delon, and Y. Gousseau, "Combining color and geometry for local image matching," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 2667–2680.
[34] F. Luisier, M. Tickoo, W. Andrews, G. Ye, D. Liu, S.-F. Chang, R. Salakhutdinov, V. Morariu, L. Davis, A. Gupta, I. Haritaoglu, M. Park, S. Guler, and A. Morde, "BBN VISER TRECVID 2014 multimedia event detection and multimedia event recounting systems," Raytheon BBN Technol., Cambridge, MA, USA, Tech. Rep., 2014.
[35] B. Victor, Z. He, S. Morgan, and D. Miniutti, "Continuous video to simple signals for swimming stroke detection with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 122–131.
[36] S. Bansod and A. Nandedkar, "Transfer learning for video anomaly detection," *J. Intell. Fuzzy Syst.*, vol. 36, no. 3, pp. 1967–1975, Mar. 2019.
[37] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in Neural Information Processing Systems*, vol. 27. Red Hook, NY, USA: Curran Associates, 2014.
[38] N. Madan, N.-C. Ristea, R. T. Ionescu, K. Nasrollahi, F. S. Khan, T. B. Moeslund, and M. Shah, "Self-supervised masked convolutional transformer block for anomaly detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 1, pp. 525–542, Jan. 2024.
[39] LSVCL. *UCSD Anomaly Detection Dataset*. Accessed: Sep. 10, 2023. [Online]. Available: http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm
[40] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent flows: Real-time detection of violent crowd behavior," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 1–6.
[41] P. Nikolaos and Vassilios. *Unusual Crowd Activity Dataset of University of Minnesota*. Accessed: Sep. 10, 2023. [Online]. Available: http://mha.cs.umn.edu/movies/crowdactivity-all.avi
[42] D. Web, "The WebDataset format," webdataset, Tech. Rep., 2023. [Online]. Available: https://github.com/webdataset/webdataset
[43] D Sport 1m. *Sports 1m Dataset*. Accessed: Sep. 10, 2023. [Online]. Available: https://www.kaggle.com/datasets/sabahesaraki/sports-1m-dataset
[44] Y.-G. Jiang, G. Ye, S.-F. Chang, D. Ellis, and A. C. Loui, "Consumer video understanding: A benchmark database and an evaluation of human and machine performance," in *Proc. 1st ACM Int. Conf. Multimedia Retr.* New York, NY, USA: Association for Computing Machinery, Apr. 2011, pp. 1–8.
[45] Y.-G. Jiang, Z. Wu, J. Wang, X. Xue, and S.-F. Chang, "Exploiting feature and class relationships in video categorization with regularized deep neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 2, pp. 352–364, Feb. 2018.
[46] C. Shih-Fu. *DVMM—Demos and Downloads*. Accessed: Sep. 10, 2023. [Online]. Available: https://www.ee.columbia.edu/ln/dvmm/newDownloads.htm
[47] S. A. Velastin and D. A. Gómez-Lira, "People detection and pose classification inside a moving train using computer vision," in *Advances in Visual Informatics* (Lecture Notes in Computer Science), H. Badioze Zaman, P. Robinson, A. F. Smeaton, T. K. Shih, S. Velastin, T. Terutoshi, A. Jaafar, and N. M. Ali, Eds. Cham, Switzerland: Springer, 2017, pp. 319–330.
[48] *Traffic Flows, Borough—London Datastore*, DOT, London, U.K., 2014.
[49] B Data. *Subway Station Entrances—BetaNYC's Community Data Portal*. Accessed: Sep. 10, 2023. [Online]. Available: https://data.beta.nyc/en/dataset/subway-station-entrances
[50] C Newyork. *Swimming Data on Data.world | 20 Datasets Available*. Accessed: Sep. 10, 2023. [Online]. Available: https://data.world/datasets/swimming
[51] Datopian. *ATP World Tour Tennis Data*. Accessed: Sep. 10, 2023. [Online]. Available: https://datahub.io/sports-data/atp-world-tour-tennis-data#data-cli
[52] K. Soomro, A. R. Zamir, and M. Shah, "UCF101: A dataset of 101 human actions classes from videos in the wild," Center Res. Comput. Vis., Univ. Central Florida, Orlando, FL, USA, Tech. Rep. CRCV-TR-12-01, Dec. 2012.
[53] C. Zhao, J. Wang, L. Li, X. Qu, and J. Xiao, "Adaptive few-shot learning algorithm for rare sound event detection," 2022, *arXiv:2205.11738*.
[54] Y. Chen and H. Jin, "Rare sound event detection using deep learning and data augmentation," in *Proc. Interspeech*, Sep. 2019, pp. 619–623.
[55] A. M. Rizzo, L. Magri, D. Rutigliano, P. Invernizzi, E. Sozio, C. Alippi, S. Binetti, and G. Boracchi, "Known and unknown event detection in OTDR traces by deep learning networks," *Neural Comput. Appl.*, vol. 34, no. 22, pp. 19655–19673, Nov. 2022.
[56] S. Yasin, S. A. Hussain, S. Aslan, I. Raza, M. Muzammel, and A. Othmani, "EEG based major depressive disorder and bipolar disorder detection using neural networks: A review," *Comput. Methods Programs Biomed.*, vol. 202, Apr. 2021, Art. no. 106007.
[57] J. Coffinet and J.-N. Kien, "Detection of rare events: A machine learning toolkit with an application to banking crises," *J. Finance Data Sci.*, vol. 5, no. 4, pp. 183–207, Dec. 2019.
[58] Y. Mao, H. Qi, P. Ping, and X. Li, "Contamination event detection with multivariate time-series data in agricultural water monitoring," *Sensors*, vol. 17, no. 12, p. 2806, Dec. 2017.
[59] M. Nakano and D. Sugiyama, "Discriminating seismic events using 1D and 2D CNNs: Applications to volcanic and tectonic datasets," *Earth, Planets Space*, vol. 74, no. 1, p. 134, Sep. 2022.
[60] T. Chalongvorachai and K. Woraratpanya, "A data generation framework for extremely rare case signals," *Heliyon*, vol. 7, no. 8, Aug. 2021, Art. no. e07687.
[61] A. A. Torres-García, O. Mendoza-Montoya, M. Molinas, J. M. Antelis, L. A. Moctezuma, and T. Hernández-Del-Toro, "Chapter 4—Preprocessing and feature extraction," in *Biosignal Processing and Classification Using Computational Learning and Intelligence*, A. A. Torres-García, C. A. Reyes-García, L. Villaseñor-Pineda, and O. Mendoza-Montoya, Eds. New York, NY, USA: Academic Press, Jan. 2022, pp. 59–91.
[62] M. A. S. Md Afendi and M. Yusoff, "A sound event detection based on hybrid convolution neural network and random forest," *IAES Int. J. Artif. Intell. (IJ-AI)*, vol. 11, no. 1, p. 121, Mar. 2022.
[63] A. Mesaros, T. Heittola, and T. Virtanen, "TUT database for acoustic scene classification and sound event detection," in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2016, pp. 1128–1132.
[64] W. Wang, C.-c. Kao, and C. Wang, "A simple model for detection of rare sound events," 2018, *arXiv:1808.06676*.
[65] S. Venkatesh, D. Moffat, and E. R. Miranda, "You only hear once: A YOLO-like algorithm for audio segmentation and sound event detection," *Appl. Sci.*, vol. 12, no. 7, p. 3293, Mar. 2022.
[66] J. R. Baldvinsson, "Rare event learning In URLLC wireless networking environment using GANs," Ph.D. thesis, Dept. Elect. Eng. Comput. Sci., KTH Royal Inst. Technol., Stockholm, Sweden, 2021.
[67] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. D. Plumbley, "Detection and classification of acoustic scenes and events," *IEEE Trans. Multimedia*, vol. 17, no. 10, pp. 1733–1746, Oct. 2015.
[68] A. Diment, A. Mesaros, T. Heittola, and T. Virtanen, "TUT rare sound events, development dataset," Univ. Technol., Tampere, Finland, Tech. Rep. 603106, 2017.

[69] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[70] K. J. Piczak, "ESC: Dataset for environmental sound classification," in *Proc. 23rd ACM Int. Conf. Multimedia*. New York, NY, USA: Association for Computing Machinery, Oct. 2015, pp. 1015–1018.
[71] H. Rahadianto, H. Tatano, M. Iguchi, H. L. Tanaka, T. Takemi, and S. Roy, "Long-term ash dispersal dataset of the sakurajima taisho eruption for ashfall disaster countermeasure," *Earth Syst. Sci. Data*, vol. 14, no. 12, pp. 5309–5332, Dec. 2022.
[72] K. Zhao, Q. Han, C.-B. Zhang, J. Xu, and M.-M. Cheng, "Deep Hough transform for semantic line detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4793–4806, Sep. 2022.
[73] Y. Chen and E. Keogh. *Time Series Classification Website*. Accessed: Sep. 10, 2023. [Online]. Available: https://www.timeseriesclassification.com/description.php?Dataset=ECG5000
[74] D. Stephen. (2018). *2018: Music And/or Speech Detection—MIREX Wiki*. Accessed: Sep. 10, 2023. [Online]. Available: https://www.music-ir.org/mirex/wiki/2018: Music_and/or_Speech_Detection
[75] M. M. Juan. *Build Software Better, Together*. Accessed: Sep. 10, 2023. [Online]. Available: https://gist.github.com/jmmauricio
[76] R. Hamaguchi, K. Sakurada, and R. Nakamura, "Rare event detection using disentangled representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9319–9327.
[77] M. Heidari, S. Mirniaharikandehei, A. Z. Khuzani, G. Danala, Y. Qiu, and B. Zheng, "Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms," *Int. J. Med. Informat.*, vol. 144, Dec. 2020, Art. no. 104284.
[78] M. Vannucci, V. Colla, G. Nastasi, and N. Matarese, "Detection of rare events within industrial datasets by means of data resampling and specific algorithms," *Int. J. Simul. Syst. Sci. Technol.*, vol. 11, no. 3, pp. 1–17, 2010.
[79] F. Haist and T. L. Jernigan, "Adolescent brain cognitive development study (ABCD)–annual release 5.0," Nat. Inst. Mental Health Data Archive (NDA), Rockville, MD, USA, Tech. Rep., 2023.
[80] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[81] L. Yann. *MNIST Dataset*. Accessed: Sep. 10, 2023. [Online]. Available: https://www.kaggle.com/datasets/hojjatk/mnist-dataset
[82] Y. Waleed A, I. Omar M. Ibrahime, and M. Taha. (May 7, 2019). *Learning Meters of Arabic and English Poems With Recurrent Neural Networks: A Step Forward for Language Understanding and Synthesis*. [Online]. Available: https://paperswithcode.com/paper/190505700
[83] B. Christian. *Web Data Commons*. Accessed: Sep. 10, 2023. [Online]. Available: https://webdatacommons.org/index.html
[84] M. Schemmer, J. Holstein, N. Bauer, N. Kühl, and G. Satzger, "Towards meaningful anomaly detection: The effect of counterfactual explanations on the investigation of anomalies in multivariate time series," Tech. Rep., Feb. 2023. [Online]. Available: https://arxiv.org/abs/2302.03302
[85] C. Ranjan, M. Reddy, M. Mustonen, K. Paynabar, and K. Pourak, "Dataset: Rare event classification in multivariate time series," Tech. Rep., May 2019. [Online]. Available: https://arxiv.org/abs/1809.10717
[86] A. Pillai, S. Nepal, and A. Campbell, "Rare life event detection via mobile sensing using multi-task learning," 2023, *arXiv:2305.20056*.
[87] D. Cemernek, "Outlier detection as instance selection method for feature selection in time series classification," 2021, *arXiv:2111.09127*.
[88] J. Hodgson, A. M. Johansen, and M. Pollock, "Unbiased simulation of rare events in continuous time," *Methodology Comput. Appl. Probab.*, vol. 24, no. 3, pp. 2123–2148, Sep. 2022.
[89] S. Aaron and Gloria. *Project Tesserae*. Accessed: Sep. 10, 2023. [Online]. Available: https://tesserae.nd.edu/
[90] DOT. *TFA Traffic Mapping Application—TDA—MnDOT*. Accessed: Sep. 10, 2023. [Online]. Available: https://www.dot.state.mn.us/traffic/data/tma.html
[91] NTLC. (2018). *2018 Yellow Taxi Trip Data | NYC Open Data*. Accessed: Sep. 10, 2023. [Online]. Available: https://data.cityofnewyork.us/Transportation/2018-Yellow-Taxi-Trip-Data/t29m-gskq
[92] A. Abbes, H. Essid, I. Farah, and V. Barra, "Rare events detection in NDVI time-series using Jarque-Bera test," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 338–341.
[93] C. A. Escobar, R. Morales-Menendez, and D. Macias, "Process-monitoring-for-quality—A machine learning-based modeling for rare event detection," *Array*, vol. 7, Sep. 2020, Art. no. 100034.

YAHAYA IDRIS ABUBAKAR is currently pursuing the Ph.D. degree in artificial intelligence with Laboratoire Images, Signaux et Systémes Intelligents (LiSSi), Université Paris-Est Créteil (UPEC), France. His research focuses on computer vision, images, and rare event detection.

ALICE OTHMANI has been an Associate Professor with Université Paris-Est Créteil, since 2017. She has been with several international institutions, such as Ecole Normale Superieure de Paris, College de France, and the Agency for Science, Technology and Research (A*STAR), Singapore. Her research work concerns developing computer vision and artificial intelligence solutions for healthcare, emotional intelligence, and psychiatry.

PATRICK SIARRY received the Ph.D. degree from University Paris 6, in 1986, and the Doctorate of Sciences (Habilitation) degree from University Paris 11, in 1994. He was first involved in the development of both analog and digital models of nuclear power plants with Electricité de France (E.D.F.). Since 1995, he has been a Professor of automatics and informatics. His primary research interests include the computer-aided design of electronic circuits and the application of new stochastic global optimization heuristics to various engineering fields. Additionally, he is interested in fitting process models to experimental data, learning fuzzy rule bases, and neural networks.

AZNUL QALID MD SABRI received the Ph.D. degree from the University of Picardie Jules Verne, Amiens, France. He is currently an Associate Professor with the Department of Artificial Intelligence, Faculty of Computer Science and Information Technology (FCSIT), Universiti Malaya. His main research interests include computer vision, robotics, and machine learning.