Jurnal Kejuruteraan 37(5) 2025: 2215-2229
https://doi.org/10.17576/jkukm-2025-37(5)-15

Abnormal Activities Detection along Oil Pipeline Using Deep Learning

Yau Alhaji Samaila[a,b]*, Patrick Sebastian[a], Syed Saad Azhar Ali[c,d], Aliyu Nuhu Shuaibu[e,f], Sulaiman Adejo Muhammad[e] & Abba Muhammad Adua[g]

[a]*Universiti Teknologi Petronas, Faculty of Engineering, Department of Electrical and Electronics Engineering, 32610, Bandar, Seri Iskandar, Perak, Malaysia.*

[b]*University of Maiduguri, Faculty of Engineering, Department of Electrical and Electronics Engineering, Maiduguri, Nigeria.*

[c]*Aerospace Engineering Department, King Fahd University of Petroleum & Minerals*

[d]*Interdisciplinary Research Centre for Smart Mobility and Logistics, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia.*

[e]*University of Jos, Department of Electrical and Electronics Engineering, Jos, Nigeria.*

[f]*Department of Electrical, Telecoms and Computer Engineering, Kampala International University, Uganda.*

[g]*Abdullahi Fodio University of Science and Technology, Faculty of Engineering,Department of Electrical and Electronics Engineering, Aliero, Kebbi State, Nigeria*

[*]*Corresponding author: yau_22000515@utp.edu.my & yausamaila002@gmail.com*

Received 27 November 2024, Received in revised form 19 February 2025
Accepted 19 March 2025, Available online 30 August 2025

ABSTRACT

Abnormal activities like oil pipeline vandalism need to be identified promptly. Manual surveillance systems for oil pipelines use ground team surveys, while CCTV Cameras are employed in semi-automated surveillance to detect those abnormal behaviours. Oil pipeline failure resulting from vandalism has detrimental effects on both humans and the environment. Despite the availability of the current technologies, escalating incidences of vandalism occur, prompting the necessity for computerized monitoring techniques. Computerized solutions that use deep learning networks require an enormous quantity of information for their implementation. The popular UCF Crime dataset is meant to detect generic vandalism and other anomalies of a similar or divergent nature. Hence, a dataset explicitly designed to complement such a model and assist in pipeline monitoring is needed. This work aims to investigate and develop a behaviour recognition model and a new dataset named Vandalism Detection Dataset 2024(VDD 24) for detecting and classifying abnormal behaviours along oil pipelines. A Modified pre-trained ResNet18 is used for feature extraction, and a Bi-directional long-short-term memory (Bi-LSTM) is employed to detect and categorize those human actions as normal or abnormal (vandalism). Digging, sawing, hammering, and stone impact are regarded abnormal, while activities such as walking, running, and cycling are defined as normal. Despite the similarity in the abnormal actions in the dataset, the model was able to detect and classify the anomaly. Experimental results reveal that our model's performance on VDD 24 is significant, with an accuracy of 82.5%. The model is further validated on the UCF-Crime dataset with an impressive performance.

Keywords: Abnormal activities; Oil pipeline; ResNet18+Bi-LSTM; deep learning; VDD 24

INTRODUCTION

Anomaly detection or Abnormal event detection identifies data points, events or patterns that deviate from the normal range. Anything that falls outside the standard or expected points is called an anomaly. Anomaly detection is also known as outlier detection, abnormal activity detection or novelty detection. Generally, activities like walking, running, standing, and sitting are typically considered normal. In contrast, odd and uncommon behaviours like stealing, robbery, fighting, theft, and vandalism are well-thought-out as abnormal. Pipeline system is a medium of transporting crude oil, natural gas, industrial chemicals, and other related products, in which unattended problems like vandalism result in unconceivable catastrophe in its

operation. Vandalism, as used in this work, means illegitimate acts that damage or destroy pipelines. It is an illegal move meant for obtaining oil products for personal consumption or for sale in the shadow economy, particularly in third-World Countries like Nigeria, where they are ubiquitous. A surveillance system is critical for safeguarding such pipelines. Since ancient times, the necessity of putting in place appropriate safety mechanisms for pipeline management has been acknowledged (Ghazal et al. 2007). Manual surveillance system using ground team surveys is time-consuming, labour-intensive and ineffective. A semi-automated surveillance system using CCTV cameras is also not reliable. A person monitoring a multi-camera (Amosa et al. 2023) can miss a more significant percentage of the activity (R. J. Nayak & Chaudhari 2023). An intelligent surveillance system (ISS) should detect abandoned objects (Samaila et al. 2020) and unusual actions or behaviours in public and restricted places for better efficiency and reliability.

An automated video surveillance system complements the surveillance process by integrating data from CCTV cameras with a trained model. Figures 1(a) and (b) show block diagrams of the semi-automated and automated surveillance schemes, respectively. Intelligent system that uses Machine learning (ML) processes such as Decision tree, k-means clustering, artificial neural network (ANN), Shallow auto-encoders, etc., detects data anomalies by learning the underlying pattern and then identify any deviation from it. Deep Learning (DL), a subtype of ML, simulates the human brain's intricate ability to make decisions using multiple layers of neural networks. CNNs, DNN and RNNs are three instances of DL (Zamani et al. 2023)(Khalil, 2024). DL requires vast data to learn an underlying pattern and detect anomalies. Thus, intelligent systems for smart video surveillance mainly depend on ML/DL (Janiesch et al. 2021).

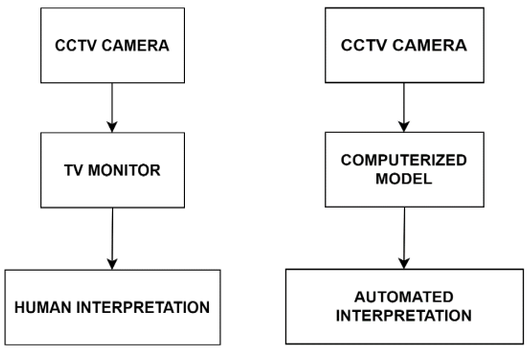


FIGURE 1. Surveillance systems (a) semi-automated surveillance system (b) automated surveillance system

ISS for video anomaly detection (VAD) comprises four primary stages- the input stage, feature extraction, model building and the output. The input data can be written, visual, audio, or data sequences. Finding pertinent feature vectors from the data (films) that can represent the characteristics of the behaviors is known as feature extraction. Feature extraction could be handcrafted/partly automated in the case of shallow machine learning or fully automated in the case of DL. The model building uses the extracted feature in the preceding step for classification, regression, etc., depending on the targeted model. To determine if a video is normal or aberrant, classification looks for certain patterns. Shallow machine learning makes the feature extraction, followed by model building/ classification. Contrarily, DL architectures such as CNN, RNN, Auto-encoders, and Generative adversarial networks (GAN), to mention a few, are primarily set up as an end-to-end system that integrates both features in one (Janiesch et al. 2021). Nevertheless, DL can also be used to extract a feature representation, which can then be fed into other machine learning algorithms such as SVMs and Long-short-term-memory (LSTM) classifiers. A probability ranging from zero to one or another entity could be the outcome. (Janiesch et al. 2021). Reassurance regarding the effectiveness of the framework can be obtained by evaluating its quality using performance indicators like precision as well as accuracy. The method of developing an analytical model is shown in Figure 2.

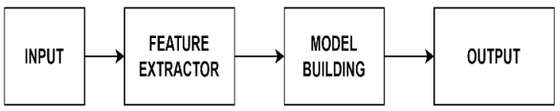


FIGURE 2. Block diagram of analytical model building for ISS

Numerous strategies are employed to detect general and specific anomalies in VAD. Sultani et al. (2018) developed a video (weakly labelled) based anomaly detection which treats normal(negative) and abnormal (positive) films as bags and video fragments. The fragments use samples in a bag. Deeper multiple instance learning (MIL) ranking loss is used for training the algorithm and detecting anomalies utilizing samples in a bag. They validated their algorithm using the UCF Crime dataset. The use of algorithms that identify particular anomalies is limited, according to the authors, because they cannot be adapted to identify other anomalous events (Sultani et al. 2018). The point is that most of the existing datasets and algorithms do not consider specific anomalies, where peculiarities of that anomaly are of utmostHence, there is a need to develop a dataset that takes specific

anomalies into cognizance, such as oil pipeline vandalism. This dataset will enable us to design and evaluate a better solution for an oil pipeline monitoring system. We offer the following in this paper:

1. We developed a Vandalism Detection Dataset 2024(VDD 24) meant explicitly for oil pipeline vandalism detection. Other researchers will utilize this dataset to carry out similar research.
2. We proposed a DL framework that specifically detects oil pipeline vandalism based on modified pre-trained ResNet18 and Bi-LSTM.

The order of the remaining portion of the paper is as outlined below: The literature pertaining to this study is included in the next part, which is followed by the research methodology. The findings and discussion are presented. This research is concluded in the final section.

RELATED WORKS

One of the most extensively studied topics in machine vision is anomaly detection, which has many facets, from context-dependent anomalies to dataset annotations (Sultani et al. 2018). VAD aims at using a computer vision system to monitor all surveillance cameras and detect anomalous events. The quest for a general method that can address every anomaly detection issue is currently ongoing. Most importantly, the identification and pinpointing of anomalies in motion picture depend on two critical factors: the types of abnormalities and the complexities of the environment (Samaila et al. 2024). Many researchers have used traditional, ML and DL methods to address human behaviour recognition/ anomaly detection (Jayaswal & Dixit 2021). Others resort to using Electronic-based methods to detect anomalies/vandalism.

A number of algorithms have been created to identify irregularities in videos. Additionally, Ofualagba and O'tega Ejofodomi (2020) created an automatic petroleum and natural gas Network Protection that uses specially made sensing units (SUs) to identify underground sound vibrations specific to ground excavation and pipeline damage. SUs are two hundred meters apart and buried underneath the earth's surface above the subterranean pipeline. The Mega version of the Arduino controller circuit uses a specially designed algorithm to process acquired seismic signals, classifying actions occurring at the earth's surface into vandalism (such as drilling and digging) and non-vandalism (such as walking and jogging). When the SU notices vandalism, it sends notification via the patrol officers' cell phones in police cars that are 20 kilometers away (Ofualagba and O'tega Ejofodomi 2020). Ghazal et al. (2007) A technique for identifying vandalism in current time video surveillance was presented by Ghazal et al. (2007). Without requiring recognition of an object, vandalism is identified by retrieving more reliable high-level events. According to the suggested approach, vandalism is committed when an object enters the location and modifies an area that is vulnerable to vandalism without authorization. The results of tests conducted both online and offline demonstrate that the suggested method is capable of differentiating between normal and vandalism actions. The integrity of the researchers' work was not supported by any clear quantitative findings. Nyajowi et al. (2021) proposed a vandalism detection technique. In order to identify the image of Criminals (humans), the researchers used a CNN-LSTM blended machine learning algorithm that feeds a picture detector to a taught photo identification network. The Image-Net dataset is used for training and assessing the algorithm being used. According to the findings, the model outperformed an independent CNN with an accuracy in training of 98% (Nyajowi et al. 2021). The authors did not use conventional data to compare their work with the existing ones. A technique for automatic pipeline identification with location definition and distant surveillance was accomplished (Samuel et al. 2014). In order to notify the managers of the pipeline via SMS and email alerts of vandals' early infiltration into the infrastructure, a passive infrared detector was employed. This allowed for the initiation of preventive measures, such as closing the pipeline valves or contacting the security patrol team, to minimize damage, leading to monetary harm and ecological deterioration. No dataset has been used, and no comparative analysis with related previous studies has been done.

Ahmed et al. (2017) offered a long-term solution and presented a multi-agent-based security system, preventing and controlling pipeline vandalism. The framework suggests that Nigerian pipeline vandalism can be managed through the use of Multi-Agent System (MAS) technology. Multiple agents that communicate over established channels and have the ability to influence their surroundings make up a multi-agent system. The MAS programming made use of a program platform called JADE (Java Agent Development Framework), which is entirely built in the Java language. Sensors built into the system identify anomalous pipe flow and the existence of metallic items in proximity to pipelines, and subsequently transmit a text message notification to the pipeline inspector's cell phone. The researchers did not provide any qualitative or quantitative evidence to support their conclusions. Jayaswal & Dixit (2021) used an algorithm that can categorize anomalous acts and created the Human Behavior Dataset 2021 (HBD21), a dataset for detecting anomalous

behavior. The authors employed an Xception model to extract the relevant characteristics from the audiovisual frames and an LSTM for behavior classification. The model was trained on HBD21 and achieved 97.25% accurateness. The authors use different data to compare their work with the cutting-edge techniques. Kommanduri and Ghorai employ inception network, followed by a fully connected layer and a classifier to determine if extracted frames are normal or aberrant. The inception encoder network is superior at understanding complex and high-level information in video footage. This network flawlessly applies its capabilities to multimedia information processing, assisting in the finding of anomalies by recognizing deviations from typical patterns. The study outcomes verify the success of their strategy, with remarkable AUC scores of 98.9%, 99.6%, and 98.1% on the Ped1, Ped2, and Avenue datasets, correspondingly. Furthermore, the technique runs at a significantly lower computing expenses, making it an appealing choice for practical uses requiring rapid and precise anomaly recognition. When unexpected abnormalities arise, the technique may encounter difficulties with significant costs. (Kommanduri & Ghorai, 2024). Using camera-based and surveillance footage, a reliable automated intelligence platform was put in place to identify and categorize particular crime types (vandalism, burglary, and arson) via SMS service in Homeowner Association (HOA) communities, parking lots, and apartment buildings (Gao et al. 2024). The clips came from a variety of data sources, including YouTube, HMDB51, and UCF. To identify the criminal scene in the motion picture, five distinct machine learning networks were trained: SSD MobileNet and YOLOv6 for Vandalism, Faster RCNN and YOLOv7 for Burglary, and Improved YOLOv5 for Arson. The concept offers communities security devoid the need for intervention from humans and creates reasonably priced surveillance cameras that can identify and categorize criminal activity. For Arson detection, the YOLOv5 achieved a mAP@50 FPS of 0.80 (as compared to the previous mAP of 0.5). For burglary, the YOLOv7 shows a better result, achieving a mAP@50 FPS of 0.87 (compared to the previous mAP of 0.61). Last of all, for vandalism, the YOLOv6 maintained an outstanding mAP@50 FPS of 0.86(compared to the previous mAP of 0.67). The authors claimed that there is no existing work for comparison (Gao et al. 2024).

Sultani et al. (2018) used weakly labelled training for detecting anomalies in videos. They use the convolutional 3-dimensional (C3D) network feature extractor-based MIL approach to get anomaly scoring for classifying the behaviours in videos. An AUC value of 75.41% was achieved on the large dataset they developed, the UCF Crime dataset. The method ignores important vital temporal dependencies. The authors advocated the use of inflated 3-dimensional (I3D) network features for improved performance (Sultani et al. 2018). An improvement of the MIL approach is presented by Tian et al. (2021), which takes into account important video temporal dependencies ignored in the MIL approach. A robust temporal feature magnitude (RTFM) is used to learn or differentiate abnormal instances (video snippets) from normal ones. Broad investigations reveal that the RTFM-enabled MIL model beats numerous contemporary methods by a significant edge, with an AUC of 91.51% on ShanghaiTech, an AUC of 83.28% on UCF-Crime, an AUC of 75.89% on XD-Violence and 98.6% on UCSD-Peds datasets. In order to tackle weakly labelled datasets challenges, such as noisy labelling and frame-level labelling in VAD, Nayak & Chaudhari, (2023) used C3D and I3D feature extractors, followed by the MIL method, to improve anomaly detection. They used the custom loss function coupled with the MIL approach to detect frame-level anomalies in video-level labelling and teach the system to differentiate between behaviours. They conducted extensive experiments on the UCF Crime dataset, employing C3D and I3D features to assess the ability of their method. The outcomes demonstrate that with I3D features, the model achieves 84.6% frame-level AUC score for the UCF Crime dataset and 92.27% frame-level AUC score for the ShanghaiTech dataset, which are comparable to previous approaches tested on similar data. The system seldom fails to generates alert when anomalies are present. Feng et al. (2021) create a multiple instance self-training framework (MIST) to efficiently modify task-driven discriminatory structures using just video annotations. The basic concept behind MIST is to create a two-step working alone technique that trains a task-specific feature encoder for finding anomalies in video footage. Each component of the framework can be replaced with any other cutting-edge component for example, C3D with I3D or a more powerful pseudo label generator to replace the multi-instance one. Furthermore, their approach can be applied to other problems, such as weakly supervised video actions localization. Simulations on the UCF Crime and ShanghaiTech datasets show the efficiency of their strategy when compared to similar works, with a frame-level AUC of 94.83% on ShanghaiTech. The method should be extensively evaluated.

To categorize human actions as normal or suspicious, Barsagade et al. (2023) use CNNs. First, the CNN is utilized to extract high-level information from images, and then the prediction is made. Walking, jogging, hand waving, and other common activities were classified as normal. Suspicious activities include running, boxing, and fighting, among others. The researchers trained their model on the KTH dataset and the CCTV-Fights dataset before creating their own data for testing. Results indicate 98.57% accuracy on the earlier mentioned dataset on their method. The

authors failed to compare their work with any of the existing state of the art. Another study suggested a DL configuration for anomaly detection by combining ResNet-50/34/18 and SRU networks. They utilized the UCF-Crime and binary datasets. Their findings indicate an average (of the three models) AUC and accuracy of 90.20% and 90.19%, respectively. They held that, the division of the dataset into UCF Crime and Binary majorly contributed to the improved performance contrasted with cutting-edge approaches. The addition of attention layers to the CNN + SRU will yield a more improved result (Qasim & Verdu 2023). Pathak et al. used four deep learning models, namely VGG16+CNN, VGG16+Bi-LSTM, Slow Fast Network, and DenseNet121, to identify abnormal actions. The Slow Fast Network (99% accuracy) and VGG16 CNN (98%) models performed better than the other existing models in classifying the anomalies from the UCF Crime Dataset. Results show that their approach outscored other existing models, with room for improvement (Pathak et al. 2024).

METHODOLOGY

The proposed method for abnormal event detection or vandalism detection (VD) along oil pipeline consists of two stages. A dataset named vandalism detection dataset (VDD 24) was developed. Several pre-processing activities were performed to prepare the data for the two steps of the method stated. The first step consists of data acquisition and development, followed by feature extraction. The feature extraction uses a modified pre-trained ResNet18 network to extract deep features from the input videos. The feature vectors obtained from the ResNet18 are fed to the Bi-LSTM to learn or train the model. The second step is the classification stage, which utilizes the Bi-LSTM network. ResNet18 is equipped with hierarchical feature extraction potentials, and Bi-LSTM's (RNN) possesses temporal consciousness. This hybrid which is similar to (Fazlića & Tahirb 2024) improves anomaly detection efficiency in time series data(video) (Syamsundararao et al. 2024.). The flow chart and architecture of the VD model are shown in Figures 3a and 3b, respectively.

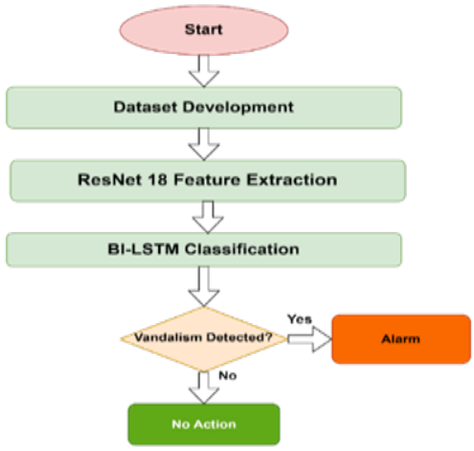


FIGURE 3. (a) Flow chart of the VD model

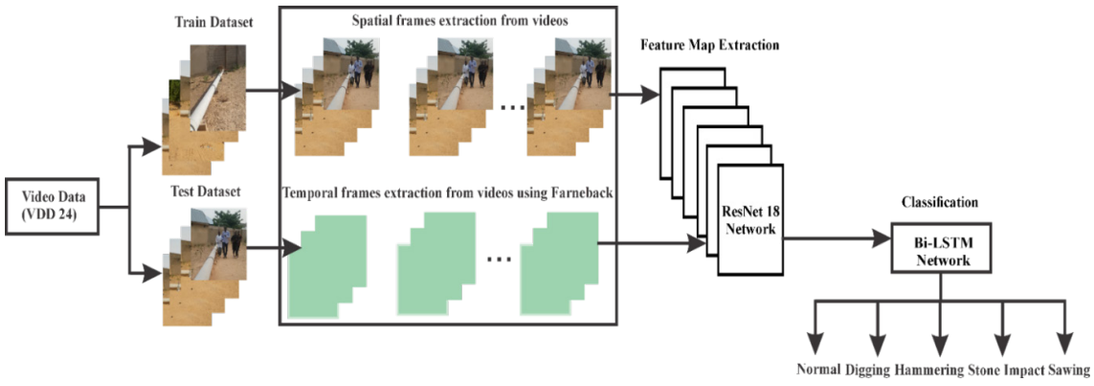


FIGURE 3. (b) Architecture of the VD model

The process for developing the proposed model are as follows:

Step 1: Pre-process the VDD 24 dataset, i.e., resize the video to 224x224x3.
Step 2: Upload the pre-processed video dataset into the MATLAB environment
Step 3: Extract temporal frames for each video using the Farneback algorithm.
Step 4: Extract spatiotemporal feature maps using the pre-trained ResNet18.
Step 5: Frame-wise annotation of the videos.
Step 6: Lastly, classify the features of the videos into five classes using the Bi-LSTM classifier.

DATASET DEVELOPMENT

A reliable benchmark dataset is vital to assess and compare the VAD system's performance. Factors such as labelled data, action type, and sample dimension are important when building a comprehensive dataset. Researchers have produced many anomaly detection datasets from surveillance cameras placed in various locations, which are divided into training and testing sets.

PREVIOUS DATASETS

The existing and prominent datasets for VAD are the UCF Crime, Shanghai Tech, UMN, CUHK Avenue, UCSD, Subway, CAVIAR datasets to mention a few. The UMN dataset consists of a single action(running), and the Avenue dataset contains short videos with some unrealistic anomalies(Sultani et al. 2018). The Shanghai dataset has only temporal ground truth annotation (R. Nayak et al. 2023). Despite being the biggest, generic and most popular dataset, the UCF Crime has significant processing requirements.

Other datasets, such as the CAVIAR, Subway, and UCSD, are not better than UCF Crime. These datasets are briefly explained here.

1. UCF CRIME: Sultani et al. (2018) A brand-new, sizable dataset consisting of 128 hours of videos was provided by Sultani et al. (2018). The dataset consists of 1900 uncut, actual events CCTV recordings that include normal activity and 13 anomalies, including robberies, vandalism, and fighting, to name a few. In VAD, it is among the most frequently utilized datasets.
2. Shanghai Tech: This all-inclusive dataset covers 13 different subjects, each captured with different lighting and camera viewpoints. It comprises around 270,000 training frames and 130 anomalies, with the pixel-level ground truth of these incidences thoroughly recorded. The introduction of fist fights and chases makes the dataset unique and suitable for real life situation (Pathak et al. 2024).
3. UMN: This is a dataset from the University of Minnesota. The dataset includes normal and anomalous crowd videos. Each video starts with normal behaviour and ends with series of abnormal behaviour. Despite the enormous amount of abnormal behaviour scenarios, only the panic one is incorporated in this dataset, which is not practical in real-life(Berroukham et al. 2023).
4. UCSD: The Peds 1 and Peds 2 security camera recordings of two different busy pedestrian path scenarios make up the UCSD dataset. Bicyclists and skaters are among the usual and unusual activities that take place on the trails. An anomalous/unusual event is sometimes defined as a pedestrian movement in an unexpected manner(Yadav and Kumar 2022).
5. Subway: This dataset contains meticulously documented incidents that were recorded at a subway station's entrance and departure gates. The departure gate video recording lasts forty-three minutes and has 64900 frames, whereas the one for the entrance gate video series is a full hour and 36 minutes long and contains 144249 frames. Walking in the wrong direction is among the acts that is considered anomalous(Berroukham et al. 2023).
6. CAVIAR: It contains video clips of people performing nine activities in two settings: the entry foyer of the INRIA Laboratories in Grenoble, France, and a supermarket in Lisbon. The videos have a resolution of 384x288 pixels, at 25fps, and are compressed using MPEG2(Chaquet et al. 2013)(Patil & Biswas, 2016).

OUR DATASET

Datasets such as the UCF Crime can be used for vandalism detection. However, it is not meant to detect oil pipeline vandalism. VDD 24 was created as a result of the growing necessity to provide a dataset that takes these anomalies

into account and to compliment the VD model. It should be noted that the VD model is a DL framework that specifically detects oil pipeline vandalism. It is based on a hybrid of ResNet18 and Bi-LSTM.

The dataset includes artificially produced videos shot using a video camera at various times of the day. Videos were taken both near and far from the oil pipeline. In the scenes used to build the dataset, trained volunteers displayed normal and anomalous behaviors. The four abnormal acts taken into consideration are sawing, digging, hammering, and stone impact. Activities like sitting, running, and crossing the pipeline are all considered normal. Samaila et al.(2023) innitially presented VDD 24 in the developmental stage; this work improved upon that work. Sample scenarios from VDD 24's typical activities are displayed in Figure 4. Figure 4(a)-(d) display activities close to or on the pipeline, which are tagged as vandalism. On the other hand, Figure 4(e)-(h) are considered normal actions. Table 1 describes the VDD 24.

FIGURE 4. Sample Scene from VDD 24 (a)-(d) show the acts of vandalism (a) sawing, (b) digging, (c) hammering, and (d) stone impact., (e)-(h) show normal actions (e) crossing the pipeline (f) running (h)hsitting and (h) cycling.

TABLE 1. Description of the VDD 24

VideoType	Video Description	AverageLength (secs)	No of Videos
Normal	Running, walking, crossing, etc.	25	147
Anomaly	Sawing (110) Digging (127) Hammering (131), Stone impact (155)	25	523
Total			670

The purpose of labelling is primarily to recognize objects and actions in a video. The annotation is carried out by experts who have a clear idea of the anomaly considered(vandalism). To avoid errors in the process, the annotators work separately first and then collaborate to validate their work. This verification is critical since the annotation impacts the dataset's health, which influences the algorithm's efficacy. The persons assigned to the task watch the videos and annotate them based on a frame basis, taking into consideration the start and end points of anomalies in the videos.

The VDD 24 collection contains 670 videos

spanning more than 4 hours and shot at 25 frames per second (average), depicting both normal actions and vandalism acts.Similar to (Senin et al. 2024), the dataset was divided as follows: 80% for training and 20% for testing.

DATA PRE-PROCESSING

The original video is 852 by 480 pixels at 25fps. A warping operation was employed to make the videos uniform. The videos were down-sampled to reduce the size to 224x224x3 (height x weight x colour layer) without losing vital information for the feature extraction phase. The spatial frames were naturally extracted from the data, while the temporal frames were extracted using the Farneback method. The Farneback method is a technique used to estimate the optical flow between two frames in a video.

A video data store was created to retain videos organized by label, which were then uploaded into the MATLAB environment via a graphic user interface. The interface allows users to interact with the data and the model with ease using graphical elements such as buttons and icons.

RESNET 18 FEATURE EXTRACTION

The extracted temporal frames, together with the spatial frames, are sent as input to the feature extraction process. Extracting features from video data is indispensable for ML and DL models since it takes place prior to the training of the classifier or regression model; hence, it grossly influences the performance of the classifier.

A modified pre-trained ResNet18 network extracts spatiotemporal features. ResNet18 is a deep CNN (DCNN), with 4 residual blocks (R. Nayak et al. 2023). In DCNN, ResNet18 is a DL model with 11 network layers, the first nine of which are convolution layers, followed by a global average pooling layer that averaged ordered execution as shown in Figure 5a. A 224x224x3 image is needed as the network's input (where 3 is the number of channels). The input is convolved with a series of filters of different dimensions. The earliest layer of ResNet18 is a 3 × 3 convolutional layers with 64 output channels and stride 1. A batch normalization layer and ReLU activation function layer trail each convolutional layer. The residual block includes two 3×3 convolutional layers with the same output stream and stride 2. Each residual block contains double the number of streams as the previous residual block. ResNet18 has four output channels: 64, 128, 256, and 512. The output for each frame is 1 × 1 × 512. After that, we flattened each feature map to 1x512. The Bi-LSTM receives a 1 × 512 feature output map as input. In this work, the ResNet18 is utilized for feature extraction rather than classification; thus, we separated the fully connected layer and the output. In other words, the classification layer is not utilized here. The Bi-LSTM performs multiclass classification. The final layer's activation function is a global average pooling layer (pool 5).

BI-LSTM CLASSIFIER

LSTM is a kind of RNN that can understand the long-term temporal dependence of data while retaining information in the form of a sequence. One of the most viable model for detecting anomalies in videos is the LSTM.The goal of LSTM is to address the issue of long-term dependency by utilizing short-term memory. In an LSTM, information flows into and out of memory via three (3) gates. The three gates are: input gate, forget gate, and output gate.

A modification of LSTM, Bi-LSTM enhances comprehension of sequential trends by learning with both

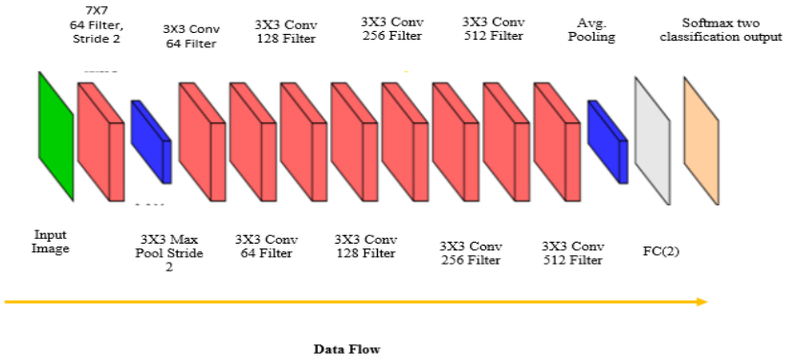
forward and backward dependencies. Two LSTM networks are stacked on top of one another in the model, which takes into account the values before and after the input before combining the outputs. In this manner, LSTM learns the effect of both prior and subsequent data for every datum. The forward sequence is used to train the first LSTM. The reverse LSTM receives the input sequence in its inverse form. In contrast to LSTM, the model's learning capacity is improved by its bidirectional nature. The computation uses values from two layers, which is very different from a typical RNN.

Equation (1) yields the value at $\hat{y}$ by calculating the values, weights, and values of conditional probability bayes derived from the states before and after input (a^{+}, a^{-}) (Tuncer & Doğru Bolat, 2022).

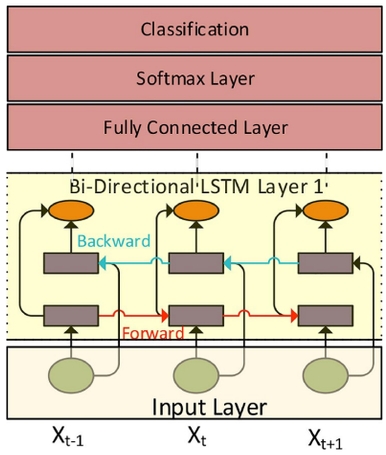
$$\hat{y}^{(t)} = g\left(W_{y}\left[a^{+(t)}, a^{-(t)}\right]\right) + b_{y} \quad (1)$$

Where a^{+}, and a^{-} are probability bayes obtained prior to and after input. $\hat{y}^{(t)}$ is the output, W and b are the weight and biases, respectively. The Bi-LSTM structure is displayed in Figure 5(b).(Tuncer & Doğru Bolat, 2022). X_{t-1}, $X_{t,}$ and X_{t+1} are the input states at time t=-1, 0 and 1, respectively.

The choice of the Bi-LSTM as the classification model over SVM and others alike (Zakariaa et al. 2022) emerged because it is better at memorizing/learning temporal dependencies (change in appearance, which translates to motion information) in both directions. This ability enables it to comprehend well the (sequential) patterns or features used in detecting and differentiating vandalism actions. The Bi-LSTM takes the 512 features vectors from the ResNet18 to classify it. In accordance with the number of classes in the VDD 24, the last layer of the Bi-LSTM contains five (5) neurons. We initially used the extracted features to train the classifier. After that, the modified ResNet18 and the Bi-LSTM were cascaded, making the model end-to-end for the rest of the process. A fully connected layer with five units was used to classify activities into five categories, and the softmax function was used as the activation function.



(a)



((b)

FIGURE 5. The hybrid VDD model (a) ResNet 18 architecture (b) Bi-LSTM structure(Tuncer & Doğru Bolat, 2022)

IMPLEMENTATION

All experiments were conducted on Windows OS with MATLAB (2024a) in the MATLAB deep learning toolbox. The train, and test proportion of the data are 80% and 20% respectively. The sequence length chosen is 450, with a 50% dropout regularization, and cross-entropy is the classification layer used. A minimum batch size of 8 and a maximum epoch of 16, with an original learning rate of 0.001, are selected in the training options. The number of dimensions (output) for the ResNet 18 is 512, and an iteration limit of 1000 was selected.

Evaluation Metrics: For this work, the confusion matrix is primarily used to evaluate the effectiveness of our algorithm.

RESULTS AND DISCUSSION

The dataset, which consists of many video folders, was uploaded into the MATLAB environment via the GUI. Afterwards, temporal frames are extracted from the videos using the Farneback method for onward feature extraction using the modified ResNet18. The Bi-LSTM uses spatial and temporal features for classification.

Figure 6 depicts a plot of model accuracy and loss. Accuracy at the top section and loss at the bottom of the Figure. The plot shows that the accuracy (validation accuracy) rose at the initial epochs and then went down at epoch 7, thus stabilizing finally at the 8th epoch. The loss (validation loss) was high and up at the first epoch, went down after that, and then rose at epoch 7 before finally stabilizing at epoch 8. The validation loss did not converge to zero. However, from the plot, we can conclude that the model neither overfit nor underfit as there is no wider variation between the training and validation. In other words, the model was able to establish an overriding trend within the data, as no training error was recorded, thus obtaining a good performance.

The network training information, such as the time elapsed, iteration, and validation accuracy, amongst others, are contained in Table 2. It can be deduced from Table 2 that the network, upon completion of training, has a final validation accuracy of 71.89% with an elapsed time of 5mins 31secs.

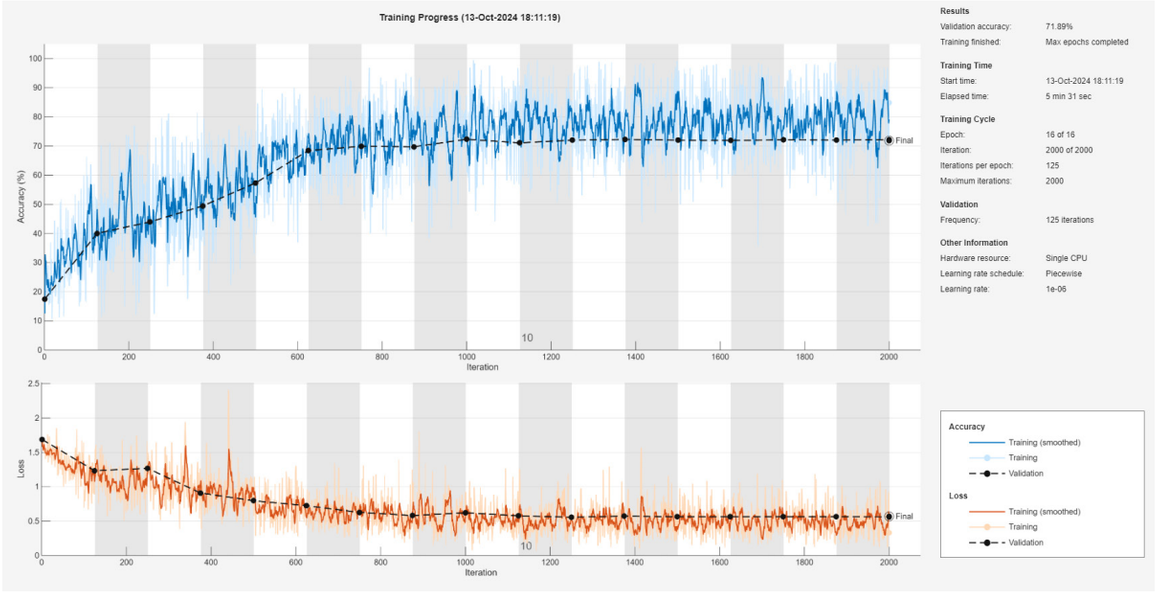


FIGURE 6. Plot of accuracy and loss of the VD model

TABLE 2. Network Training Information

Elapsed Time	Final Epoch/iteration	Final Validation accuracy	Learning Rate
5mins 31secs	16/2000	71.89%	0.000001

Figures 7(a) and 7(b) are frames showing sawing and hammering (anomaly) detected correctly by the model. Incorrect detection is noticed in Figure 7(c), where the model detected stone impact as sawing. The similarity in the abnormal actions could be the cause of this error. The train, validation and test confusion matrix of the model with true class versus the predicted class for the five actions (5x5 matrix) is shown in Figure 8. The model correctly detected 79% of the digging class, while the remaining were misclassified to other class. The algorithm successfully detected 85% of the hammering class, failing in detecting the rest of that anomaly class. Normal action recorded the best detection rate of 98% of the assigned class. Stone impact and sawing recorded the lowest correct detection of 77% and 76% respectively. It can be seen from the confusion matrix of the test set in Fig. 8(c) and the behaviour results in Table 3 that the true positive rate (sensitivity) for the normal action(s) is 0.976 for sawing is 0.757, for hammering is 0.848, for digging is 0.795 and for stone impact is 0.772. Hammering has the highest sensitivity in the anomalous class because the actions are clearer as the movement of the arms is well projected. The results from Table 3 also indicate that the model does not miss any of the five classes by producing high precision in all classes despite the similarity in the abnormal actions. The model is capable of learning generalized representation, thus capturing all abnormal classes with exactitude. Additionally, the F1-score in Table 3 depicts that the model is successfully able to understand all those activities of varying kinds, with the highest F1-score of 0.978 for normal action and then 0.825 for digging. This score is followed by hammering with an F1- score of 0.803 and stone impact with 0.774. Sawing recorded the lowest F1-score of 0.764. These results signify that the hybrid VD model has a better generalization capability to learn real-life temporal features of activities of varying forms.

Table 4 shows the overall performance of the model: the training accuracy is 92.6% (highest), the validation accuracy is 86% and the test accuracy recorded 82.5%. The result indicates a good performance of the model with an accuracy of 82.5% on VDD 24. The accuracy obtained means that the model either failed or made a wrong detection for the remaining percentages 17.5 times out of 100). The F1 scores for the model are as follows: the training set has 0.93, the validation set also has 0.86 and 0.83 is recorded for the test set. The accuracy and the FI-scores for the model signify that the model neither overfits nor underfits. The reason why the model did not perform even better could be due to similarity in the actions, leading to difficulty for the model to differentiate between those actions.

Due to limited system's computational power/ capacity, it took several hours for the feature extraction process to be completed but just 5 minutes 31 seconds to train the classifier as earlier shown in Figure 6.

(a) (b) (c)

FIGURE 7. Model detection results (a) and (b) correct detection, (c) false detection

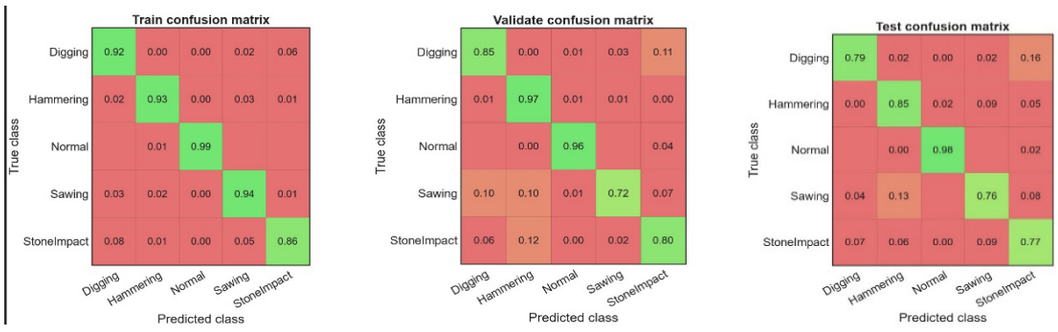


FIGURE 8. Confusion matrix of the VD model: train, validation, test on VDD 24 dataset

TABLE 3. Behaviour Results of the Model (Test Set)

Behaviour	Precision	Recall	True Positive Rate	False Positive Rate	F1-Score
Digging	0.858	0.795	0.795	0.030	0.825
Hammering	0.762	0.848	0.848	0.054	0.803
Stone impact	0.776	0.772	0.772	0.079	0.774
Normal	0.980	0.976	0.976	0.005	0.978
Sawing	0.771	0.757	0.757	0.052	0.764

TABLE 4. Overall Performance of the model on VDD 24

Set	Accuracy	F1 score
Train	0.926	0.926
Validation	0.860	0.859
Test	0.825	0.829

MODEL VALIDATION ON THE STATE-OF-THE ART DATASET

The model was tested on one of the renowned state-of-the-art dataset, the UCF Crime, for validation (Sultani et al. 2018). We select videos with actions closely similar to the ones contained in the VDD 24 dataset. A total of 248 videos of varying lengths comprising 5 anomalies, assault, burglary, road accident, explosion and vandalism, were used for the validation. Figure 9 shows the confusion matrix for the train, validation and test set. The burglary class has the highest proper detection/ classification rate, with an error of 20%. Assault is properly identified 74 times out of 100. 65% of road accidents are correctly categorized, with the remainder mistaking

for other behaviors. Vandalism and explosions had the lowest detection rates of all activities, with 51% and 49% correct detections, respectively. Based on the above results, it is important to note that, while the model was specifically created for oil pipelines, its application can be expanded to other abnormal situations.

All the anomalies mentioned were detected with an appreciable precision, as shown in Table 5. Despite the variations in the action type in this dataset compared to those of VDD 24, results indicate an impressive performance of 72% detection accuracy of the proposed method on the UCF Crime dataset as captured in Table 6.

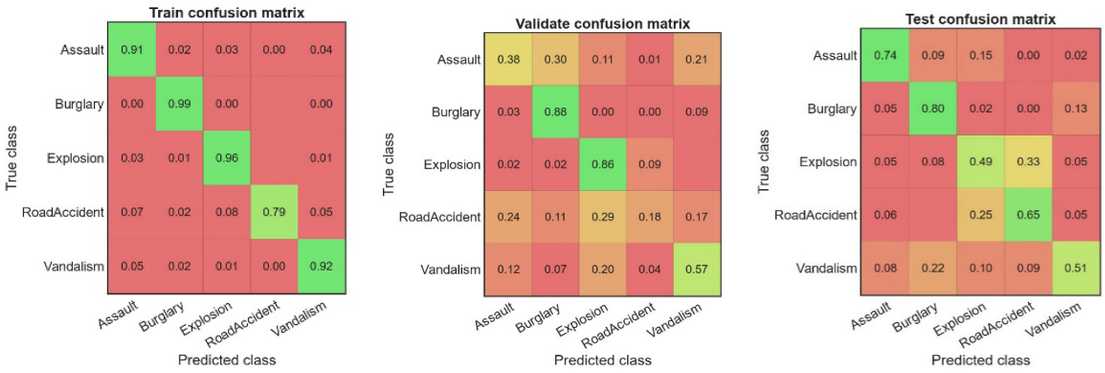


FIGURE 9. Confusion matrix of the VD model: train, validation, test on UCF Crime Dataset

TABLE 5. Behaviour Result of the model (Test Set) on UCF Crime dataset

Anomalies	Precision	Recall	True Positive	False Positive	F1
Assault	0.712	0.797	0.797	0.057	0.752
Burglary	0.773	0.824	0.824	0.165	0.797
Explosion	0.681	0.694	0.694	0.056	0.687
Road Accident	0.652	0.443	0.443	0.016	0.527
Vandalism	0.659	0.583	0.583	0.092	0.619

TABLE 6. Overall Performance of the model on the UCF Crime

	Accuracy	F1 score
Train	0.920	0.841
Validation	0.721	0.523
Test	0.720	0.677

CONCLUSION

A specific dataset and algorithm that detects abnormal activities along oil pipeline is lacking in the field of computer vision. We propose a hybrid of modified pre-trained ResNet18 and Bi-LSTM anomaly detection model alongside a novel dataset named VDD 24 that explicitly detects oil pipeline vandalism. A modified ResNet18 and a Bi-LSTM were employed for feature extraction and classification, respectively. The extracted feature vectors from the ResNet18 were tapped before the classification stage and then fed to a separate Bi-LSTM for classification into five (5) different actions: normal, sawing, hammering, digging and stone impact- corresponding to the actions in VDD 24. The results show a good performance with an average accuracy of 87%, which demonstrates that the proposed model has learnt and performed well for those complex and similar anomalous activities. This performance also reveals that the model is robust, as it was able to perform well on this new dataset. Experimental validation of the algorithm on UCF Crime dataset yielded an impressive performance with an accuracy of 79%. The accuracy obtained on the UCF Crime is lower than that obtained on the VDD 24. This is an indication that, though the model can detect other anomalies, but it is more suitable for detecting vandalism. The dataset will serve as a reference point for other researchers to draw their insight and trends, thus facilitating advancement in the field of VAD. The proposed model was also able to detect oil pipeline vandalism for occluded or partially blocked targets as well. The model will assist security personnel in focusing on vandalism areas, thereby minimizing pipeline failures and saving costs, lives, and properties. This research would be extended to consider merging some of the actions that are too similar, performing extensive validation on other state-of-the-art dataset.in the future.

ACKNOWLEDGEMENT

Appreciation to UTP Malaysia and PTDF Nigeria for supporting in conduct of this research.

DECLARATION OF CONFLICTING INTEREST

None.

DATA AVAILABILITY

Available on request.

REFERENCES

Ahmed, M. K., Umar, A. Y., & Bute, M. S. 2017. Multi-agent based architectural framework for the prevention and control of oil pipeline vandalism. In *2017 International Conference on Computing Networking and Informatics (ICCNI)* (pp. 1-8). IEEE. doi: 10.1109/ICCNI.2017.8123808.

Amosa, T. I., Sebastian, P., Izhar, L. I., Ibrahim, O., Ayinla, L. S., Bahashwan, A. A., Bala, A., & Samaila, Y. A. 2023. Multi-camera multi-object tracking: A review of current trends and future advances. *Neurocomputing* 552: 126558. https://doi.org/https://doi.org/10.1016/j.neucom.2023.126558

Barsagade, K., Tabhane, S., Satpute, V., & Kamble, V. 2023. Suspicious activity detection using deep learning approach. *1st IEEE International Conference on Innovations in High Speed Communication and Signal Processing, IHCSP 2023*, 1–6. https://doi.org/10.1109/IHCSP56702.2023.10127155

Berroukham, A., Housni, K., Lahraichi, M., & Boulfrifi, I. 2023. Deep learning-based methods for anomaly detection in video surveillance: A review. *Bulletin of Electrical Engineering and Informatics* 12(1): 314–327. https://doi.org/10.11591/eei.v12i1.3944

Chaquet, J. M., Carmona, E. J., & Fernández-Caballero, A. 2013. A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding* 117(6): 633–659. https://doi.org/https://doi.org/10.1016/j.cviu.2013.01.013

Fazlića, H., & Tahirb, N. M. 2024. Deep Learning-Based Audio-Visual Speech Recognition for Bosnian Digits. *Jurnal Kejuruteraan* 36(1): 147–154. https://doi.org/10.17576/jkukm-2024-36(1)-14.

Feng,J. C., Hong, F. T., & Zheng, W. S. 2021. Mist: Multiple instance self-training framework for video anomaly detection. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition* (pp.14009-14018). https://doi.org/10.48550/arXiv.2104.01633

Gao, J., Shi, J., Balla, P., Sheshgiri, A., Zhang, B., Yu, H., & Yang, Y. 2024. Camera-based crime behavior detection and classification. *Smart Cities* 7(3): 1169–1198. https://doi.org/10.3390/smartcities7030050.

Ghazal, M., Vázquez, C., & Amer, A. 2007. Real-time automatic detection of vandalism behavior in video sequences. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 1056–1060. https://doi.org/10.1109/ICSMC.2007.4414038

Janiesch, C., Zschech, P., & Heinrich, K. 2021. Machine learning and deep learning. *Electronic Markets* 31(3): 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jayaswal, R., & Dixit, M. 2021. A framework for anomaly classification using deep transfer learning approach. *Revued'Intelligence Arti icielle* 35(3). https://doi.org/10.18280/ria.350309.

Khalil, M. S. 2024. Deducting abnormalities in chest x-rays using Gabor Filters and Deep Neural Network (DNN). *Jurnal Kejuruteraan* 36(5):1965–1972. https://doi.org/10.17576/jkukm-2024-36(5)-16.

Kommanduri, R., & Ghorai, M. 2024. Anomaly detection in video surveillance: A supervised inception encoder approach. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-18604-2

Nayak, R. J., & Chaudhari, J. P. 2023. Real-world anomaly detection in video using spatio-temporal features analysis for weakly labelled data with auto label generation. *International Journal of Electrical and Computer Engineering Systems* 14(5): 565–573. https://doi.org/10.32985/ijeces.14.5.8

Nayak, R., Pati, U. C., & Das, S. K. 2023. A comprehensive review of datasets for detection and localization of video anomalies: a step towards data-centric artificial intelligence-based video anomaly detection. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-17889-z

Nyajowi, T., Oyie, N., & Ahuna, M. 2021. CNN real-time detection of vandalism using a hybrid -LSTM deep learning neural networks. *IEEE AFRICON Conference*, *2021-Septe*, 1–6. https://doi.org/10.1109/AFRICON51333.2021.9570902

Ofualagba And O'tega Ejofodomi, G. 2020. Automated Oil and Gas Pipeline Vandalism Detection System. *SPE Nigeria Annual International Conference and Exhibition, Virtual,* August2020.*SPE-203695-MS* https://doi.org/10.2118/203695-MS.

Pathak, A., Jayaswal, R., & Mahajan, S. 2024. Exploring deep learning techniques for abnormality detection: A comparative analysis on UCF crime dataset. *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, *2024*(3): 392–401. www.ijisae.org

Patil, N., & Biswas, P. K. 2016. A survey of video datasets for anomaly detection in automated surveillance. *2016 Sixth International Symposium on Embedded Computing and System Design (ISED):* 43–48. https://doi.org/10.1109/ISED.2016.7977052

Qasim, M., & Verdu, E. 2023. Video anomaly detection system using deep convolutional and recurrent models. *Results in Engineering* 18: 101026. https://doi.org/https://doi.org/10.1016/j.rineng.2023.101026

Samaila, Y. A., Rabiu, H., & Mustapha, I. 2020. Real-time detection of abandoned object using centroid difference method. *Arid Zone Journal of Engineering, Technology and Environment* 16(1):48–57. https://azojete.com.ng/index.php/azojete/article/view/174

Samaila, Y. A., Sebastian, P., Shuaibu, A. N., Muhammad, S. A., & Shuaibu, I. 2023. Vandalism detection in videos using convolutional feature extractor and LSTM classifier. In *International Conference on Electrical, Control & Computer Engineering* (pp. 585-597). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-3847-2_48

Samaila, Y. A., Sebastian, P., Singh, N. S. S., Shuaibu, A. N., Ali, S. S. A., Amosa, T. I., Mustafa Abro, G. E., & Shuaibu, I. 2024. Video anomaly detection: A systematic review of issues and prospects. *Neurocomputing* 591: 127726. https://doi.org/10.1016/J.NEUCOM.2024.127726

Samuel, A., Obodoeze, C., Chukwujekwu, F., & Ekene, O. F. 2014. Oil pipeline vandalism detection and surveillance system for niger delta region. *International Journal of Engineering Research & Technology (IJERT): 3*(7): 1–11. www.ijert.org

Senin, S. F., Yusufb, K., Yusufa, A., & Rohima, R. 2024. Machine learning application for concrete surface defects automatic damage classification. *Jurnal Kejuruteraan* 36(1): 21–27. https://doi.org/10.17576/jkukm-2023-36(1)-03.

Sultani, W., Chen, C., & Shah, M. 2018. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp.6479-6488). https://doi.org/10.48550/arXiv.1801.04264.

Syamsundararao, T., Gorintla, S., Nitya, E., S Raju Battula, R. S., Kongala, L., Verma, A., & Kiran, A. n.d. Anomaly Detection in Time Series Data Using Deep Learning. In *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE* 21s. www.ijisae.org

Tian, Y., Pang, G., Chen, Y., Singh, R., Verjans, J. W., & Carneiro, G. 2021. Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4975–4986. DOI: 10.1109/ICCV48922.2021.00493

Tuncer, E., & Doğru Bolat, E. 2022. Classification of epileptic seizures from electroencephalogram (EEG) data using bidirectional short-term memory (Bi-LSTM) network architecture. *Biomedical Signal Processing and Control* 73: 103462. https://doi.org/https://doi.org/10.1016/j.bspc.2021.103462

Yadav, R. K., & Kumar, R. 2022. A survey on video anomaly detection. *2022 IEEE Delhi Section Conference (DELCON):* 1–5. https://doi.org/10.1109/DELCON54057.2022.9753580

Zakariaa, N. K., Tahir, N. M., Jailania, R., & Taherd, M. M. 2022. Anomaly gait detection in ASD children based on markerless-based gait features. *Jurnal Kejuruteraan* 34(5): 965–973. https://doi.org/10.17576/jkukm-2022-34(5)-25

Zamani, N. S. M., Hoe, E. Y. C., Huddin, A. B., Zaki, W. M. D. W., & Abd Hamid, Z. 2023. Deep learning for an automated image-based stem cell classification. *Jurnal Kejuruteraan* 35(5):1181–1189. https://doi.org/10.17576/jkukm-2023-35(5)-18