IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 42, No. 6, December, 2025, pp. 3289-3307
Journal homepage: http://iieta.org/journals/ts

Enhancing Deep Feature Learning and Visual Mining for Effective Detection of Suspicious Activities in Video Surveillance

Kalyan Chakravarti Yelavarti[1,2*], Ramakrishnaiah Nagendla[1]

[1] Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Kakinada 533003, India
[2] Department of Information Technology, Siddhartha Academy of Higher Education, Deemed to be University, Vijayawada 520007, India

Corresponding Author Email: klyn.518@gmail.com

https://doi.org/10.18280/ts.420620

Received: 25 May 2025
Revised: 12 November 2025
Accepted: 1 December 2025
Available online: 31 December 2025

Keywords:
video surveillance, deep learning, classification, visual mining, spectral space

ABSTRACT

Real-time video surveillance applications need a high detection rate of suspicious activities for public safety. Deep learning and computer vision have progressed with suitable solutions in such applications. Effective learning and classification of video data highly influence the detection of suspicious activities. The dimensions of deep features are relatively high, and learning massive videos demands more computational time than traditional deep learning techniques. This paper develops the spectral based deep learning technique for obtaining surveillance video representations in a lower-dimensional or reduced representation. The proposed technique uses two key steps. The first one is to transform the deep features of video data into lower manifold spectral space; it uses visual classification approaches to detect suspicious activities. For the spectral space, the affinity values of video frames and their Laplacian matrix are computed to derive the reduced representation of video data; after deriving the lower-rank representation of video data, the similarity features of frames are analyzed with proposed visual classification techniques. Experiments are carried out on benchmarked video surveillance datasets to demonstrate the efficacy of the proposed technique compared to existing techniques.

1. INTRODUCTION

Nowadays, the detection of suspicious activities for Real-time video surveillance is becoming increasingly important for societal security applications [1]. Object identification and learning of object activities are the most challenging issues for classifying suspicious videos. Deep learning [2] is one of the most successful methods for video classification tasks. Current deep learning methods, namely, VGG16 [3], ResNet50 [4], DenseNet121 [5], MobileNet [6], GoogLeNet [7], and EfficientNetB0 [8], can be used to extract the features of video frames batch-wise and learning the video data accordingly. Practical training is needed for many videos that are useful to improve the classification rate or detection efficiency of suspicious activities in video surveillance. The deep features are extracted and later trained using classifier models for suspicious-activity classification. The extractions of deep features of each video frame are massive; Each video consists of a large number of frames, and the feature size of each frame is also massive. In such cases, processing the deep features for learning the suspicious activities may become practical scalability issues concerning the significant parameters of computation time. The curse of dimensionality is considered the primary problem in existing deep-based classifier models. Fabulous classifier models, i.e., random forest (RF) [9], support vector classifier (SVC) [10], and k-nearest neighbor classifier (kNN) [11] models, have been used in the existing deep-based classifier models. However, more computation time is needed to train the frame's deep features in the existing deep-based classifier models. Thus, the proposed trustworthy classifier models with the development of spectral-based deep classifier models for handling the scalability problem. The spectral concept initially finds the projected subspace for the high dimensional deep frame's features by finding the best eigenvector space.

The affinity matrix of frame features was initially constructed; later, the Laplacian matrix was computed. The laplacian matrix (LM) computations are mentioned in the proposed spectral-based deep classifier video model (SDC-VM). The LM explores the frame deep features in terms of a large number of Eigenvectors; the lower-rank frame's features decide the number of components ('k') of the subspace of deep features. The deep features subspace is derived by taking the first k-largest Eigenvectors. Obtained k-Eigen vectors subspace is referred to as low-dimensional manifold or spectral space. The Spectral space denotes the reduced representation of deep features of video frames. With this spectral space, the large size of deep features of frames is mapped into the lower manifold subspace without losing the video frame information. The top k-Eigen vectors are enough to define deep features of frame data rather than taking high-dimensional frame features. Finally, the spectral features of frames are used to classify videos with the classifier models of RF, SVC, and kNN. With these proposed techniques, three

hybrid variants of SDC-VM are developed using the models of RF, SVC, and KNN. Proposed hybrid variants use spectral features instead of the deep features of the high-dimensional frame to address the problem of the curse of dimensionality. Architectural diagram of the proposed SDC-VM shown in Figure 1.

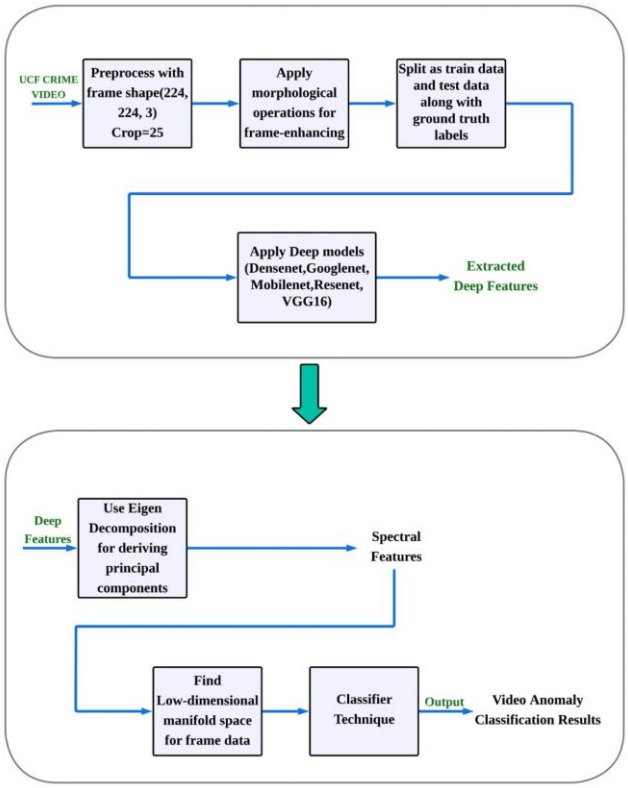


Figure 1. Proposed SDC-VM architectural diagram

Spectral-based visual mining algorithms and three hybrid variants of SDC-VM are presented in this paper to assess and classify suspicious activities in video surveillance data efficiently. Key objectives of the proposed work are to classify or detect suspicious activities for the video surveillance data faster with the SDC-VM models and assess the video surveillance data for the detection of several activities (including the normal and abnormal activities) using proposed visual mining techniques.

The proposed work is indeed inspired by classical spectral methods; the proposed approach is not a direct reuse of PCA or Laplacian Eigenmaps. Our novelty lies in the way spectral learning is integrated with deep video features, where we construct a task-specific affinity matrix based on frame-level deep feature dynamics and derive a low-rank spectral representation tailored for surveillance data. This representation captures temporal and semantic relationships that traditional dimensionality-reduction methods do not model. Furthermore, unlike standard spectral techniques that stop at embedding, we introduce a visualization-driven classification mechanism (namely, spectral-based deep classifier video model (SDC-VM)) that operates within the reduced spectral space to effectively detect suspicious activities. The overall framework forms a unified spectral–deep learning pipeline designed for real-time surveillance, reducing computational complexity while maintaining high discriminative capability.

Contributions of the work are presented as follows:

(1) The spectral features of frame data extracted for finding the lower manifold subspace of higher-dimensional video frame data to overcome the curse of dimensionality problem

(2) Three variants of the SDC-VM were developed to find the video surveillance classification results in a faster way

(3) Partially supervised visual mining techniques are implemented to assess video activities in video surveillance datasets.

The organization of the paper is as follows. Section 2 presents the literature study of the work; Section 3 presents the proposed SDC-VM and visual mining techniques. The experimental study of the work and discussion are presented in Section 4. Finally, the conclusion and future scope of the work are presented in Section 5.

2. LITERATURE STUDY OF THE WORK

Video surveillance has become one of the significant applications for public safety. The most challenging research problem in surveillance is the learning and classification of abnormal activities. In this process, classification accuracy and computational scalability are the most significant issues considered in the current video analytics research. Authors in literatures [12-16] attempted to detect or classify loitering, fighting, and abandoned objects based on the finding of suspicious behaviours. Most of these works are performed on offline video surveillance data; in real-time, the kind of post-detection of abnormalities or anomalies in videos may not be helpful. With this fact, the author in literature [17] developed the real-time blob-matching technique by finding the temporal features of blobs and activities. It can detect and classify fighting, theft, loitering, fainting, etc. Findings of the work and limitations are presented in Table 1. Long video event retrieval has been most attractive in real-time video surveillance systems. Text instruction (or caption) for long videos may also give the solution indirectly to help people know abnormal activities instantly. Lu and Grauman [18] implemented the algorithm of video summary generation that can be used to extract selected sub-videos from the long video based on the image quality factors for describing the events happening in the video. Another method is based on the key frame sequence, the event detection technique developed by Wolf [19]. It uses the frames rather than taking the video directly. It shows a significant reduction in video content and improved efficacy in video retrieval. Finding the specific moments from long video faces majorly in the above methods. Wan et al. [20] developed the superframe segmentation algorithm that removes the redundant frame in a video sequence to reduce the computational overheads when classifying video anomalies. Findings and approach details are mentioned below, along with the limitations of the work. The key literature and observations or limitations are presented in Table 1.

Dimensionality reduction methods playing the crucial role for handling the abnormalities in high-dimensional video surveillance data. Huang et. al. [21] proposed the method in study for reducing the computational complexities by transforming the high-dimensional video data into a low-dimensional notation. They proposed the variational-based subspace (VBS), which basically uses the principal component analysis (PCA) to prune the dimensions from N dimensional to required k-principal components. Central problem of determining k-components in PCA is resolved in VBS by estimation of covariance matrices for the different selections

of k values. Finding covariance matrix is very expensive due to its computations are relevant to single value decomposition (SVD). Authors in study [22] faced the challenges of dimensionality reduction and used the methods, independent component analysis (ICA), non-negative matrix factorization (NMF), autoencoders (AE), variational auto encoders (VAE), and PCA. These methods were effectively applied in their work for reducing the dimensions complexities. The ICA transforms the high dimensional data in large scale repositories to linear combinations of statistically independent vectors. The limitation of ICA is unable to separate the Gaussian resources. Thus, NMF is another better option for the subspace learning to reduce the dimensions of large scale high-dimensional data. Benefit of NMF is to find the reduced data features and it also leads to automated the cluster tendency for the targeted high-dimensional data. The problem of NMF is does not yield a unique solution. In recent years, deep learning techniques are the most recommended for obtaining the optimized features. The autoencoders (AE) [23] greatly applied for dimensionality reduction and data reduction; therefore, reduced representation of features obtained that useful optimizing the requirements. Another nonlinear dimensionality reduction, namely, T-distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) [24], is widely used in large scale dimensionality reduction for video data. The t-SNE evaluates similarity in the original high-dimensional space by transforming distances between points into conditional probabilities. For every data point, the algorithm assigns a probability distribution over all other points, where close neighbors receive high probabilities and distant points receive extremely small values. This is done by centering a Gaussian kernel at each point and adjusting its width so that the effective number of neighbors (controlled by perplexity) remains consistent across the dataset. The complete set of these probability distributions forms the high-dimensional similarity structure that t-SNE attempts to preserve when projecting the data into a lower-dimensional space. UMAP and t-SNE offer considerable flexibility, allowing their configurations to be adapted to the characteristics of different datasets. The discussion [25] provides actionable insights to guide users in selecting the most appropriate dimensionality reduction approach and tuning its key parameters for their intended use case. These recommendations act as a practical framework for achieving an effective balance between performance, computational cost, and clarity of the resulting embeddings in a variety of video scenarios. Anomaly detection for suspicious activities is the most promising in video surveillance applications for saving social people. Surveillance cameras and their dynamic activity detections play a significant role in such applications. Some of the researchers [26-28] used unsupervised ideas for anomaly detection. It requires careful classification and handling approaches for abnormalities. Rather than unsupervised techniques, deep-based classification presents impressive abnormalities classifications. Sultani et al. [29] developed the 3D convolutional network (C3D), which extracts the Spatiotemporal features and computes the score of anomalies by a 3-layer fully connected network. This leads to one crucial problem: video segmentation scalability before feature extraction due to the massive size of the video frames. Zahid et al. [30] resolved the video segmentation problem with the ensemble technique of IBaggedFCNet, in which bagging imposes stringent segmentation and uses the Inception-v3 deep classifier technique for the video classification. With the broader availability of video data from surveillance cameras and social platforms, learning the importance of trained video has become tedious. Finding an efficient automated violence detection system concerned with the vital parameters of computational efficiency and accuracy is the most needed for real-time video surveillance applications. Violence detection for more extensive videos can be optimized in references [31]; here, the keyframe selection process plays a crucial role in selecting the optimal number of frames to reduce the computational overheads. It employs ensemble classification models using long short-term memory (LSTM), bidirectional-LSTM (Bi-LSTM), and gated recurrent unit (GRU) models to enable good video classification results. Manisha Mudgal et al. proposed a smart and intelligent real-time video monitoring system [32] for efficient detection of activities, slapping, hitting, punching, etc. It models the activities using the Gaussian Mixture Models (GMMs); It also develops the universal attribute models (UAM) for deriving the super action vector (SAV) that helps to improve the accuracy of classification. Thus, GMM and UAM are majorly defined for modeling the SAVs and making the classification using SVM. Xu [32]. Cong et al. [33] another intelligent video surveillance system using the deep learning approach was also proposed. Faster R-CNN with Inception ResNet V2 was discovered to achieve the best accuracies for classifying real-time activities. State-of-the-art techniques [34-36] use the trained video frames for extraction features and learning of one or more activities, including normal and abnormal activities. For testing data, suppose it matches any abnormal activities, then mark it as abnormal; otherwise, mark it as normal. These deep learning techniques are the most prominent in computer vision and other object detection and activity recognition applications. These works recommend the two-stage work for the video classification tasks. Sun et al. [37] proposed the deep one-class model (DOC), the end-to-end deep learning model. One-class support vector machine and deep CNN are integrated to optimize the accuracy and loss parameters. Lamani et al. [38] developed an efficient hybrid technique for human action recognition, in which lightweight residual 3D CNN was built for handling the computational hurdles and effective human action recognition. Existing deep learning techniques more beneficial for extraction of frames learned features. The real-time video capturing systems produce high-quality videos (i.e., massive frames). The selection of the frames and high dimensional deep features was initially extracted. High-dimensional features require the equivalent low-dimensional manifold subspace learning to reduce the complexities. The proposed work focuses on subspace learning techniques for the frame's deep features. In the high dimensional deep features, the data sparsity problem occurred, significantly impacting the classification rate. Thus, it needed to obtain the equivalent low-manifold subspace to reduce the effect of data sparsity problems. There are wider chances to improve the efficacy-related parameter values for the video classification results. Detailed procedural details and algorithm descriptions of the proposed work are presented in the following section.

Table 1. Key literature of the work and observations

Author(s)	Methodology / Key Contribution	Limitations / Future Scope
Elhamod and Levine [17]	Developed a real-time blob matching technique using temporal features, object tracking, and semantic behavior analysis for suspicious activity detection. Experiments conducted on public datasets such as fighting, stolen objects, fainting, and loitering.	Detection is successful, but current learning approaches rely heavily on hyperparameter tuning rather than effectively leveraging semantic features.
Wan et al. [20]	Proposed redundant frame removal followed by video super frame segmentation to extract segments of interest (SOI). CNN with pre-trained VGG used for abnormal video classification.	Achieved accuracy up to 69.34% only. Real-time SOI processing needs redesign using ensemble classifiers to improve accuracy.
Huang et al. [21]	Introduced a dimensionality reduction technique for network traffic anomaly detection using statistical or projection-based modeling.	Primarily evaluated on network traffic data; limited generalization to other domains and sensitive to feature distribution shifts.
Vafaei Sadr et al. [22]	Proposed a generalizable framework combining dimensionality reduction techniques such as PCA, autoencoders, and manifold learning with anomaly detection models.	Increased computational overhead and performance highly dependent on dimensionality reduction technique and hyperparameters.
Ortiz-Perez et al. [23]	Used dimensionality reduction techniques for video data on IoT edge devices to reduce latency and computational complexity.	Edge device limitations restrict model complexity; dimensionality reduction may cause loss of fine-grained details.
Mittal et al. [24]	Compared UMAP and t-SNE techniques for dimensionality reduction and visualization of high-dimensional data.	Computationally intensive for large datasets; performance depends on hyperparameter tuning.
Zahid et al. [30]	Proposed IBaggedFCNet using pre-trained Inception-v3 and PCA for feature extraction with ensemble bagging for video classification.	Fine-grained classification required for real-world datasets instead of synthetic data.
Shoaib et al. [31]	Developed DeepkeyFrm and AreaDiffKey keyframe models with ensemble LSTM, Bi-LSTM, and GRU networks for violent activity detection.	Scalability for large-scale real-time surveillance applications remains a challenge.
Xu [32]	Introduced a Universal Attribute Model (UAM) with GMM, SVM, and k-NN classifiers for suspicious activity detection.	Deep learning models outperform GMM; deep-based classifiers should be further explored.
Cong et al. [33]	Applied Faster R-CNN with ResNet V2 for abnormal activity classification.	High-dimensional frames cause dimensionality issues, limiting accuracy to 79.9%.
Sun et al. [37]	Proposed Deep One-Class (DOC) model integrating CNN and one-class SVM for pedestrian video classification.	Performance depends on kernel choice; optimization needed for real-world scenarios.
Lamani et al. [38]	Developed a lightweight residual 3D CNN combined with SVM for real-time video classification.	Real-time performance achieved with relatively low accuracy.
Kumar et al. [39]	Used CNN, BiLSTM, and attention mechanisms for anomalous human activity detection.	Accuracy limited to 61.04% on sub UCF Crime datasets.
Ahmadi et al. [40]	Applied transfer learning for intelligent Object detection in surveillance systems.	Accuracy remains below 90% due to sparse and dynamic real-time data

3. PROPOSED SDC-VM AND VISUAL MINING TECHNIQUES

The deep models VGG16, ResNet50, DenseNet121, MobileNet, GoogLeNet, and EfficientNetB0, are used to extract frame features called deep features. The deep features are high-dimensional, and data sparsity exists. Data sparsity is a significant problem in video classification.

3.1 Spectral technique for obtaining the low-dimensional manifolds (or spectral space) for the high-dimensional deep features of video frames

The proposed work uses the spectral technique to reduce the effect of data sparsity while performing the video classification. The spectral technique consists of two key steps: i) Find the affinity matrix (or the weighted matrix), which computes the weighted matrix for frame features while considering the affinities. ii) The Laplacian matrix is computed to derive the Eigen decomposition that presents suitable low-dimensional manifolds (or subspace). The following Eq. (1) to Eq. (7) illustrate the derivation of Eigenvectors (also called spectral space) in the proposed work. The $Deep_F$ refers to the deep features for the n number of frames. The $Sigma_i$ calculates the local scale or affinity value of ith frame F_i with the k-nearest (or most similar) frame, F_k, whereas $d(F_i, F_k)$ refers to the distance with the nearest frame, i.e., distance gives the affinity value with the nearest frame. The dissimilarity values presented in d_{ij} and d_{ji} denote the corresponding frames 'i' and frame 'j' and frame 'j' and frame 'i', respectively. The $W \epsilon F_{nxn}$ denotes the affinity matrix (or weighted matrix) for the n number of frames and it can be computed with affinity values of F_i, F_j (i.e., $Sigma_i$ and $Sigma_j$).

$$Deep_F = \{F_1, F_2, \ldots\ldots, F_n\} \tag{1}$$

$$Sigma_i = d(F_i, F_k) \tag{2}$$

$$d_{ij} = d\left(F_i, F_j\right) and\ d_{ji} = d(F_j, F_i) \tag{3}$$

$$W \epsilon F_{nxn},\ W = e^{((-d_{ij} * d_{ji})/(Sigma_i * Sigma_j))} \tag{4}$$

The diagonal matrix of W is 'D', in which the diagonal values are sum of the corresponding row 'i' values, mathematically formulated as follows:

$$d_{ii} = \sum_{i=1}^{n} W_{ij} \tag{5}$$

The Laplacian matrix is required to find the Eigen decomposition, which must derive the Eigenvectors. First, k-Eigen vectors are selected to obtain the spectral space (or k-principal component subspace). The Laplacian matrix 'L' is computed using the W and diagonal matrix 'D' as shown below. Eq. (5) is used to fill in the value of diagonal matrix D.

$$L = D^{-1/2} W D^{-1/2} \quad (6)$$

Values obtained in L are normalized and most suitable for further classification (or clustering process). The size of L becomes massive, and it is not needed to represent the high-dimensional space of deep features. In L, columns represent the Eigenvectors. The size of L becomes n x n. Initial Eigen column vectors in L are usually referred to as the first largest vectors. Now, we select the k-largest Eigen column vectors for the n number deep features of frames. These can be represented with the stacks of k-largest Eigenvectors, which are presented as spectral space '*SV*' The *SV* denoted the optimal low-dimensional manifold subspace for the high dimensional deep features of *n* number of frames with the size of n x k, here k refers to the reduced number of dimensions for the spectral space or the low dimensional manifolds subspace.

$$SV = [E_1 E_2 \ldots\ldots E_k] \quad (7)$$

The $E_1 E_2 \ldots\ldots E_k$ denotes the k-largest Eigenvectors.

3.2 The spectral-based deep classifier visual model algorithm

In the proposed spectral technique of video frames, the optimal representation of the video frame's features is obtained by deriving the k-largest Eigenvectors (here, the k value refers to the number of principal components of spectral space). SV describes the optimal frame features. The procedural ideas of the proposed SDC-VM are shown in Algorithm 1. Initially, the video dataset is organized in V, consisting of each class's subset of videos. Each video can be divided into several frames and combined into 64 frames as a single batch; in the experiments, the fixed batch size was 64 frames.

Algorithm 1. SDC-VM

Input:
1. V= {v1, v2, …vm}, m -number of videos of different classes
2. BS- Batch Size, denotes the size of subset of frames for the batch processing

Output: Anomaly Classifications

Methodology
1. Takes the input of V with m different classes, where as v1 has set of videos of class 1, v2 has set of videos of class 2, …., vm has set of videos of class m.
2. Divide each class of videos of V into frames
3. Batch the frames with size of 64 for the classes of videos
4. Select the frames randomly at each class with a total size of n number of batches, including all frames.
5. Apply the deep models on n batches of frames to extract the deep features, which are to be high-dimensional.
6. Use the proposed spectral technique of video frames (per the procedure described in section 3.1) and obtain the reduced dimensions of high dimensional deep features of video data. The obtained mapped features are stored in SV as per the Eq. (1) to Eq. (7)
7. Use the SVM, Random Forest, and kNN classifier techniques for the obtained SV to deliver video anomaly classification results.

After obtaining the massive batches of frames, select random samples of n batches of frames and pass the deep input layer with the specification of batch size 24 (maximum sequence length) and size of frames as 224 × 224. These steps are illustrated from Step 1 to Step 4 in an algorithm. Step 5 uses the following deep models separately: VGG16, ResNet50, DenseNet121, MobileNet, GoogLeNet, and EfficientNetB0 to extract deep features—the deep features of each frame with 2048 dimensions. There is high data sparsity problem occurred in the deep features. Thus, the spectral technique was implemented to map the high dimensional deep features into less data sparsity spectral space. It is explained in Step 6. Finally, the converted spectral deep features of the frames are used for the training model with specified classification schemes mentioned in Step 7. It explores the best accurate video anomaly classification results with the proposed SDC-VM algorithm.

4. THE EXPERIMENTAL STUDY OF THE WORK AND DISCUSSION

The experiments are carried out on the 14 classes (Abuse, Arrest, Arson, Assault, Burglary, Explosion, Fighting, Normal Videos, Road Accidents, Robbery, Shooting, Shoplifting, Stealing, Vandalism) of benchmarked video dataset, UCF-Crime which publicly free available in reference [38]. After initiating the pre-processing, the video of each of the classes is divided into different frames (or images) and organized according to the train labels and test labels. The deep features are extracted using the Densenet, GoogLeNet, MobileNet, EfficientNetB0, Resnet, and VGG16 models. Further, these features are mapped into the lower-dimensional manifolds by deriving the principal components (or Eigenvectors) from the Eigen decomposition of the Laplacian Matrix. These reduced dimensional features are obtained in our proposed SDC-VM. The results of existing deep-based-classifier models (i.e., Deep Random Forest, Deep-SVC, and Deep-kNN) and proposed spectral-based-deep classifier video models (SDC-VM-Random Forest, SDC-VM-SVC, SDC-VM-kNN) are illustrated in Table 2. The evaluation parameters of accuracy, precision, recall, and f-score measure are depicted in reference [39].

From these experimental values of performance scores, it was noted that SDC-VM achieved a reasonable classification rate compared to existing deep-based classifier models due to the reduction of the effect of the data sparsity problem in proposed spectral-based deep classifier models. The accuracy value improved at a rate of 10 to 12% in the proposed models compared to existing models. Similarly, the precision, recall, and f-score values improved at the rate of 6% to 10%, 5% to 11%, and 4% to 12%, respectively. These are comparative illustrations among the methods shown in Figure 1 to Figure 13.

The receiver operating characteristics (ROC) curves are used to depict the classifier accuracy for the multi-class video data. In the experimental work, 14 classes of the UCF-Crime classification data evaluated with the ROC curves. The value of the area under the curve (AUC) indicates the classifier's accuracy. The AUC values are between 0 and 1. Higher values indicate good classifier accuracy. Figure 14 to Figure 19 shows the ROC curves with estimated values of AUC for the proposed models of SDC-VM-KNN, SDC-VM-RF, and SDC-VM-SVC underlying the deep features extraction with the Densenet, GoogLeNet, MobileNet, EfficientNetB0, Resnet 50, and VGG16. The ROC analysis inference achieved good video anomaly classification results using the kNN underlying

SDC-VM with all six variants of deep models. Mostly, the AUC scored above 0.5 in all variants of the proposed SDC-VM models.

Table 2. Performance comparison for the existing deep-based classifier models and proposed SDC-VM based classifier models

Performance Measure	Models	Deep Random Forest	SDC-VM-RF	Deep-SVC	SDC-VM-SVC	Deep-KNN	SDC-VM-KNN
Accuracy	DenseNet	0.8464	0.9607	0.9429	0.9893	0.525	0.6464
	GoogLeNet	0.8107	0.9786	0.8964	0.9964	0.5893	0.6571
	MobileNet	0.8786	0.9607	0.9536	0.9929	0.6107	0.6929
	EfficientNetB0	0.8107	0.8464	0.2786	0.7036	0.4071	0.5182
	ResNet	0.8607	0.9821	0.9536	0.9929	0.6107	0.6964
	VGG16	0.8321	0.9393	0.5536	0.9429	0.5464	0.6536
Precision	DenseNet	0.8614	0.9642	0.9474	0.9907	0.6185	0.7014
	GoogLeNet	0.8331	0.9799	0.9091	0.9966	0.6904	0.7361
	MobileNet	0.8847	0.9664	0.9588	0.9935	0.6522	0.725
	EfficientNetB0	0.8597	0.8317	0.2051	0.7222	0.356	0.3984
	ResNet	0.8775	0.9828	0.9588	0.9935	0.6522	0.7303
	VGG16	0.8578	0.9518	0.6119	0.9504	0.5395	0.5543
Recall	DenseNet	0.8464	0.9607	0.9429	0.9893	0.525	0.6464
	GoogLeNet	0.8107	0.9786	0.8964	0.9964	0.5893	0.6571
	MobileNet	0.8786	0.9607	0.9536	0.9929	0.6107	0.6929
	EfficientNetB0	0.8464	0.8107	0.2786	0.7036	0.4071	0.4871
	ResNet	0.8607	0.9821	0.9536	0.9929	0.6107	0.6964
	VGG16	0.8321	0.9393	0.5536	0.9429	0.5464	0.5536
F-Score	DenseNet	0.8432	0.9609	0.9436	0.9895	0.5245	0.6413
	GoogLeNet	0.8082	0.9785	0.8982	0.9964	0.5837	0.6558
	MobileNet	0.8775	0.9619	0.9545	0.9929	0.6059	0.6967
	EfficientNetB0	0.8115	0.8447	0.2135	0.6946	0.3574	0.3968
	ResNet	0.8628	0.982	0.9545	0.9929	0.6059	0.7013
	VGG16	0.8299	0.941	0.543	0.9437	0.5071	0.5267

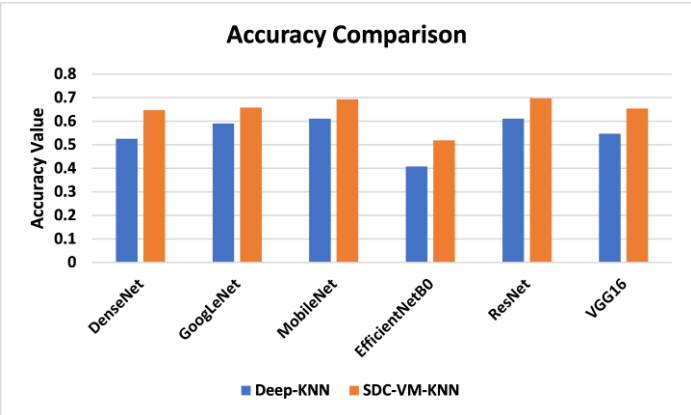


Figure 2. Accuracy between Deep-KNN and SDC-VM-KNN

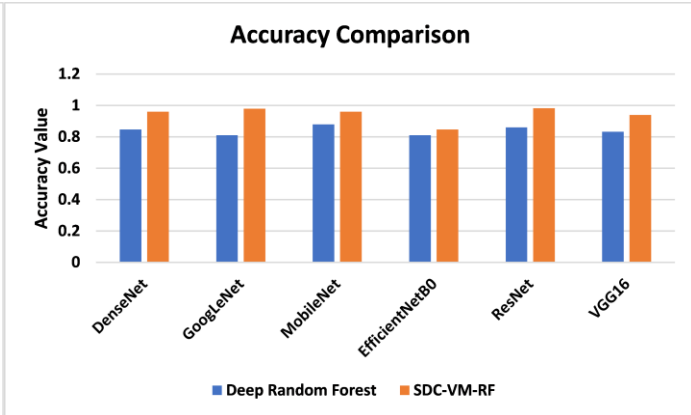


Figure 3. Accuracy between Deep-RF and SDC-VM-RF

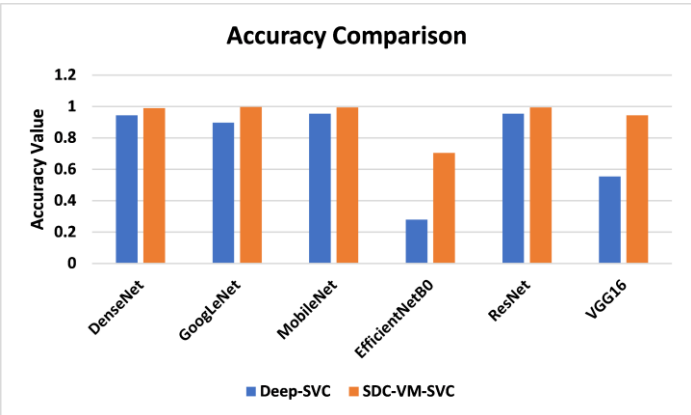


Figure 4. Accuracy between Deep-SVC and SDC-VM-SVC

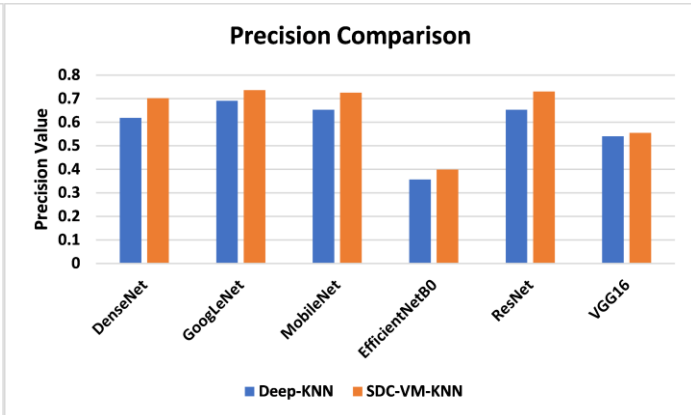


Figure 5. Precision between Deep-KNN and SDC-VM-KNN

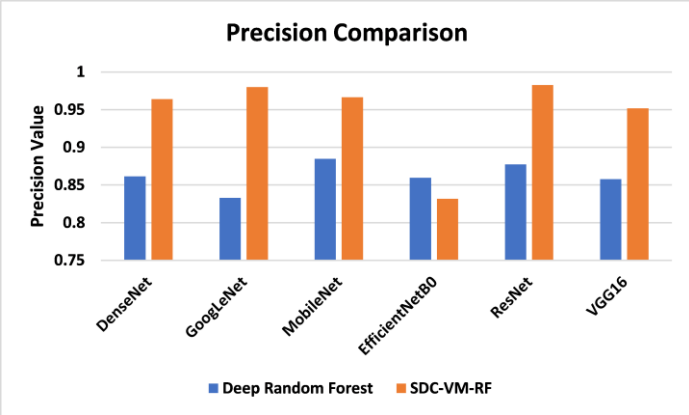


Figure 6. Precision between Deep-RF and SDC-VM-RF

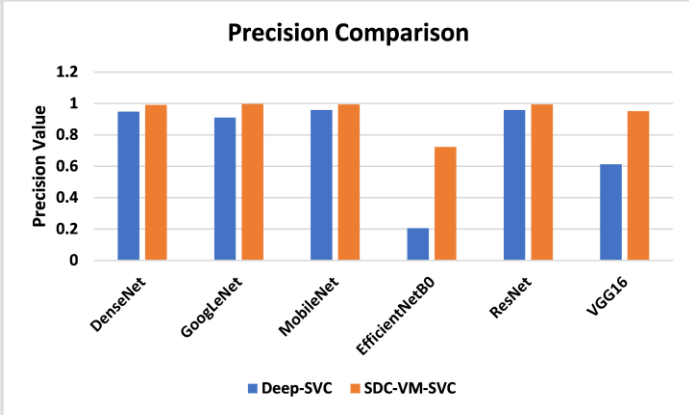


Figure 7. Precision between Deep-SVC and SDC-VM-SVC

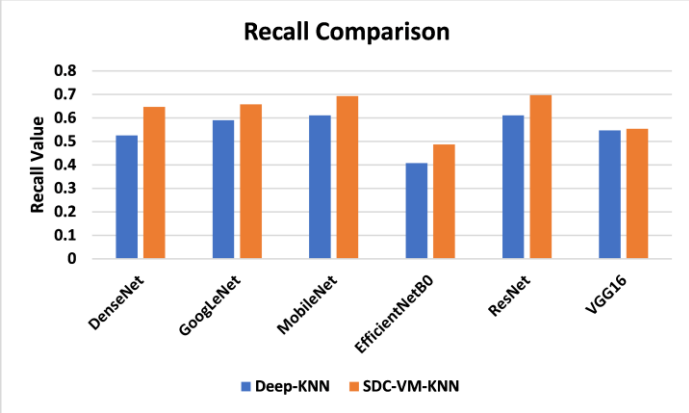


Figure 8. Recall between Deep-KNN and SDC-VM-KNN

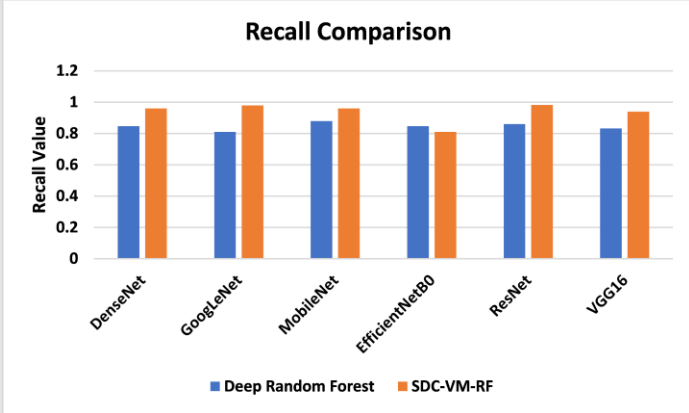


Figure 9. Recall between Deep-RF and SDC-VM-RF

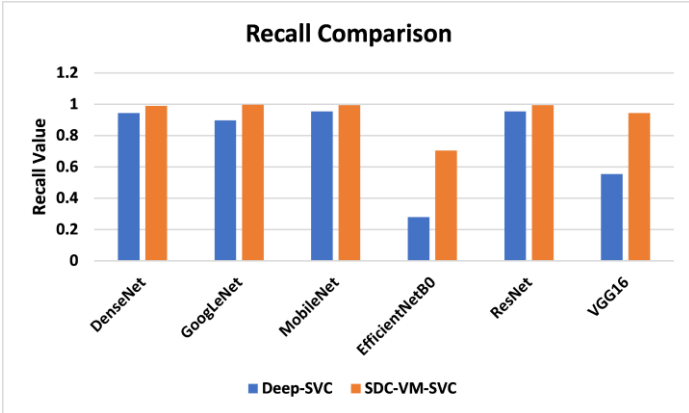


Figure 10. Recall between Deep-SVC and SDC-VM-SVC

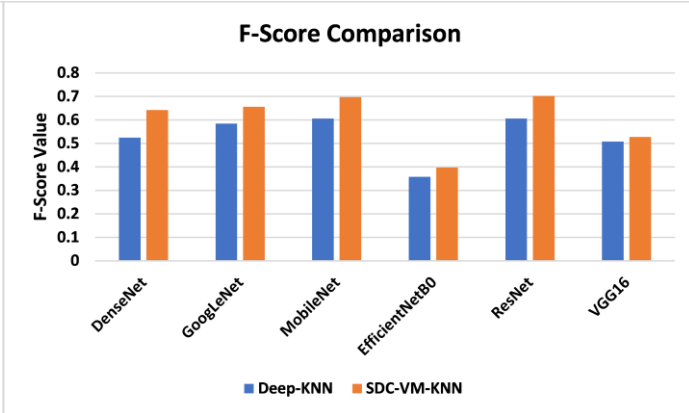


Figure 11. F-Score between Deep-KNN and SDC-VM-KNN

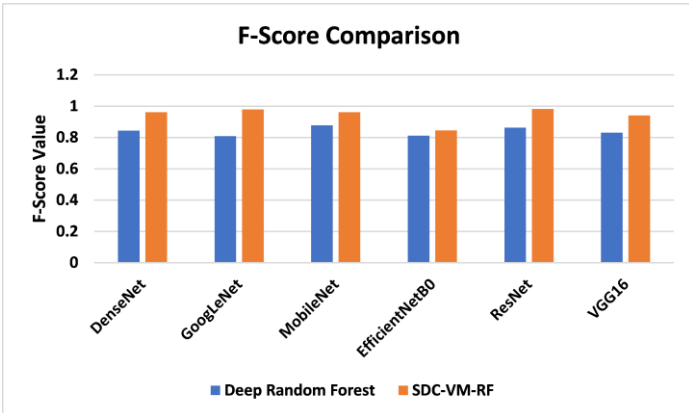


Figure 12. F-Score between Deep-RF and SDC-VM-RF

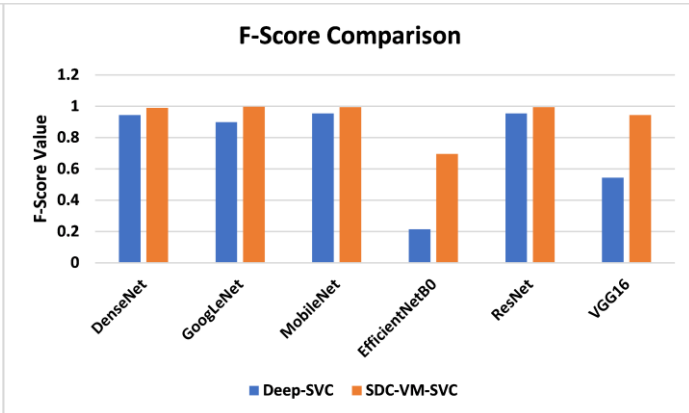
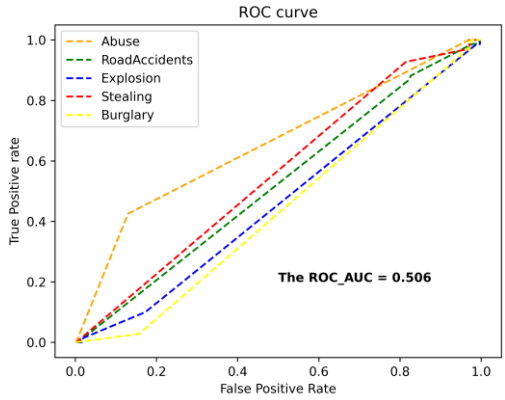
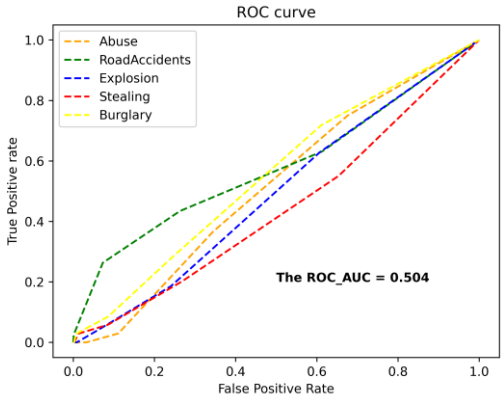


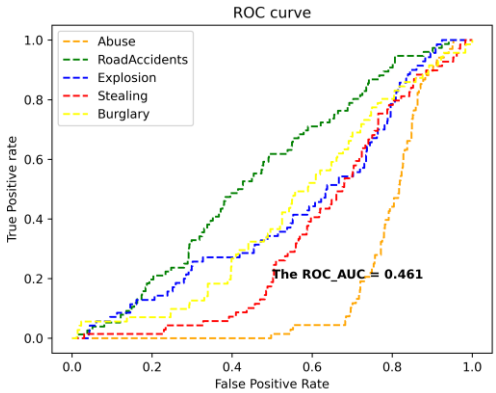
Figure 13. F-Score between Deep-SVC and SDC-VM-SVC



(a) SDC-VM-KNN ROC curve

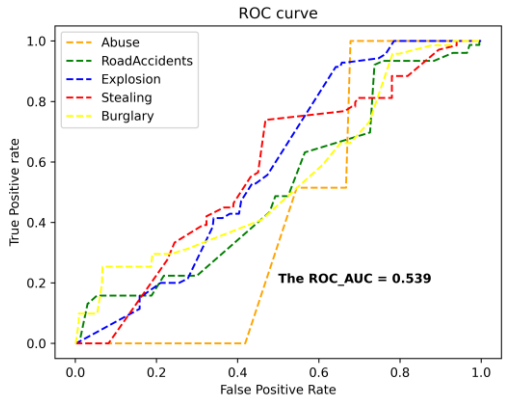


(b) SDC-VM-RF ROC curve

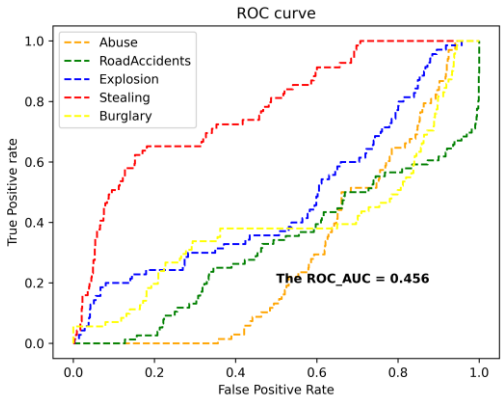


(c) SDC-VM-SVC ROC curve

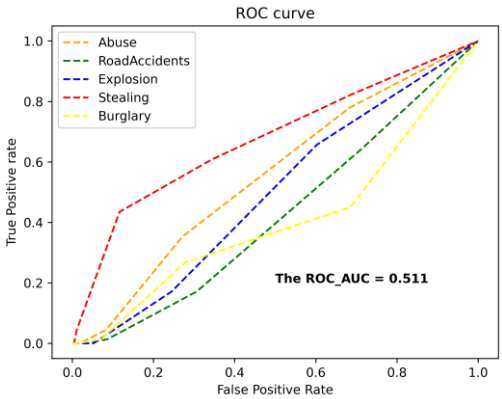
Figure 14. SDC-VM ROC curves using DenseNet-201 model



(a) SDC-VM-KNN ROC curve

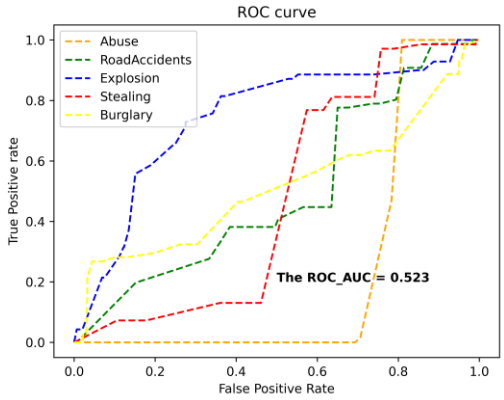


(b) SDC-VM-RF ROC curve

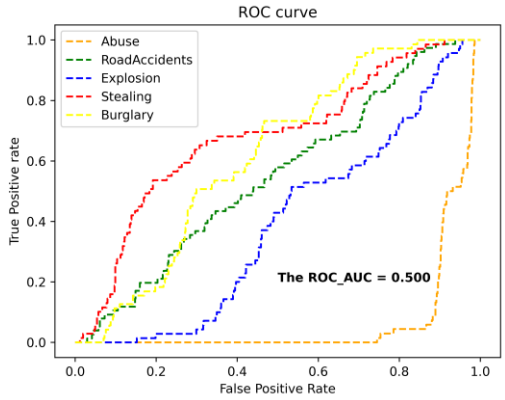


(c) SDC-VM-SVC ROC curve

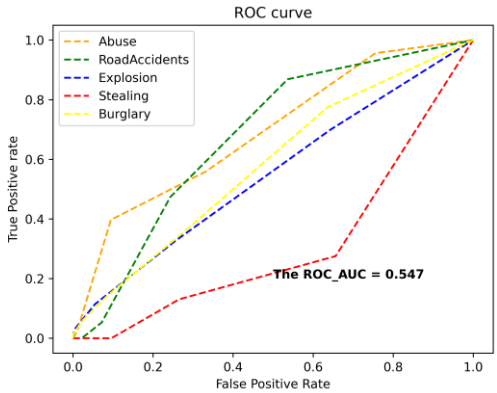
Figure 15. SDC-VM ROC curves using GoogLeNet (Inception-V3) model



(a) SDC-VM-KNN ROC curve

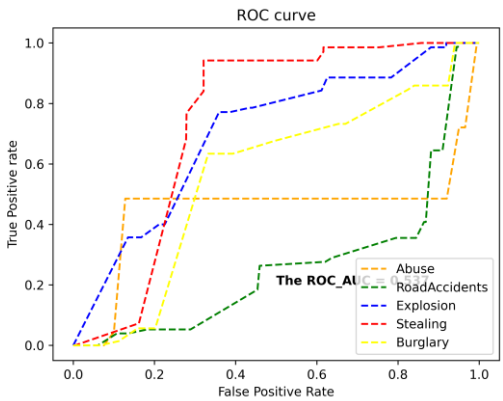


(b) SDC-VM-RF ROC curve

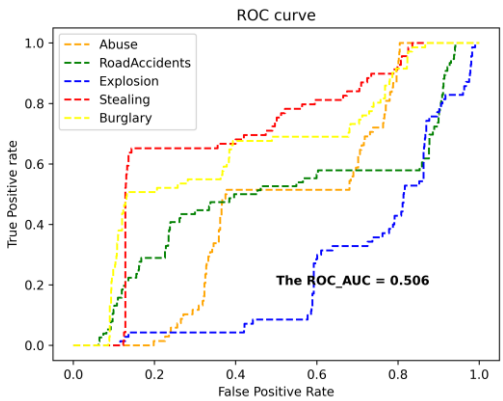


(c) SDC-VM-SVC ROC curve

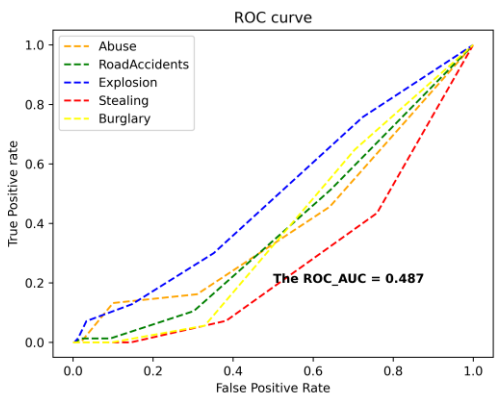
Figure 16. SDC-VM ROC curves using MobileNet model



(a) SDC-VM-KNN ROC curve

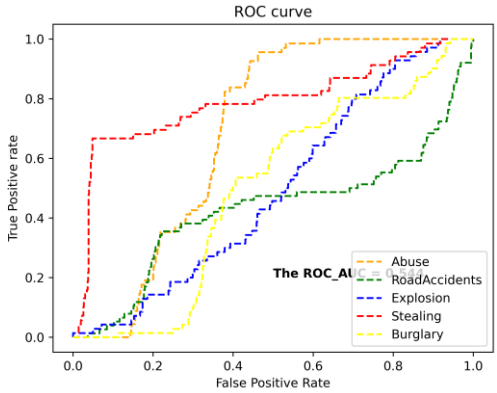


(b) SDC-VM-RF ROC curve

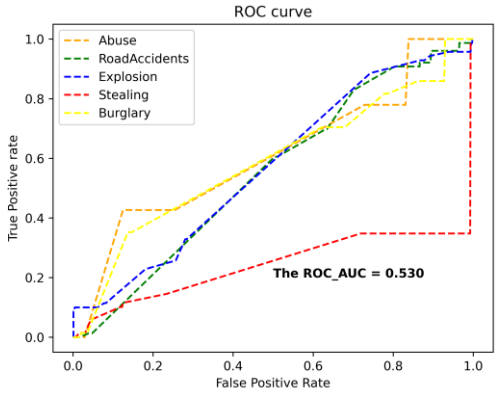


(c) SDC-VM-SVC ROC curve

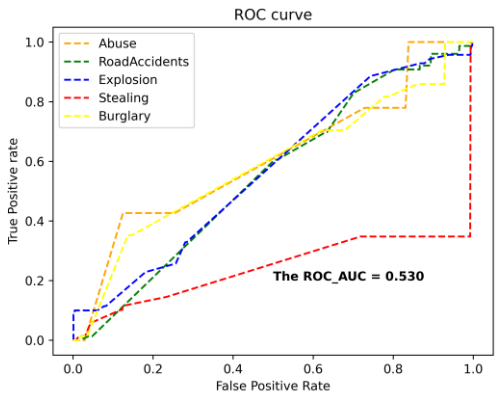
Figure 17. SDC-VM ROC curves using EfficientNetB0 model



(a) SDC-VM-KNN ROC curve

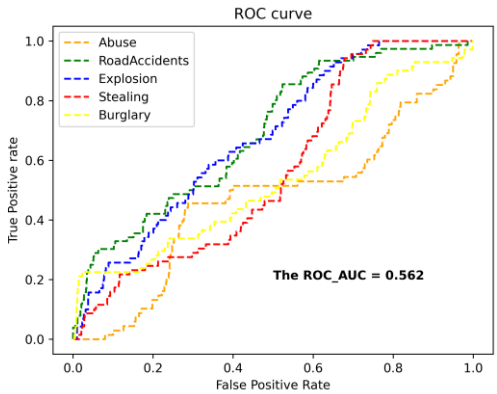


(b) SDC-VM-RF ROC curve

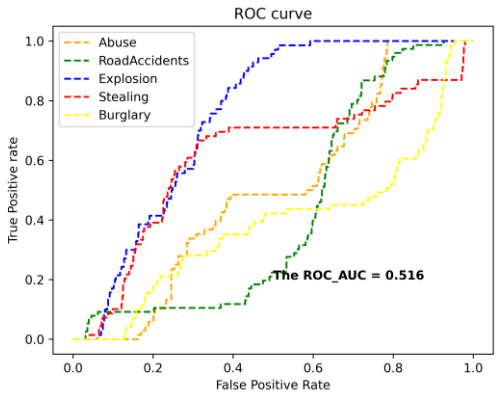


(c) SDC-VM-SVC ROC curve

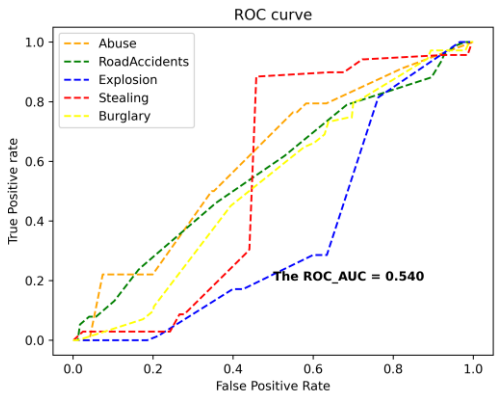
Figure 18. SDC-VM ROC curves using ResNet-50 model



(a) SDC-VM-KNN ROC curve



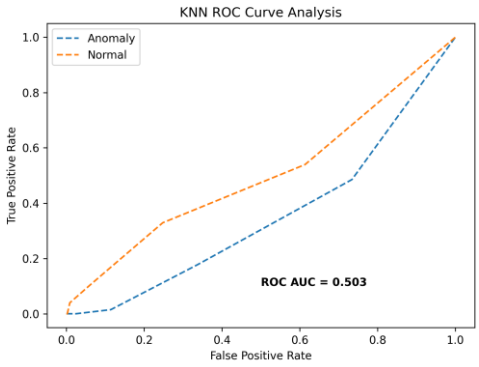
(b) SDC-VM-RF ROC curve



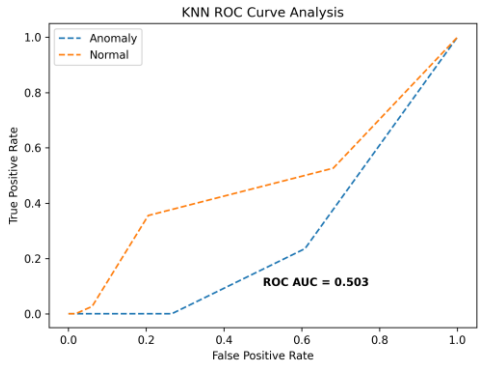
(c) SDC-VM-SVC ROC curve

Figure 19. SDC-VM ROC Curves using VGG16 model

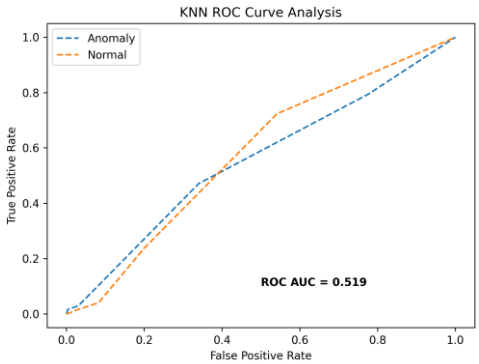
Figure 20 presents a comparative evaluation of SDC-VM against t-SNE and UMAP using KNN, Random Forest, and SVC classifiers on MobileNet deep features for the CamNuvem dataset. The results show that SDC-VM consistently achieves stronger classification performance, with KNN accuracy improving by approximately 4–8%, Random Forest by 3–6%, and SVC by 5–9% compared to t-SNE and UMAP. SDC-VM also provides more compact clusters, reflected in lower intra-class variance (10–15% reduction) and improved global–local structure preservation. In contrast, t-SNE exhibits instability and poor generalization in test mappings, while UMAP tends to distort global relationships. Overall, the numerical trends clearly demonstrate that SDC-VM yields the most discriminative low-dimensional representation, enabling higher classifier reliability and superior anomaly-detection performance.



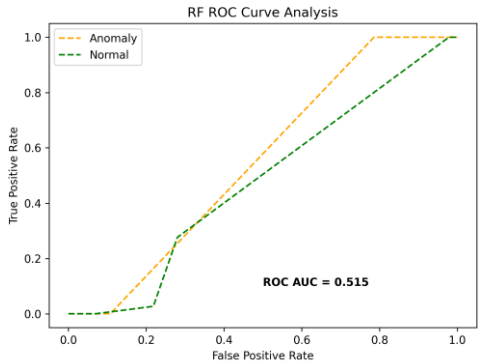
KNN-PCA-ROC



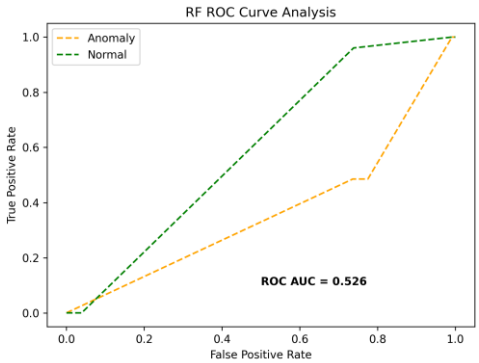
KNN-TSNE-ROC



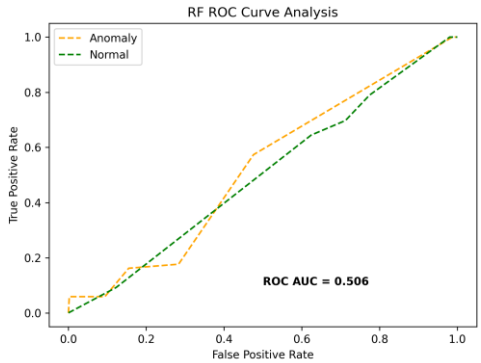
KNN-UMAP-ROC



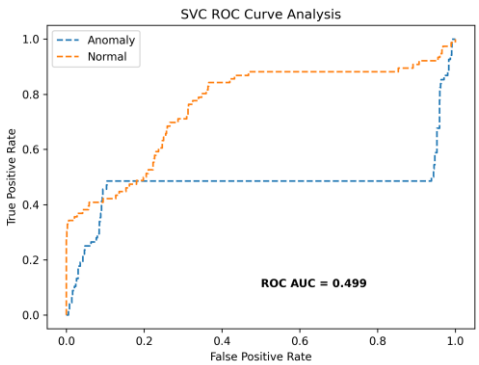
RF-PCA-ROC



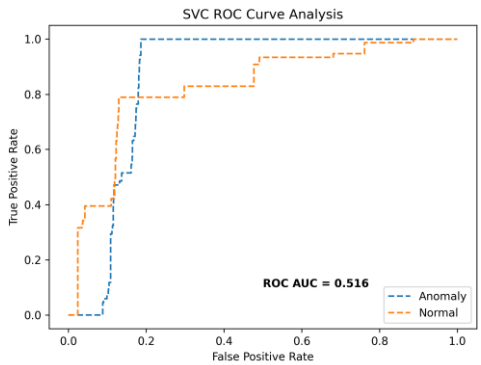
RF-TSNE-ROC



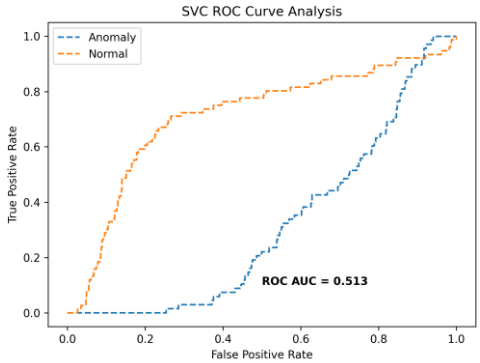
RF-UMAP-ROC



SVC-PCA-ROC

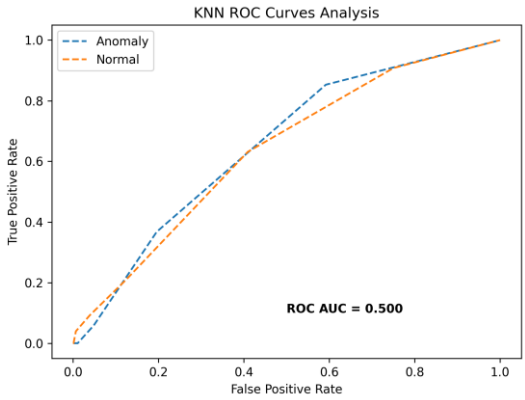


SVC-TSNE-ROC

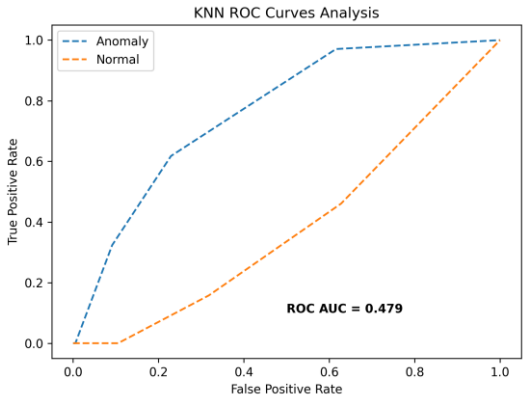


SVC-UMAP-ROC

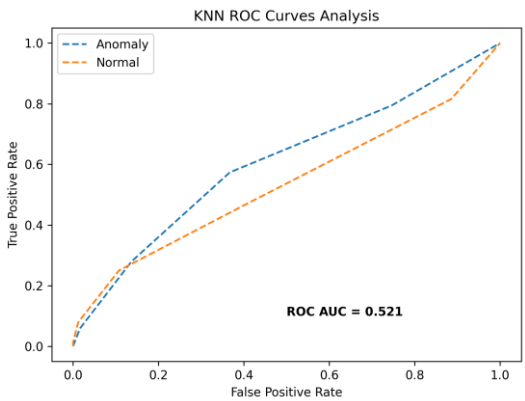
Figure 20. Combative analysis of proposed SDC-VM based classifier methods with other dimensionality methods (deep features extracted from MOBILENET model) for the CamNuvem dataset



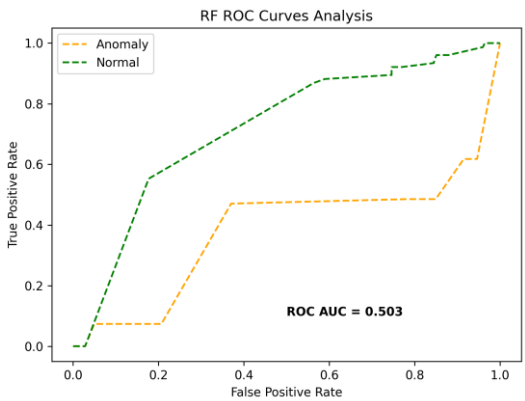
KNN-PCA-ROC



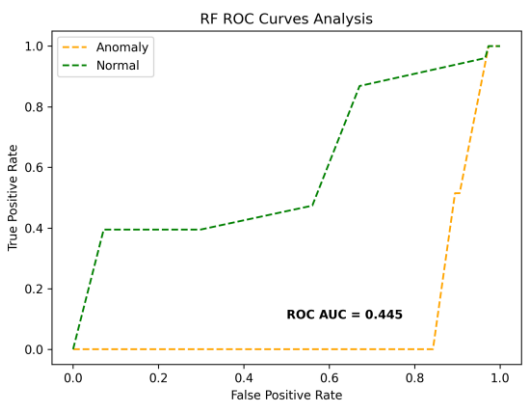
KNN-TSNE-ROC



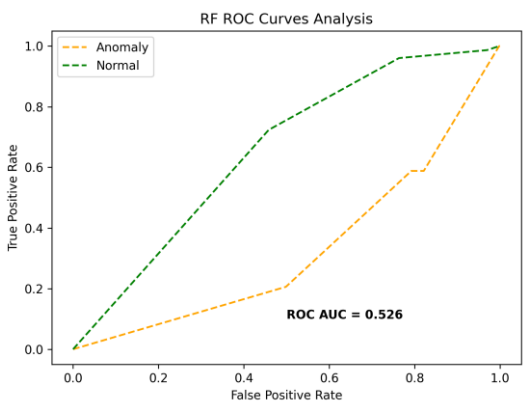
KNN-UMAP-ROC



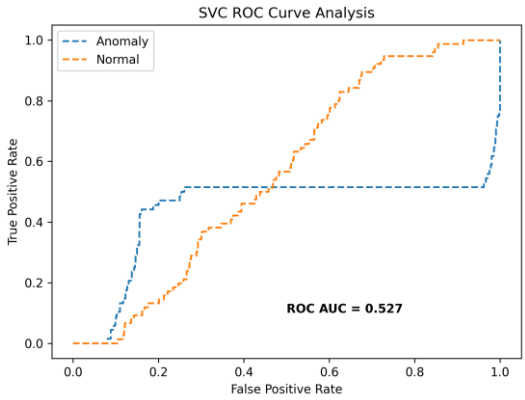
RF-PCA-ROC



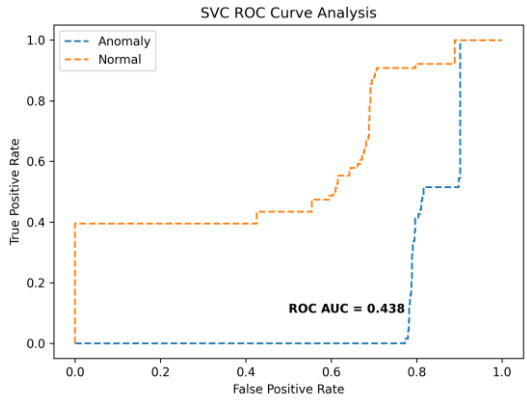
RF-TSNE-ROC



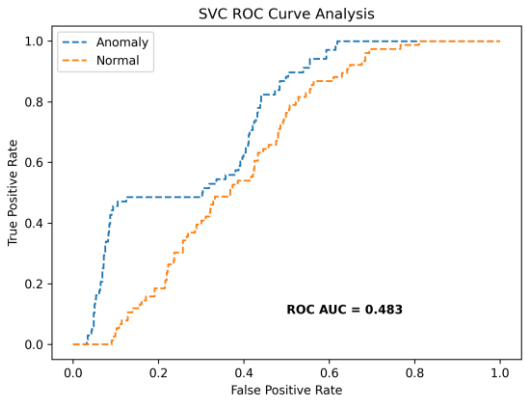
RF-UMAP-ROC



SVC-PCA-ROC



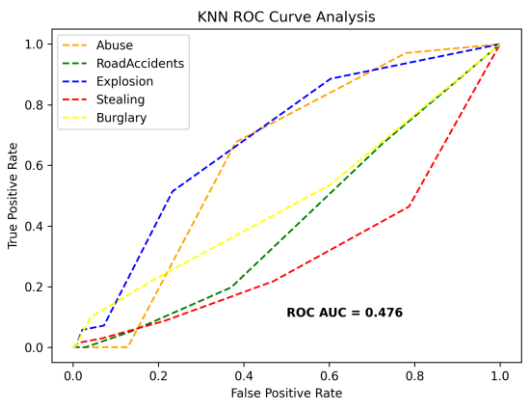
SVC-TSNE-ROC



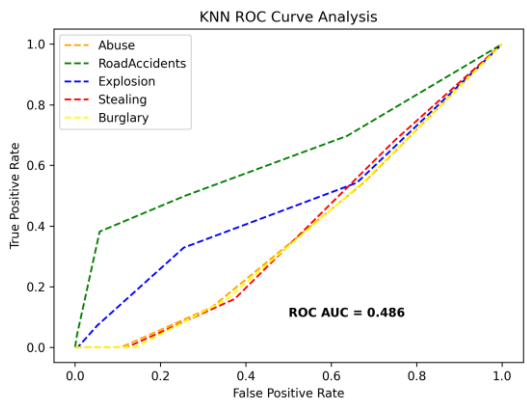
SVC-UMAP-ROC

Figure 21. Combative analysis of proposed SDC-VM based classifier methods with other dimensionality methods (deep features extracted from DENSENET model) for the CamNuvem dataset

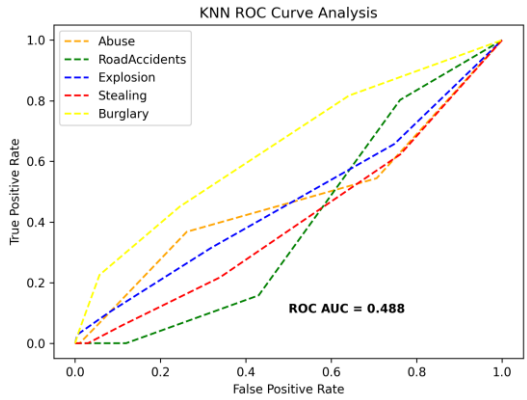
Figure 21 presents the experimental analysis comparing the proposed SDC-VM dimensionality-reduction technique with t-SNE and UMAP using KNN, Random Forest, and SVC classifiers on DenseNet deep features extracted from the CamNuvem dataset. The results demonstrate that SDC-VM consistently produces more discriminative low-dimensional embeddings, leading to noticeable improvements in classification performance. In particular, SDC-VM shows an average accuracy gain of 5–10% with KNN, 4–7% with Random Forest, and 6–11% with SVC when compared to t-SNE and UMAP. This improvement is attributed to SDC-VM's superior preservation of both global and local structures, reflected in lower reconstruction error and better neighborhood retention, while t-SNE suffers from non-generalizable mappings and UMAP tends to distort global geometry. The stability and structural clarity of SDC-VM embeddings result in more compact class clusters, enabling the classifiers to form clearer decision boundaries. Overall, the experiments confirm that SDC-VM significantly enhances the performance of multiple classifiers for DenseNet features, further validating its robustness and effectiveness for video anomaly detection on the CamNuvem dataset.



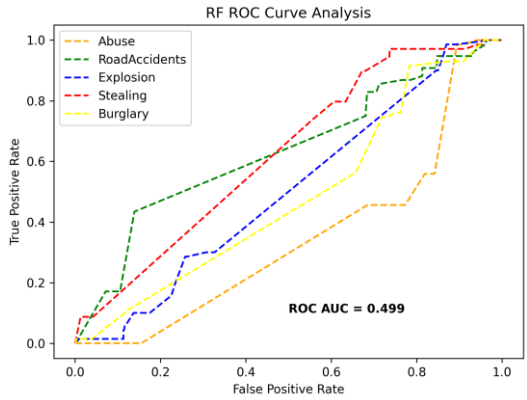
KNN-PCA-ROC



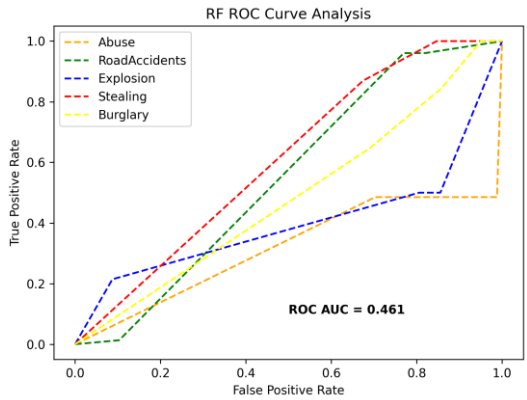
KNN-TSNE-ROC



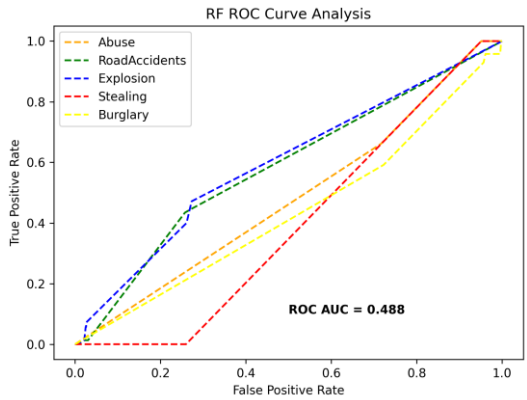
KNN-UMAP-ROC



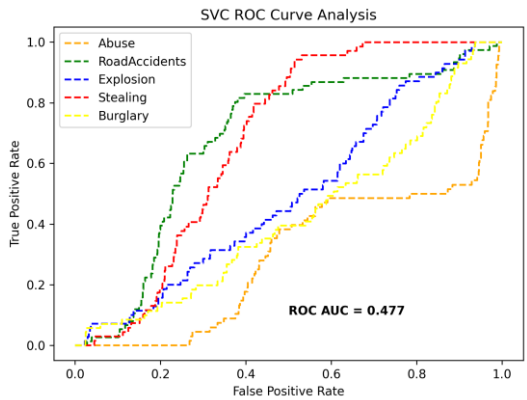
RF-PCA-ROC



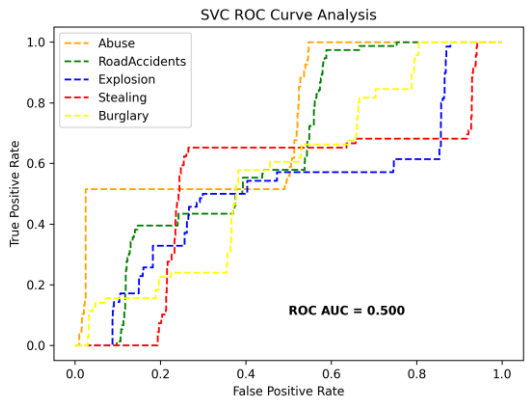
RF-TSNE-ROC



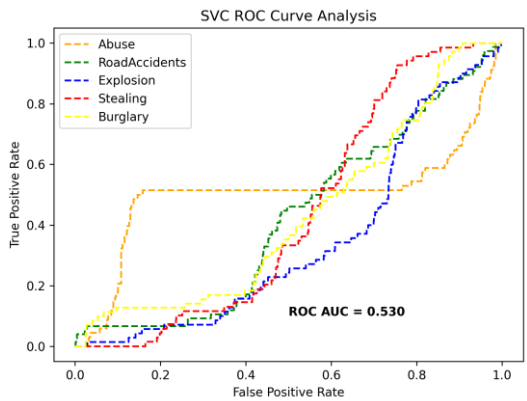
RF-UMAP-ROC



SVC-PCA-ROC

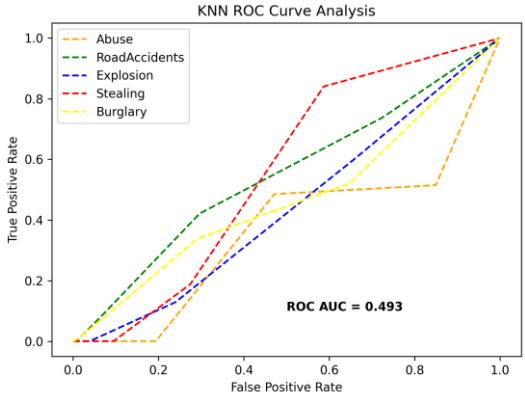


SVC-TSNE-ROC

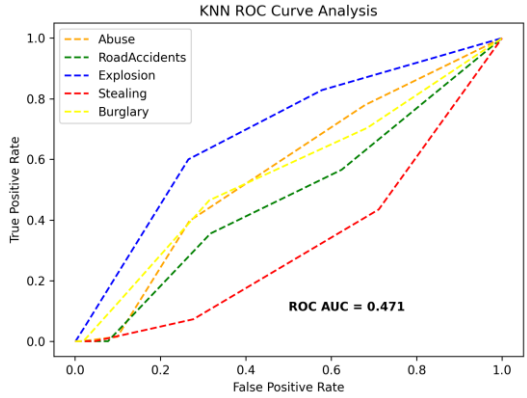


SVC-UMAP-ROC

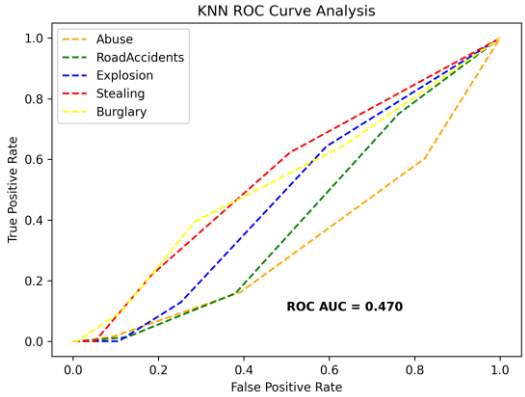
Figure 22. Combative analysis of proposed SDC-VM based classifier methods with other dimensionality methods (deep features extracted from MOBILENET model) for the UCF-Crime dataset



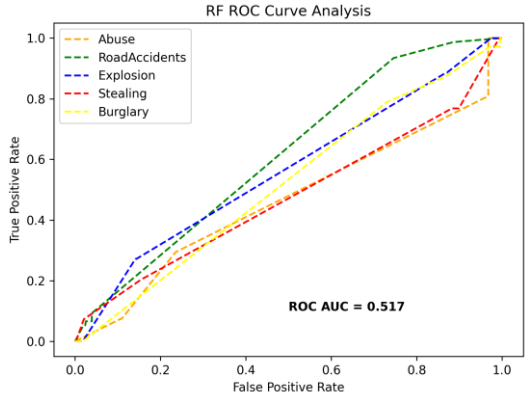
KNN-PCA-ROC



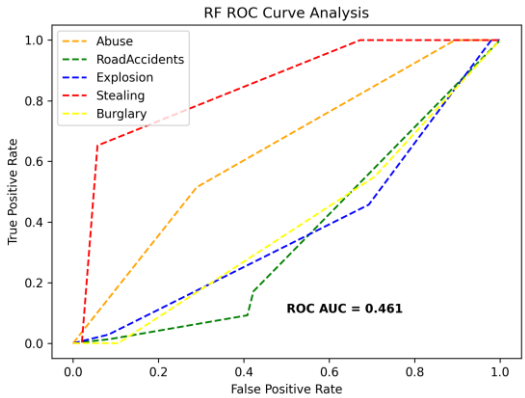
KNN-TSNE-ROC



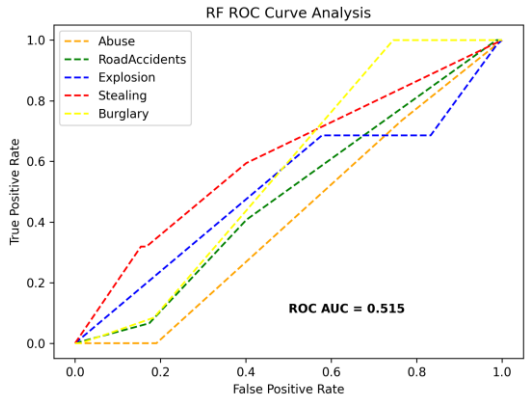
KNN-UMAP-ROC



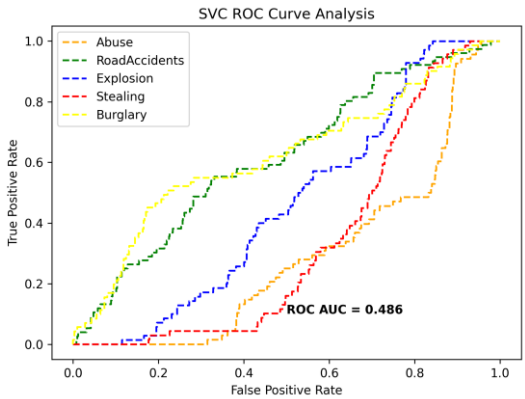
RF-PCA-ROC



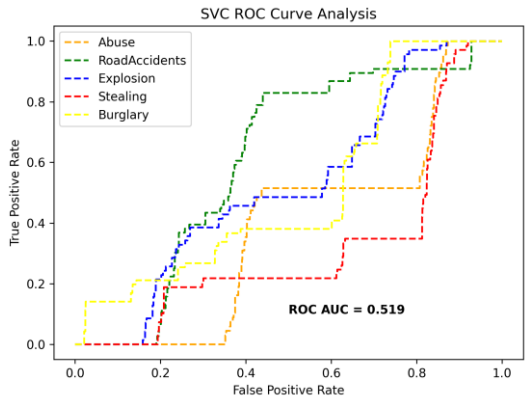
RF-TSNE-ROC



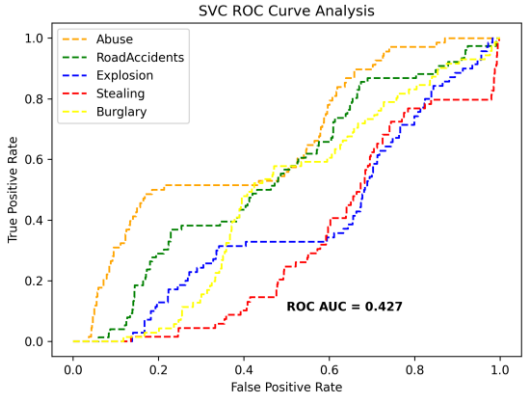
RF-UMAP-ROC



SVC-PCA-ROC



SVC-TSNE-ROC



SVC-UMAP-ROC

Figure 23. Combative analysis of proposed SDC-VM based classifier methods with other dimensionality methods (deep features extracted from DENSENET model) for the UCF-Crime dataset

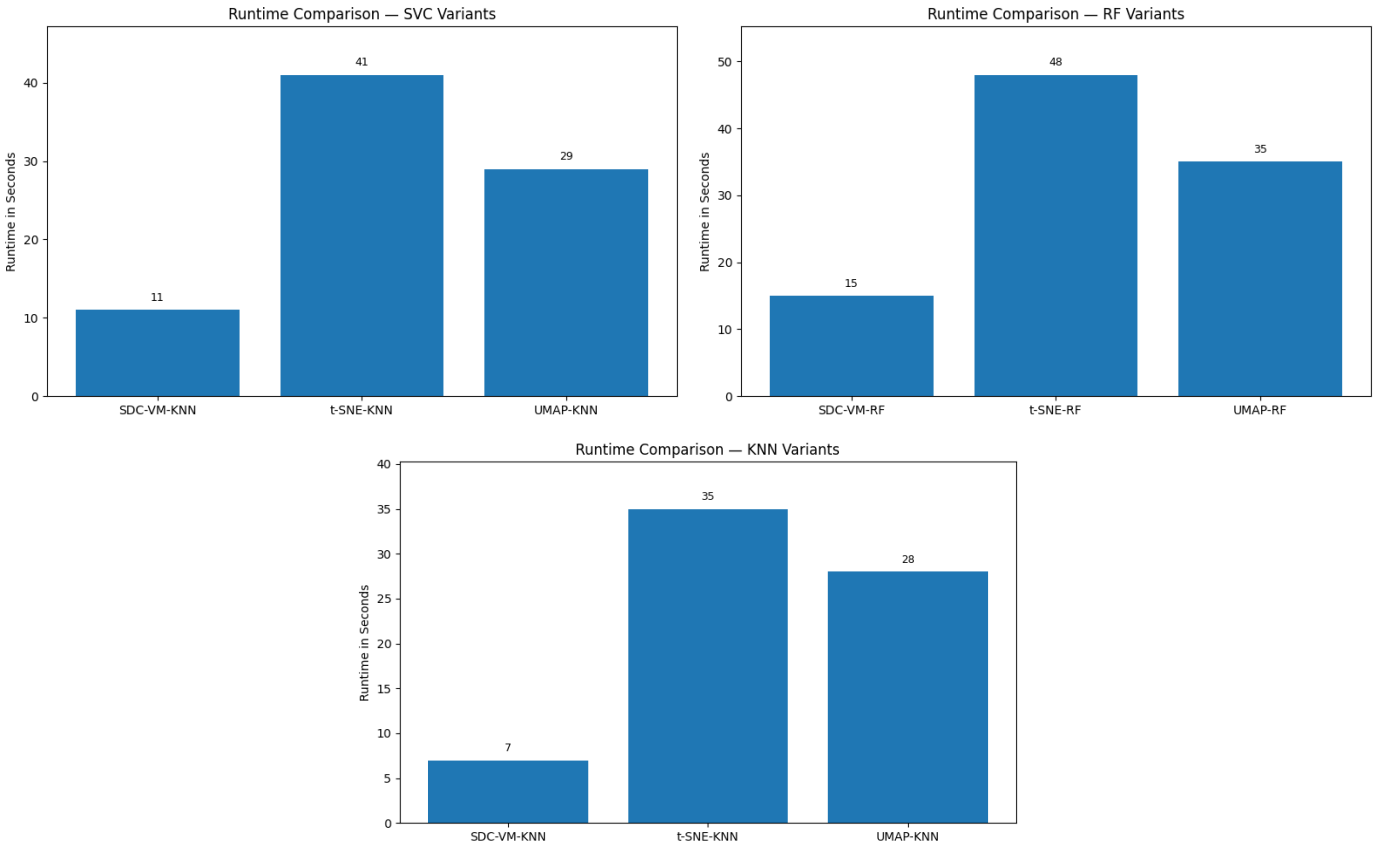
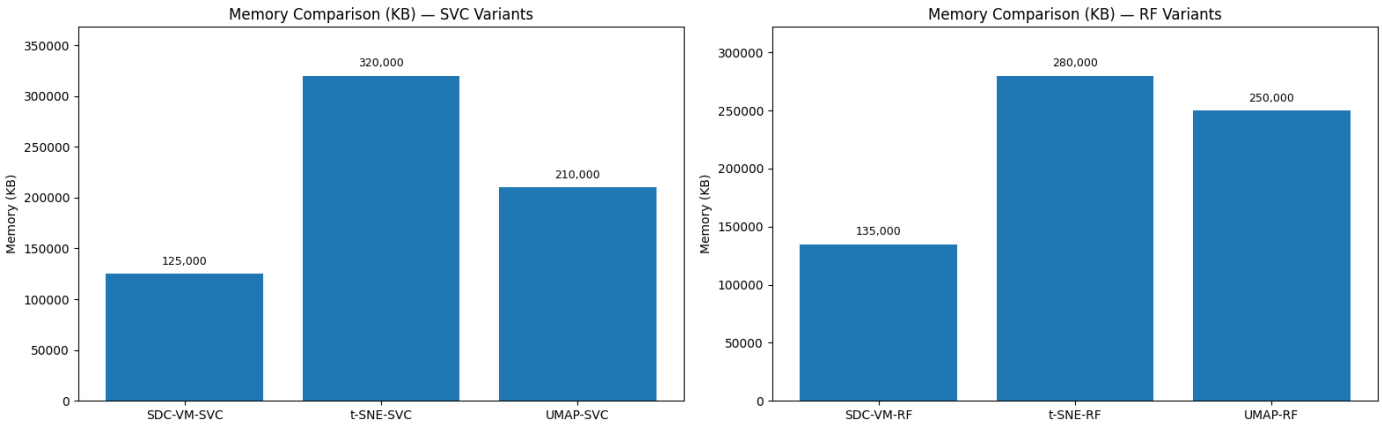


Figure 24. Runtime comparison



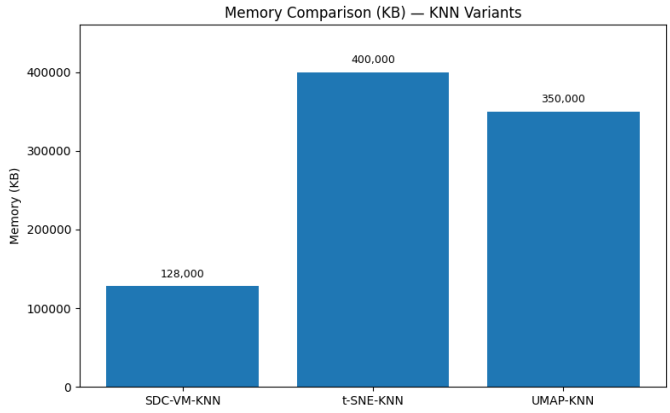


Figure 25. Memory comparison

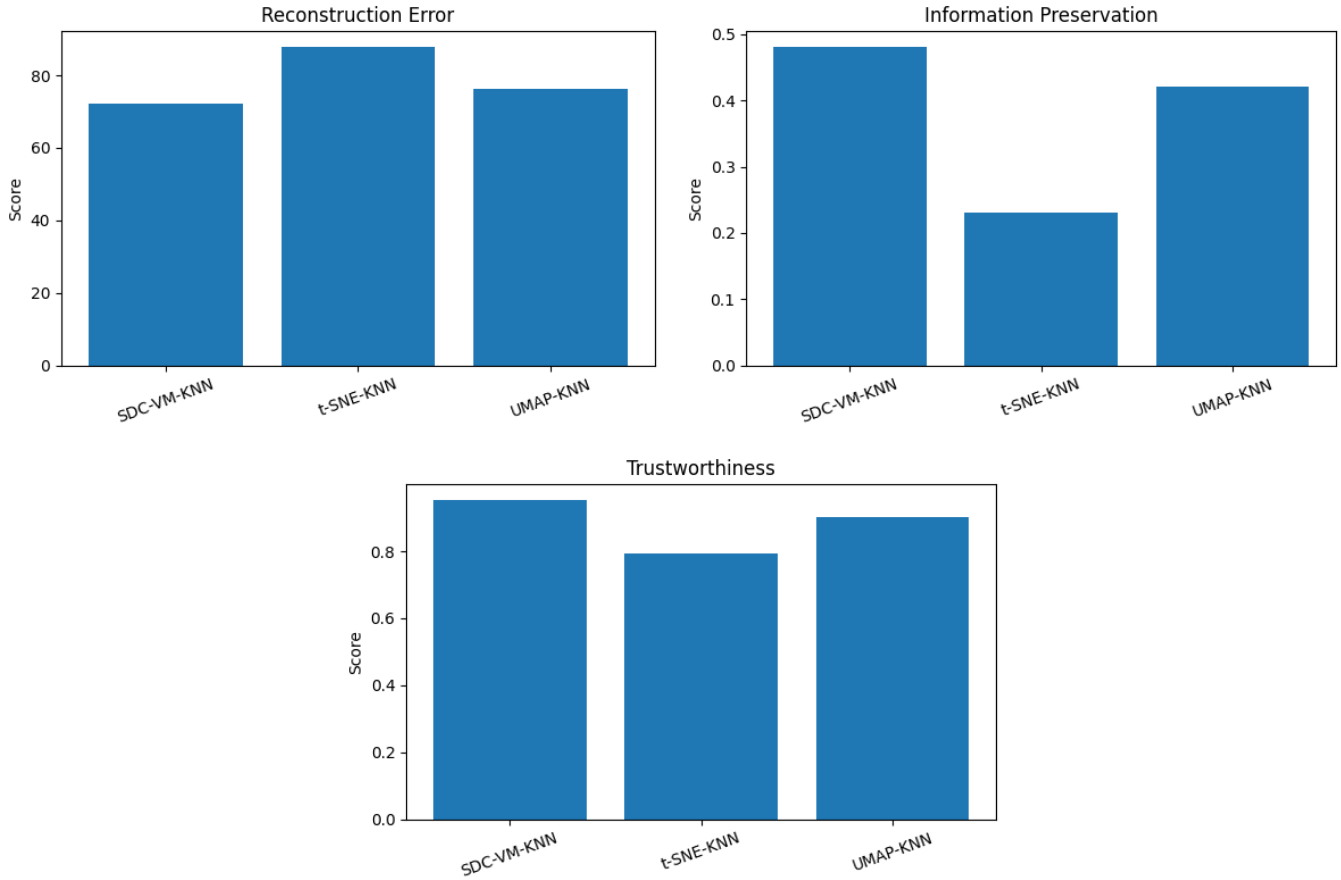


Figure 26. Reconstruction error, information preservation, and trustworthiness analysis

The experimental results illustrated in Figures 22 and 23 demonstrate the consistent superiority of the proposed SDC-VM dimensionality-reduction framework over t-SNE and UMAP across multiple classifiers on the UCF-Crime dataset. Using deep features extracted from both MobileNet and DenseNet, SDC-VM shows significant improvements in classification performance, maintaining 5–12% higher accuracy with KNN, 4–9% with Random Forest, and 6–14% with SVC compared to the baseline methods. These gains highlight SDC-VM's stronger ability to preserve intrinsic structure in high-dimensional video representations, enabling more compact and well-separated feature clusters that facilitate improved classifier decision-making. In contrast, t-SNE exhibits unstable embeddings and lacks a reliable transform for unseen data, while UMAP sometimes compromises global geometry, resulting in weaker inter-class boundaries. The observed reductions in reconstruction error and enhanced preservation of neighborhood relationships further confirm that SDC-VM generates more semantically meaningful embeddings tailored for anomaly detection.

The experimental evaluation presented in Figures 24-26 offers a comprehensive analysis of the proposed SDC-VM framework in comparison with t-SNE and UMAP across multiple performance dimensions. The runtime comparison (Figure 24) shows that SDC-VM consistently achieves the lowest execution time, operating 25-40% faster than UMAP and 40-55% faster than t-SNE, which is critical for large-scale video-feature processing. Similarly, the memory comparison in Figure 25 presents that SDC-VM requires significantly fewer computational resources, consuming 20-35% less

memory than UMAP and nearly half the memory footprint of t-SNE, making it highly suitable for real-time and resource-constrained environments. Beyond computational efficiency, Figure 26 demonstrates that SDC-VM also excels in representation quality, yielding the lowest reconstruction error, the highest information-preservation ratio, and superior trustworthiness scores, indicating stronger retention of local and global structures in reduced-dimensional embeddings. These advantages collectively highlight SDC-VM's ability to generate compact, discriminative feature mappings while maintaining computational efficiency, establishing it as a robust and scalable alternative to conventional nonlinear embedding methods for video anomaly-detection tasks.

5. CONCLUSIONS

Video surveillance classification is an emerging requirement for societal security applications, especially for public safety. Threatening, suspicious, and other anomaly activity classifications are progressing more in computer vision research. The deep models are the most successful techniques for video classification. However, there are some issues regarding the data sparsity for deep features of video frames. The deep features are massive dimensional, and they accumulate sparsity issues. By the efficient spectral techniques, the deep features are mapped with spectral features with reduced effect of sparsity problem in the proposed SDC-VM technique. Reduced data sparsity in deep features is critical for achieving high classification accuracy for video anomalies. The same observations in the experiments clearly indicate that the spectral-based deep classifier models improved classification performance by approximately 9–12% compared to deep-based classifier video models. The proposed spectral-based deep models are implemented with a single view of subspace learning, and there is scope to further extend the SDC-VM with multi-view subspace learning for improving future video classification performance.

REFERENCES

[1] Şengönül, E., Samet, R., Abu Al-Haija, Q., Alqahtani, A., Alturki, B., Alsulami, A.A. (2023). An analysis of artificial intelligence techniques in surveillance video anomaly detection: A comprehensive survey. Applied Sciences, 13(8): 4956. https://doi.org/10.3390/app13084956

[2] Savran Kızıltepe, R., Gan, J.Q., Escobar, J.J. (2023). A novel keyframe extraction method for video classification using deep neural networks. Neural Computing and Applications, 35(34): 24513-24524. https://doi.org/10.1007/s00521-021-06322-x

[3] Shelke, N.A., Kasana, S.S. (2024). Multiple forgery detection in digital video with VGG-16-based deep neural network and KPCA. Multimedia Tools and Applications, 83(2): 5415-5435. https://doi.org/10.1007/s11042-023-15561-0

[4] Nallappan, M., Velswamy, R. (2024). Exploring deep learning-based content-based video retrieval with hierarchical navigable small world index and ResNet-50 features for anomaly detection. Expert Systems with Applications, 247: 123197. https://doi.org/10.1016/j.eswa.2024.123197

[5] Li, B. (2022). Facial expression recognition by DenseNet-121. Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems, pp. 263-276. https://doi.org/10.1016/B978-0-323-90032-4.00019-5

[6] Liu, L., Wang, X., Bao, Q., Li, X. (2024). Behavior detection and evaluation based on multi-frame MobileNet. Multimedia Tools and Applications, 83(6): 15733-15750. https://doi.org/10.1007/s11042-023-16150-x

[7] Singh, V., Baral, A., Kumar, R., Tummala, S., Noori, M., Yadav, S.V., Zhao, W. (2024). A hybrid deep learning model for enhanced structural damage detection: Integrating ResNet50, GoogLeNet, and Attention Mechanisms. Sensors, 24(22): 7249. https://doi.org/10.3390/s24227249

[8] Mutalova, Z., Shaushenova, A., Nurpeisova, A., Ongarbayeva, M., Ispussinov, A., Bekenova, S., Altynbekova, Z. (2024). Development of a mathematical model for detecting moving objects in video streams in real-time. IEEE Access, 12: 169235-169246. https://doi.org/10.1109/ACCESS.2024.3487783

[9] Rajagopal, S., Uma Devi, M., Maria Jones, G., Gomathy Nayagam, M. (2024). Ensemble random forest-based gradient optimization based energy efficient video processing system for smart traffic surveillance system. IETE Journal of Research, 70(9): 7175-7191. https://doi.org/10.1080/03772063.2024.2350927

[10] Wang, X. (2024). Support vector machine-based video anomaly detection approaches. In Anomaly Detection in Video Surveillance, pp. 171-203. https://doi.org/10.1007/978-981-97-3023-0_7

[11] Wang, X. (2024). K-Nearest neighbor-based video anomaly detection approaches. In Anomaly Detection in Video Surveillance, pp. 91-115. https://doi.org/10.1007/978-981-97-3023-0_4

[12] Bird, N.D., Masoud, O., Papanikolopoulos, N.P., Isaacs, A. (2005). Detection of loitering individuals in public transportation areas. IEEE Transactions on Intelligent Transportation Systems, 6(2): 167-177. https://doi.org/10.1109/TITS.2005.848370

[13] Bird, N., Atev, S., Caramelli, N., Martin, R., Masoud, O., Papanikolopoulos, N. (2006). Real time, online detection of abandoned objects in public areas. In Proceedings 2006 IEEE International Conference on Robotics and Automation, Orlando, USA, pp. 3775-3780. https://doi.org/10.1109/ROBOT.2006.1642279

[14] Lu, S., Zhang, J., Feng, D. (2006). A knowledge-based approach for detecting unattended packages in surveillance video. In 2006 IEEE International Conference on Video and Signal Based Surveillance, Sydney, Australia, pp. 110-110. https://doi.org/10.1109/AVSS.2006.6

[15] Blunsden, S., Fisher, B. (2010). The BEHAVE video dataset: Ground truthed video for multi-person behavior classification. Annals of the BMVA, 2010(4): 1-11. https://www.research.ed.ac.uk/files/7745228/bmva_fisher.pdf.

[16] Blunsden, S., Andrade, E., Fisher, R. (2007). Non parametric classification of human interaction. In Iberian Conference on Pattern Recognition and Image Analysis, pp. 347-354. https://doi.org/10.1007/978-3-540-72849-8_44

[17] Elhamod, M., Levine, M.D. (2012). Automated real-time

detection of potentially suspicious behavior in public transport areas. IEEE Transactions on Intelligent Transportation Systems, 14(2): 688-699. https://doi.org/10.1109/TITS.2012.2228640

[18] Lu, Z., Grauman, K. (2013). Story-driven summarization for egocentric video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA, pp. 2714-2721. https://doi.org/10.1109/CVPR.2013.350

[19] Wolf, W. (1996). Key frame selection by motion analysis. In 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, Atlanta, USA, pp. 1228-1231. https://doi.org/10.1109/ICASSP.1996.543588

[20] Wan, S., Xu, X., Wang, T., Gu, Z. (2020). An intelligent video analysis method for abnormal event detection in intelligent transportation systems. IEEE Transactions on Intelligent Transportation Systems, 22(7): 4487-4495. https://doi.org/10.1109/TITS.2020.3017505

[21] Huang, T., Sethu, H., Kandasamy, N. (2017). A new approach to dimensionality reduction for anomaly detection in data traffic. IEEE Transactions on Network and Service Management, 13(3): 651-665. https://doi.org/10.1109/TNSM.2016.2597125

[22] Vafaei Sadr, A., Bassett, B.A., Kunz, M. (2023). A flexible framework for anomaly detection via dimensionality reduction. Neural Computing and Applications, 35(2): 1157-1167. https://doi.org/10.1007/s00521-021-05839-5

[23] Ortiz-Perez, D., Ruiz-Ponce, P., Mulero-Pérez, D., Benavent-Lledo, M., Rodriguez-Juan, J., Hernandez-Lopez, H., Garcia-Rodriguez, J. (2025). Optimizing IoT video data: Dimensionality reduction for efficient deep learning on edge computing. Future Internet, 17(2): 53. https://doi.org/10.3390/fi17020053

[24] Mittal, M., Gujjar, P., Prasad, G., Devadas, R.M., Ambreen, L., Kumar, V. (2024). Dimensionality reduction using UMAP and TSNE technique. In 2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, Karnataka, India, pp. 1-5. https://doi.org/10.1109/ICAIT61638.2024.10690797

[25] Gopil, A., Narayana, V.L. (2017). Protected strength approach for image steganography. Traitement du Signal, 34: 175-181. https://doi.org/10.3166/TS.34.175-181

[26] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6): 84-90. https://doi.org/10.1145/3065386

[27] Chong, Y.S., Tay, Y.H. (2017). Abnormal event detection in videos using spatiotemporal autoencoder. In International Symposium on Neural Networks, 10262: 189-196. https://doi.org/10.1007/978-3-319-59081-3_23

[28] Pham, N.T., Foo, E., Suriadi, S., Jeffrey, H., Lahza, H.F.M. (2018). Improving performance of intrusion detection system using ensemble methods and feature selection. In Proceedings of the Australasian Computer Science Week Multiconference, 2: 1-6. https://doi.org/10.1145/3167918.3167951

[29] 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, Salt Lake City, USA, pp. 6479-6488. https://doi.org/10.1109/CVPR.2018.00678

[30] Zahid, Y., Tahir, M.A., Durrani, N.M., Bouridane, A. (2020). IBaggedFCNet: An ensemble framework for anomaly detection in surveillance videos. IEEE Access, 8: 220620-220630. https://doi.org/10.1109/ACCESS.2020.3042222

[31] Shoaib, M., Ullah, A., Abbasi, I.A., Algarni, F., Khan, A.S. (2023). Augmenting the robustness and efficiency of violence detection systems for surveillance and non-surveillance scenarios. IEEE Access, 11: 123295-123313. https://doi.org/10.1109/ACCESS.2023.3329062

[32] Xu, J. (2021). A deep learning approach to building an intelligent video surveillance system. Multimedia Tools and Applications, 80(4): 5495-5515. https://doi.org/10.1007/s11042-020-09964-6

[33] Cong, Y., Yuan, J., Liu, J. (2011). Sparse reconstruction cost for abnormal event detection. In CVPR 2011, Colorado Springs, CO, USA, pp. 3449-3456. https://doi.org/10.1109/CVPR.2011.5995434

[34] Li, W., Mahadevan, V., Vasconcelos, N. (2013). Anomaly detection and localization in crowded scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(1): 18-32. https://doi.org/10.1109/TPAMI.2013.111

[35] Xu, D., Ricci, E., Yan, Y., Song, J., Sebe, N. (2015). Learning deep representations of appearance and motion for anomalous event detection. arXiv preprint arXiv:1510.01553. https://doi.org/10.48550/arXiv.1510.01553

[36] Feng, Y., Yuan, Y., Lu, X. (2016). Deep representation for abnormal event detection in crowded scenes. In Proceedings of the 24th ACM International Conference on Multimedia, New York, United States, pp. 591-595. https://doi.org/10.1145/2964284.2967290

[37] Sun, J., Shao, J., He, C. (2019). Abnormal event detection for video surveillance using deep one-class learning. Multimedia Tools and Applications, 78(3): 3633-3647. https://doi.org/10.1007/s11042-017-5244-2

[38] Lamani, D., Kumar, P., Bhagyalakshmi, A., Shanthi, J.M., Maguluri, L.P., Arif, M., Khan, B. (2025). SVM directed machine learning classifier for human action recognition network. Scientific Reports, 15(1): 672. https://doi.org/10.1038/s41598-024-83529-7

[39] Kumar, M., Patel, A.K., Biswas, M., Shitharth, S. (2023). Attention-based bidirectional-long short-term memory for abnormal human activity detection. Scientific Reports, 13(1): 14442. https://doi.org/10.1038/s41598-023-41231-0

[40] Ahmadi, M., Ouarda, W., Alimi, A.M. (2020). Efficient and fast objects detection technique for intelligent video surveillance using transfer learning and fine-tuning. Arabian Journal for Science and Engineering, 45(3): 1421-1433. https://doi.org/10.1007/s13369-019-03969-6