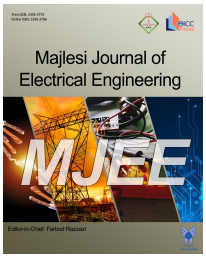


Majlesi J. Electr. Eng. 2026; 20(1) : 73-95

Majlesi Journal of Electrical Engineering (MJEE)

https://dx.doi.org/10.57647/mjee.2026.2001.06

Original Research

A Comprehensive Approach for Gun, Mask and Suspicious Activity Detection

Sunpreet Kaur Nanda[1], Deepika Ghai[2,*], Prashant Ingole[3], Sandeep Kumar[4], Suman Lata Tripathi[5]

[1]*Electronics and Telecommunication Engineering Department, P.R.Pote College of Engineering and Management, Amravati-444602, India*
[2]*School of Electronics and Electrical Engineering, Lovely Professional University, Phagwara-144411, India*
[3]*Department of Information Technology, Prof. Ram Meghe Institute of Technology and Research, Amravati-444602, India*
[4]*Department of CSE, Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune-440008, India*
[5]*Department of Electronics and Communication Engineering, Symbiosis Institute of Technology (SIT), Symbiosis International (Deemed University), Pune-412115, India*

[*]**Corresponding author:** money.ghai25@gmail.com

Article History

Received:
14 June 2024

Revised:
19 November 2024

Accepted:
17 February 2025

Published in Issue:
31 March 2026

Abstract:

Video forensics includes understanding how to examine and identify crime in the footage. The process is constantly evolving with new technology and innovations. The world of video forensics experts is growing. The realm of video forensics is simply a growing community of specialists linked with the digital video forensics sector. The State-of-the-art examinations and crimes consistently cross global and language fringes these days. With rapidly advancing technology, video has emerged as the foremost and indispensable tool in the fight against those who break the law, capturing them in the act. The computerized video forensics industry is rapidly expanding. Rapid technological advancements have made video a powerful and essential tool for law enforcement. Technology changes quickly, and innovators work with computerized images to catch criminals. Video forensics examination helps to know how accurate the input video is. In this paper, the proposed method used a soft computing technique, i.e. YOLOv3, to detect suspicious persons, the guns, or the masks by extracting frames and features from a video. In the dataset, it compares the extracted edge with images and generates output with bounding boxes for suspicious persons, the guns, or the masks. The realm of video forensics and its outcomes are also examined by this paper. When tested on different datasets, the proposed method outperforms existing techniques. For both the models, such as YOLO and customized Convolutional Neural Network (CNN), execution measurements were taken and are shown to supersede the customized CNN with its identification of guns and masks. The accuracy for YOLO design is 100% for both guns and mask detection, respectively, whereas accuracy for the customized CNN with guns and mask detection is 61.54% and 61.55%, respectively. Experimental results show that the proposed methodology outperforms the other existing methods.

Keywords: Video forensics; You Only Look Once; Customized Convolutional Neural Network; Soft computing techniques; Gun detection; Mask detection; Suspicious persons

Cite this article: Nanda SK, Ghai D, Ingole P, Kumar S, Tripathi SL. A Comprehensive Approach for Gun, Mask and Suspicious Activity Detection. Majlesi J. Electr. Eng. 2026;20(1): 73-95. https://doi.org/10.57647/mjee.2026.2001.06

1. Introduction

In investigating a crime, a process includes gathering and examining the information from past cases. Forensic science is used in many fields like cyber forensics [1, 2], video forensics [3, 4], digital forensics [5, 6], etc. One example of video forensics is the examination of videos of crime-related activities [7]. Video evidence captures

the criminals today using masks and guns efficiently and accurately. Video forensics detects hidden video objects using scientific methods, such as computer algorithms. Today's hottest topics include security at crowded and lonely locations. With soft computing techniques, one can analyze video footage without making mistakes that usually arise from manual analysis. Computer-based criminology aims to bring the use of computers and digital methods to assessment disciplines of criminology. Computers can be used in encryption, making entries or entire models secret, examining entries utilizing different computer systems, or changing any collected data. Identifying a suspect in a video of a crime scene can be challenging. Recognizing someone's face or seeing the weapon used for a crime is difficult because of the low-quality video. To address this problem, researchers are developing new methods to improve the quality of videos and photos. The open data can create connections between ancient and current topics. Internet of Things (IoT) devices allow social research to be conducted on a larger scale, incorporating the lessons learned from the past with the latest technological developments [8, 9, 10]. Video forensic investigation systems are developed to detect critical points such as weapons, faces, and arson. It is used in criminal cases and involved in the documentation of the scene, collecting physical evidence for the prosecution, and tracking down suspects. Forensic science is the utilization of different skills to answer legal issues. Other specific regions are making strides in illicit tasks. Criminological science is a mixture of criminology and computing. It benefits from research using Personal Computers (PCs), electronic methods, and theories on a particular problem [11]. Digital forensics, by the by, needs criminology and computing for mutually beneficial purposes. Video-based identification has several troubles. These difficulties include the video's lousy quality, the subject's size, and the angles. To solve this problem, data needs to be collected to make advancements in solving these problems. Authorities leverage open-source assets such as CCTV footage, online image repositories, or historical pathways to track multiple leads that progress through registration procedures [12]. They intend to develop techniques utilizing innovation to gather information more efficiently and effectively than previous methods. Video forensics is increasingly used to prove someone's identity or if the data input is valid. It is becoming necessary on a global level, and it will continue to be needed. This technology helps prevent problems before they happen and catch those already committing crimes before they are caught. The Indian Crime Investigation Association has also taken initiatives to create better video legal sciences for its use [13]. Video documentation is essential for all case parts, including documenting evidence and catching every detail that might not be clear to the average eye. As video shops replace other witness testimony, these devices are becoming the main witnesses of events in their customers' lives due to their ability to record nearly everything. Many concerns correlate with the use of this device, so these must receive more attention and exploration. Due to the development in data technology, electronic devices are a vital part of everyday life. These devices are used to record not only financial transactions but also for recording every event. These machines are investigated for security reasons in metropolitan areas.

With YouTube, computerized recordings of day-to-day life have become accessible. With the accessibility of video-altering programming, it has been made easier for these recordings to be created [14]. However, because it is still being determined how often video clips are being edited or changed to make them seem as if they were real, in reality, they are not. There must be mechanisms set up that will screen any altered videos. Video editing technology is the most well-known use for video content alteration. Using this technology, specialists can effortlessly extract an object from a video sequence, integrate an item from another video source, or incorporate a computer-generated object through design software. Video editing technology helps broaden our visual experience through its use in specific fields. As computerized video altering becomes more accessible in the future, public trust in videos is eroding with the word of technology. Although a few instances of fraud have been uncovered, the scientific community needs to think about how to ensure a video is correct. One option to prove the validity of videos is advanced watermarking [15, 16]. There are a couple of sorts of watermarks, yet fragile and semi-sensitive watermarks can be utilized to ensure the recording. Fragile watermarking works by inserting unpretentious information, which will be changed in the event that there is a work to adjust the video. After that, the install data can be separated to verify the video's accuracy. Other than barring the watermark, the change actually means higher tension in the video. This doesn't affect the quality of the video but might discourage viewers. A watermark should be embedded around the recording hour to avoid losing quality. Watermarking techniques could be more effective against modern video alteration techniques. However, it aspires to generate software to differentiate between real and altered videos. The three modules of our algorithm consist of aligning, matching, and transforming. To start recognizing false videos, one must identify specific modules in advance. These modules can help to detect falsification even in new recordings without watermarks.

Deep CNNs for video forensics can be examined using machine learning. With the insights from this discovery, the detection will become more precise and have a higher capacity to deny alerts. The theory of Inception and Residual, which uses deep learning for auditing writing, can be evaluated [17, 18]. Both are efficient and excellent models. The exhibition combines two pieces of human behavior through the two-dimensional design to create a unique experience. The exhibit shows ordinary and strange human behavior in public spaces. The following is a list of this paper's objectives:

i) Using YOLOv3, to create a training and testing framework for video-based suspicious activity de-

https://dx.doi.org/10.57647/mjee.2026.2001.06

tection.

ii) Design and foster a cover and firearm distinguishing proof framework from dubious action recordings.

iii) To approve the presentation of the proposed framework regarding exactness, accuracy, and review, with other conditions of workmanship techniques on different benchmark datasets.

In recent years, video forensics has emerged as a powerful tool in criminal investigations, leveraging video footage to detect crime-related activities and identify suspects. However, despite significant advancements, there remain critical challenges in effectively analyzing video evidence, especially when the quality of the footage is poor or when the crime occurs in crowded or isolated environments. Video recordings often suffer from low resolution, improper angles, or obstructions, making it difficult to identify individuals, weapons, or other critical objects accurately. Additionally, the ease with which videos can be altered using modern video editing technologies raises concerns about the authenticity of video evidence. While traditional forensic techniques involve manual analysis, which is time-consuming and error-prone, developing more automated, computer-based methods can greatly improve the accuracy and efficiency of video forensic investigations. Existing methods for video-based identification face significant limitations in terms of video quality, video distortion, and the complexity of recognizing objects or individuals from poorly captured footage [19, 20, 21, 22, 23, 24, 25, 26, 27, 28]. This gap in current forensic methods highlights the need for more sophisticated approaches to effectively handle video quality issues and detect tampered footage, ensuring the reliability of video evidence in legal contexts. The research gap exists in developing advanced techniques for video forensic analysis that are listed as follows:

1. Improve the quality of low-resolution footage to enhance object and face recognition.
2. Detect and verify the authenticity of videos in light of increasingly sophisticated video manipulation techniques.
3. Create more efficient and accurate systems for identifying critical objects (e.g., weapons, faces) and suspicious activities in videos, particularly in challenging environments.

This research aims to address these gaps by exploring the application of deep learning techniques, such as Convolutional Neural Networks (CNNs), for video forensics, alongside advanced methods like watermarking and object detection frameworks (e.g., YOLOv3), to improve the quality, authenticity, and accuracy of video-based evidence analysis. The goal is to develop a more reliable and automated forensic system capable of detecting suspicious activities and verifying the validity of video footage in criminal investigations. This work will make forensic video analysis more robust, efficient, and applicable to real-world law enforcement scenarios.

This research addresses the key challenges and gaps identified in video forensics by introducing novel solutions and methodologies that enhance video-based investigations' accuracy, reliability, and efficiency. Specifically, the contributions of this work are:

- Development of an Improved Video Forensics Framework: This proposal proposes a comprehensive framework for detecting and analyzing suspicious activities in video footage using advanced machine learning techniques, particularly You Only Look Once (YOLOv3) for real-time object detection.
- Weapon and Face Detection System: Designs and implements a system for the identification of weapons and faces from suspicious activity recordings, leveraging state-of-the-art object detection algorithms to identify critical evidence from low-quality videos.
- Quality Enhancement of Low-Resolution Videos: Introduces methods to enhance the quality of low-resolution or distorted video footage, allowing for more accurate recognition of faces, objects, and other relevant details, thus improving the overall effectiveness of forensic analysis.
- Video Authenticity Verification: Explores the use of advanced watermarking techniques (e.g., fragile watermarking) for ensuring the authenticity and integrity of video evidence, allowing for the detection of alterations, and providing a mechanism to verify the credibility of video recordings.
- Benchmarking and Evaluation: Conducts an extensive evaluation of the proposed system on multiple benchmark datasets to validate its performance in terms of accuracy, precision, recall, and F1 score, comparing it with current state-of-the-art methods and demonstrating its superior effectiveness in handling real-world forensic scenarios.
- Integration of Deep Learning for Robust Detection: Implements and tests Deep Convolutional Neural Networks (CNNs), specifically leveraging Inception and Residual Networks for high-level video feature extraction and suspicious activity detection, improving the robustness and scalability of forensic systems.

This research provides significant advancements in video forensics by addressing critical challenges in suspicious activity detection, weapon and face identification, video quality enhancement, and authenticity verification. The proposed framework, leveraging YOLOv3, deep learning models, and advanced techniques like fragile watermarking, demonstrates a comprehensive approach to improving the accuracy, reliability, and efficiency of forensic investigations. Through extensive benchmarking, the system has shown superior performance over existing methods, making it a valuable tool for real-world security

https://dx.doi.org/10.57647/mjee.2026.2001.06

and forensic applications. The integration of cutting-edge deep learning technologies ensures scalability and robustness, positioning this work at the forefront of modern video forensics. These contributions aim to bridge the existing gaps in video forensics, providing a more reliable, automated, and scalable solution for criminal investigations.

The remaining parts of the paper are laid out as follows: section 2 examines the connected business related to the video legal sciences setting. The proposed technique for video crime locations is examined in section 3. The outcomes and conversations i.e. the results and discussions of the video criminology examination are talked about in section 4. The conclusions are mentioned in section 5.

2. Related work

This section reviews relevant literature in the field of video forensics, categorizing the studies into thematic groups based on their focus areas. These groups highlight the gaps in the current research and justify the need for the proposed approach to improve the identification of critical objects (such as masks, guns, and suspicious behaviors) in video footage.

2.1 Object detection and activity recognition in videos

A significant body of research has focused on detecting and recognizing various objects and activities in video footage using deep learning techniques. Early works focused on basic activity recognition, such as walking, running, or clapping. These foundational methods have been important for understanding human movement patterns but are limited when it comes to identifying threatening behaviors like fighting or the presence of dangerous objects.

Wan et al. [19] (2017) discussed advances in video forensics for security purposes, particularly focusing on video manipulation detection. Their work on automatic jump-cut identification laid a foundation for ensuring video authenticity, an important aspect of forensic video analysis. However, their work does not address the identification of specific objects, such as weapons or faces, which is crucial in criminal investigations.

Kaur et al. [20] (2017) focused on the forensic analysis of mobile phones, providing tools for analyzing data such as SMS logs, internet searches, and account histories. While this study was valuable for digital forensics involving mobile devices, it does not explore video footage analysis or object detection, leaving a gap for improving forensic investigations using video surveillance.

Kaushala et al. [21] (2018) investigated moving object identification in videos using various machine-learning techniques. Their findings, especially the success of neural networks, point toward the potential of using deep learning for object detection. However, the study did not address the challenge of detecting specific threats like weapons, nor did it focus on improving the quality of low-resolution video footage, which is often encountered in real-world scenarios.

These studies indicate that while object detection and activity recognition have been explored, threat detection in specific crime-related activities remains underexplored. Identifying behaviors indicative of violent crime, such as fighting or gun usage, is crucial in forensic video analysis. This gap forms the basis for our proposal to improve object recognition in videos, specifically focusing on masks, guns, and suspicious movements.

2.2 Fire and anomaly detection in videos

Another branch of video forensics has focused on detecting specific events, such as fire or other anomalous activities in video footage. While these methods are useful in certain contexts (such as fire safety or surveillance), they are not directly applicable to criminal activity detection.

Muhammad et al. [22] (2018) worked on detecting fires in videos using CNNs, achieving 93.55% accuracy on their custom dataset. This study highlights the effectiveness of CNNs in detecting specific events but is limited by its focus on fire detection, rather than the identification of criminal activities like weapon use or suspicious behavior.

Bajestani et al. [23] (2018) focused on R-CNN algorithms to detect and recognize objects in video scenes. Their method achieved an accuracy of true positives = 0.56 and false positives = 0.36 on several standard datasets, which is promising. However, their study was primarily focused on general object detection, not on detecting dangerous objects like guns or identifying suspicious human activities associated with criminal behavior.

These studies contribute to object recognition in video but do not directly address criminal threat detection or suspicious activity identification. Our proposal addresses this gap by focusing on detecting specific threatening behaviors like fighting, the use of weapons, and mask-wearing in videos, particularly in low-quality footage.

2.3 Security solutions for video forensics and IoT networks

Research in security solutions for Internet of Things (IoT) networks has also impacted video forensics, particularly in terms of improving the overall efficiency of security systems that rely on video surveillance. These works focus on optimizing video analysis and addressing video authenticity but do not directly handle object or activity recognition.

Conti et al. [24] (2018) proposed a security solution for IoT networks using Software-Defined Networking (SDN-IoT), which enhances efficiency in managing security for IoT-based systems. This research is important for improving the overall infrastructure for video forensics but does not directly contribute to the analysis or detection of objects or suspicious activity in video footage.

Chen et al. [25] (2018) developed a CNN-based technique for image manipulation detection, showing improvements in identifying altered images with higher accuracy compared to other methods. While this is important for verifying the integrity of video evidence,

it does not address the specific challenges related to detecting objects (such as guns or masks) in real-world video footage or identifying threatening behaviors.

These studies highlight the importance of security infrastructure for video forensics but do not provide solutions for object detection or behavior analysis. Our approach combines the strengths of deep learning algorithms for detection with advanced forensic tools for video authenticity, making it more suitable for real-world criminal investigations.

2.4 Face mask detection and COVID-19-related research

During the COVID-19 pandemic, several studies focused on face mask detection, a problem relevant to both health and security applications [29]. However, these studies primarily focus on identifying whether people are wearing masks and are not concerned with detecting criminal activities such as the use of weapons or suspicious movements.

Ayyappa et al. [30] (2021) developed a system for mask detection using Fuzzy C-Means (FCM) and Backpropagation Neural Networks (BPNN), aiming to detect people who are not wearing masks during the pandemic. While this method is effective for face mask detection, it does not extend to detecting weapons or identifying suspicious behavior in video footage, which are essential for forensic video analysis in criminal investigations.

While mask detection is a relevant research area, it is only one part of the larger forensic investigation process. Our research intends to combine mask detection with other critical object detection techniques, such as identifying guns and suspicious behaviors, which are key to preventing and investigating criminal activities.

2.5 Credit card fraud detection and feature engineering

Hybrid Outlier-based Bagging Algorithm (HOBA) is a feature engineering methodology designed to enhance credit card fraud detection by combining outlier detection, feature selection, and bagging within a deep learning architecture. It improves detection accuracy by identifying key features that distinguish fraudulent transactions from legitimate ones. Genetic Algorithms (GA) are optimization techniques inspired by natural selection, used to evolve solutions by iteratively selecting, combining, and mutating candidate solutions. In credit card fraud detection, GA is applied for feature selection, improving model performance by identifying the most relevant features for accurate predictions. Zhang et al. [31] (2021) focused on a hybrid feature engineering approach integrated with deep learning to boost fraud detection capabilities. Ileberi et al. [32] (2022) utilized genetic algorithms to enhance the feature selection process, optimizing machine learning models for fraud detection. Both papers contribute novel methodologies to the growing field of credit card fraud detection, highlighting the importance of feature selection and innovative algorithmic approaches.

2.6 Human activity recognition

The field of Human Activity Recognition (HAR) using deep learning techniques has seen significant advancements in recent years, with various novel architectures proposed to enhance the accuracy and efficiency of activity classification. Below is a review of key contributions in this area:

In this work, Mim et al. [33] (2023) proposed GRU-INC, a hybrid model combining Gated Recurrent Units (GRU) with Inception and Attention mechanisms to address the challenges in human activity recognition. GRUs, which are particularly effective at processing sequential data, are paired with Inception modules for feature extraction, while attention mechanisms enable the model to focus on important parts of the input sequence. This approach significantly improves the model's performance in recognizing complex activities from sensor data, showing better generalization compared to traditional methods. The use of GRU helps capture temporal dependencies, and the Inception-attention module improves both accuracy and interpretability.

Khatun et al. [26] (2022) introduced a Deep CNN-LSTM with a Self-Attention model for HAR using data from wearable sensors. The model combines the Convolutional Neural Network (CNN) for automatic feature extraction with the Long Short-Term Memory (LSTM) network to capture long-range temporal dependencies in sequential data. The addition of a self-attention mechanism enhances the model's ability to focus on critical features in the input sequence, leading to improved activity recognition performance. This hybrid approach leverages the strengths of both CNNs (for spatial feature extraction) and LSTMs (for temporal dependencies), while self-attention enhances the model's capability to interpret important contextual information.

Sarcar et al. [27] (2021) explored the use of a CNN-LSTM model to detect violent arm movements, a specific application within HAR. By combining CNNs for feature extraction with LSTMs for temporal analysis, the model can accurately identify violent movements from sensor data. This approach is particularly useful in contexts such as security surveillance or health monitoring, where detecting abnormal activities, like aggressive arm movements, is critical. The hybrid architecture improves sensitivity and specificity, making it suitable for real-time applications in monitoring and safety systems.

2.7 Real-time surveillance and anomaly detection

Real-time surveillance and anomaly detection using deep learning techniques enable the automatic identification of unusual behaviors or threats in dynamic environments, such as crowds or security settings. These systems leverage models like CNNs and RNNs to process video or sensor data, ensuring rapid and accurate detection of anomalous events for enhanced safety and security. Rezaee et al. [28] (2024) provide a survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance. Although not directly related to HAR, this work is highly relevant

as it focuses on detecting anomalies in large crowds, which can involve human activity patterns. The authors highlight various deep learning methods, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), for real-time detection of unusual or suspicious behaviors in crowd surveillance systems. By leveraging deep learning techniques, this approach can quickly and accurately identify threats or abnormal patterns in a crowd, improving safety in surveillance systems. The studies reviewed reveal several important findings and limitations:

- General object and activity detection in video footage has been explored but is often limited to basic human activities or object recognition, without focusing on specific threatening behaviors such as fighting, gun usage, or other criminal activities.
- Event detection, such as fire detection, while useful in certain contexts, does not directly address the identification of weapons or suspicious actions that are critical for crime investigations.
- While video manipulation detection and IoT network security solutions have improved video forensics in terms of video integrity, they do not solve the problem of detecting dangerous objects or criminal behavior in videos.
- Mask detection has been widely studied, but this is usually in the context of public health (i.e., during the COVID-19 pandemic), and there is limited research on combining mask detection with the identification of guns or other suspicious activities.

Given these gaps, our proposal aims to bridge these issues by developing a comprehensive system that can effectively detect masks, guns, and suspicious movements in video footage, particularly in challenging scenarios such as low-quality videos or crowded environments. This system leverages advanced deep learning techniques, such as YOLOv3 for real-time object detection and CNNs for feature extraction, to provide a more accurate, efficient, and scalable solution for video forensics in criminal investigations.

3. Proposed methodology

Crime is increasing daily in various fields, such as banks, parks, hospitals, etc. Hence, video forensic techniques, an effective and efficient technique for identifying criminals, are needed. The proposed methodology uses YOLO architecture to identify masks, guns, and suspicious activity from videos. It is essential to investigate and assess significant learning structures to identify early inconsistencies in accounts. In the composition review, the pre-trained model achieves high-quality results at a modest computational cost. The video forensic approach can increase the accuracy of video recognition by analyzing Deep CNNs. It also live-tracks criminals' every step and delivers raw data to law enforcement officials. Distinguishing norms from abnormalities relies on machine learning and Artificial Intelligence (AI). The YOLO architecture outperforms other AI architectures in making accurate predictions and is used in different fields to identify objects. Figure 1 shows the block diagram of the proposed methodology of the video forensics system.

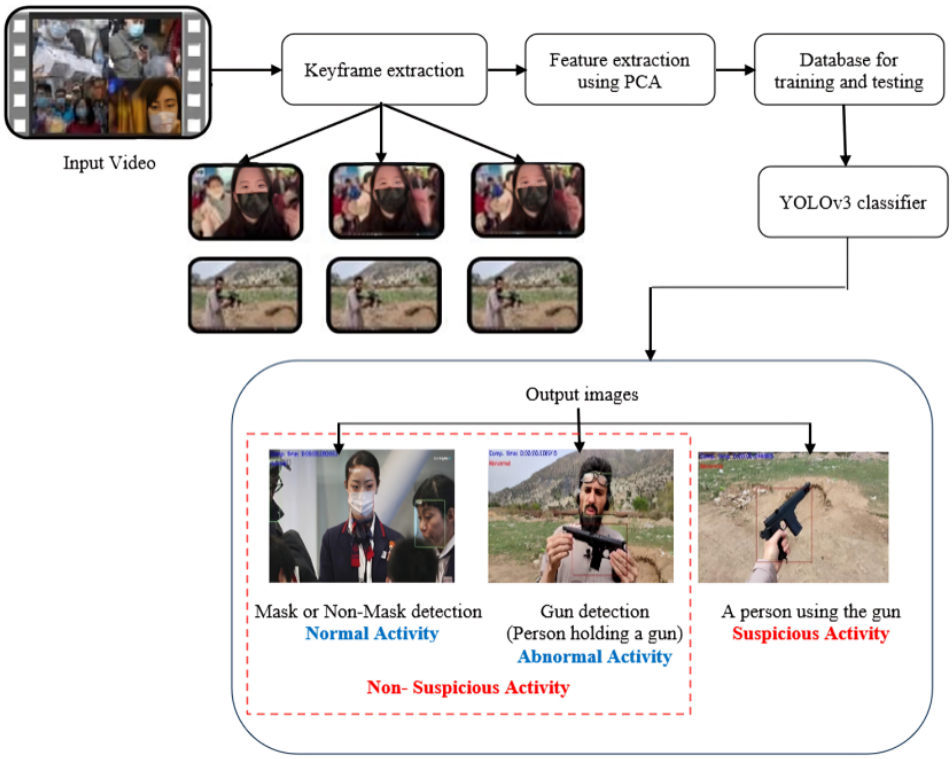


Figure 1. Block diagram of proposed methodology of the video forensics system.

https://dx.doi.org/10.57647/mjee.2026.2001.06

3.1 Input video

The proposed system can detect masks, guns, and suspicious behavior in the video. The input video can be captured through closed-circuit television (CCTV) or a camera and mobile phones.

3.2 Keyframe extraction

Camera data (e.g., position and orientation) can be used to find keyframes within the video. This is to find different types of information about the video (like drenching, brightness/contrast, vibration, dark/light, and region of the movement). The issues with keyframe extraction are that many pre-processing steps are required, and generic frame extraction methods must perform better [34, 35].

3.3 Feature extraction using PCA

The features that result from feature extraction are the product of a combination of the current features and will contain a condensed and more simplified set of information. Features can be extracted from the data. These new features highlight the essential information in the dataset rather than not being able to read into it. Principal Component Analysis (PCA) is a form of machine learning used to extract the most essential features. It will allow the user to pick only the most relevant components and ignore noise while generating data. PCA can be used to find and disclose data inconsistencies and make exceptions to rules [36]. When PCA is used on a dataset with three or more layers, it is more effective than when it is used on a single-layer or 2-layer dataset. PCA with three or more layers can account for 100% of the variance of the data and can thus achieve perfect reconstruction.

3.4 Database for training and testing

The extracted frames and features are used to create a complete database with which the system is trained and tested to identify weapons, masks, or suspicious persons. Features extracted from the data can be used for various purposes, such as detecting multiple objects.
The step-by-step flow of network classifier architecture is shown in figure 2:

1. Input Video Frame: A video frame (or a pre-processed image) is passed into the model.

2. Feature Extraction: The CNN layers extract relevant features from the input image. The convolutional filters learn to recognize patterns like edges, corners, textures, and more complex patterns as the image progresses through the network.

3. Pooling: As the feature maps progress through the layers, max-pooling reduces their spatial dimensions, focusing on the most critical features and making the model less sensitive to small changes in position.

4. Detection (YOLOv3 Mechanism): Once feature extraction is complete, the model applies the detection head (YOLO-style) to predict bounding boxes for the objects of interest. This includes detecting objects like masks, guns, and suspicious movements.

5. Output Layer: The final output consists of bounding boxes with class labels and confidence scores. If an object is detected (e.g., a gun or a mask), the system can determine its position in the image and categorize it accordingly.

The working of network classifiers for video forensics are listed as follows:

- Object Detection: The model can be trained to classify and locate multiple objects (guns, masks) in video frames. For each frame, the model predicts bounding boxes around these objects and assigns class labels (e.g., gun, mask).

- Suspicious Activity Detection: For behaviors such as fighting or unusual movements, additional layers or temporal analysis (e.g., using LSTMs or 3D CNNs) can be added to analyze movement patterns over time, not just within single frames.

- Mask + Gun + Suspicious Activity: The network can be trained to identify combinations of objects, such as detecting a masked individual carrying a gun, which is a common scenario for video forensics in criminal investigations.

3.5 YOLOv3 Classifier

A specific instance of a theory is known as a classifier. Classifiers are used to classify items into their respective categories. YOLOv3 identifies objects in live feeds. It is a deep convolutional neural network with learned features that learn from real-time data frames. The YOLO CNN is an object detection system for real-time processing. It is much faster than the other networks and still has high accuracy [37, 38, 39, 40]. Predictions are informed by global context instead of just the central

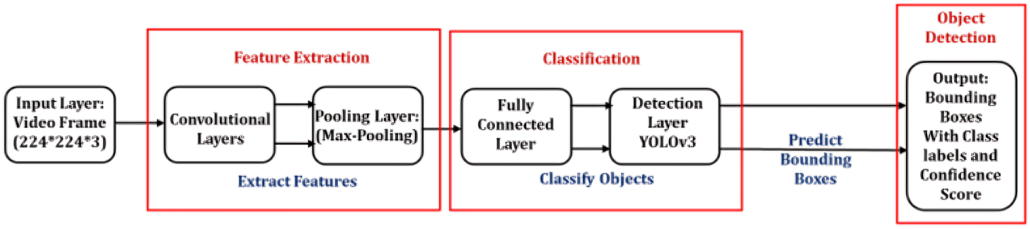


Figure 2. Network classifier architecture.

https://dx.doi.org/10.57647/mjee.2026.2001.06

object. Surveillance footage is often analyzed with computer vision techniques. YOLO segmented a mask from others in a live feed; lower-scoring regions were classified as no mask, and high-scoring regions were labeled as masks. The algorithm first maps the image by dividing it into grids. Based on the predefined classes of the objects with the highest scores, many boxes are predicted for each grid cell. The limit boxes, created by bunching the components of the ground truth boxes from the first dataset to find the most widely recognized shapes and sizes, accompany a precision score of how exact that expectation ought to be [41]. It perceives objects in a bouncing box as just a single item. A block diagram of the proposed method for distinguishing masks, guns, and suspicious people is displayed in figure 3.

The existing method for image recognition, regardless of consequences, relies on and has proven successful with PASCAL VOC benchmark datasets using a Convolutional Neural Network (CNN). It comprises 24 convolutional layers followed by two fully connected layers. By its motivation, the layers are isolated into four sections:

- Convolution and Pooling layers are structures found in artificial intelligence models and are placed 20 times consecutively.
- A helpful assortment with an objective for pregetting ready for request 224 × 224 is used.
- Detection, reduction, and convolution work in video forensics.
- The most advanced convolutional layers are ready for more learning power. In the final four convolutional layers, the connection is crucial, complemented by two associated layers.
- Objects need to have extra granular information for objects to appear in the enlightening category.
- The last layer heavily relies on linear boundaries, whereas the preceding convolutional layers were initialized using broken ReLU functions.
- The data image is sized at 448 by 448 pixels, with the output being the items listed as required.

Customized CNN engineering is outlined for the distinguishing proof and grouping of pictures for the facemasks of the people [42, 43, 44, 45, 46, 47, 48]. This philosophy is mainly founded on Convolutional Neural Networks (CNNs). CNN recognizes and classifies images in light of recently educated ascribes. It's incredibly fruitful while procuring and assessing the fitting elucidates of pictures in a multifaceted system. Artificial Neural Networks (ANN), which join numerical models through neurons and associated networks, copy the human sensory system working. The meaning of the ANN is found through the directed learning strategy. CNN alludes to the headway of ANNs, which are typically centered around structures of repeating designs in spaces like example, recognition, or representation [49]. The principal element of ANNs is that, in contrast with conventional Feed Forward Neural Networks (FFNNs), they require fewer neurons because of the layering procedure used [50]. The summed-up CNN design, which mirrors the proposed system for the recognition of facial coverings, can be referenced in figure 4.

The generalized CNN engineering is comprised of a few layers. These are the convolutional layers with ReLu activation function, the most incredible pooling layers, and the complete associated layers. The elements are separated by using the convolutional layer and the pooling layer [51, 52, 53]. The removed highlights are passed into the straightened layer to get into a solo layered exhibit to pass into completely associated layers to distinguish, regardless of whether the perceived face had a veil. CNN architecture is proposed for the

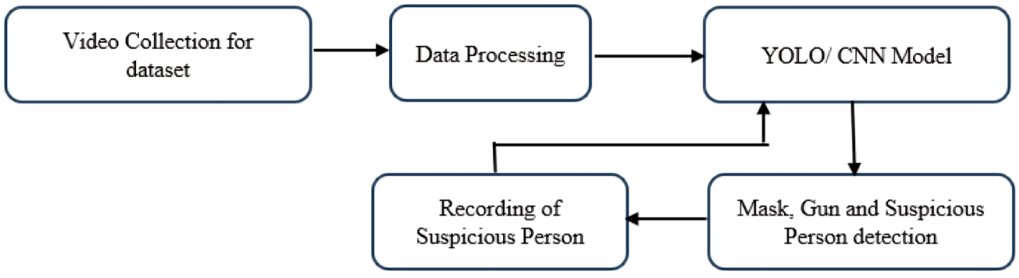


Figure 3. Block diagram of the proposed methodology for the detection of suspicious persons, guns, and masks.

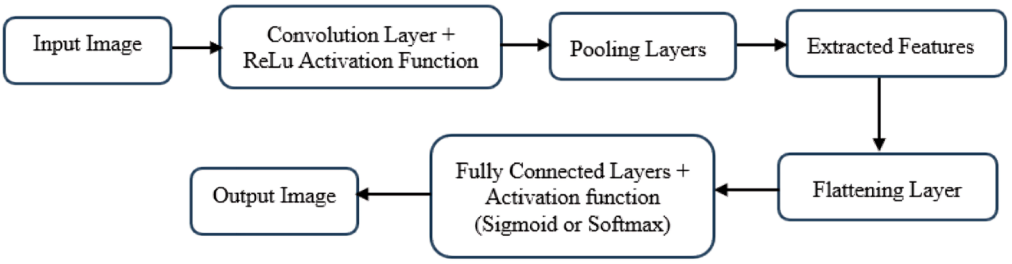


Figure 4. Generalized CNN Architecture.

https://dx.doi.org/10.57647/mjee.2026.2001.06

identification of guns as well as masks. The proposed engineering comprises five convolutional layers, three greatest pooling layers, four completely associated layers, and, at last, clump standardization appended after each layer aside from the result layer; in this manner, eight group standardization layers were used. Bunch standardization is fundamental for settling the advancing inside every unit of the profound learning network by normalizing the comparing mean and fluctuation [54]. Because of these layers, the presentation of the profound brain organization will be improved. Other than these layers, dropout layers are additionally added in the proposed CNN design to stay away from the overfitting of the model. Different boundaries that are used for building the proposed CNN design are referenced in Table 1; the engineering of the proposed CNN model is addressed with layer-wise data, as shown in figure 5.

The latest generation of cybercrime is a battle that people and organizations fight daily. The rise of this type of assault has made it challenging to stay on top of these crimes because criminals are using new strategies. One example is how these criminals act as normal posts but can easily cover up their digital intrusion under an outer layer that looks like what the user would click on [55, 56, 57]. This system uses AI for deep learning-based recognition, which means it trains data by giving input video to be trained to identify masks, guns, and suspicious persons. In addition, the same database used to identify masked persons can also be used to find them in those input videos. Some hazards come with running a network, and the way to battle these is by using someone who can cut through the staggering amounts of data and find potential digital evidence. Forensic experts can help in finding out this information in the network [58, 59, 60, 61]. Decision-making is the final step in evaluating videos in forensics investigations. The decision may be presented to law enforcement officials as proof of authenticity. After the AI has analyzed the video, it sets up a case with its different categories. There are two types of a mask: “M” for those who wear their mask up or off their heads, and ”NM” for those who aren’t masked. Whenever a procedure is finished, move to the next step or stop if nothing is left to be done. To suppress the boxes returned by the model, you need to create the following files:

- obj.names: This file contains the classes.

Table 1. Details of the parameters considered for the proposed CNN Architecture.

Parameter	Details
Optimizer function	Stochastic Gradient Descent
Learning rate	0.001
Activation function in intermediatory layers	ReLu
Dropout	0.3
Activation function in the final layers	Sigmoid
Loss function	Categorical Cross-Entropy

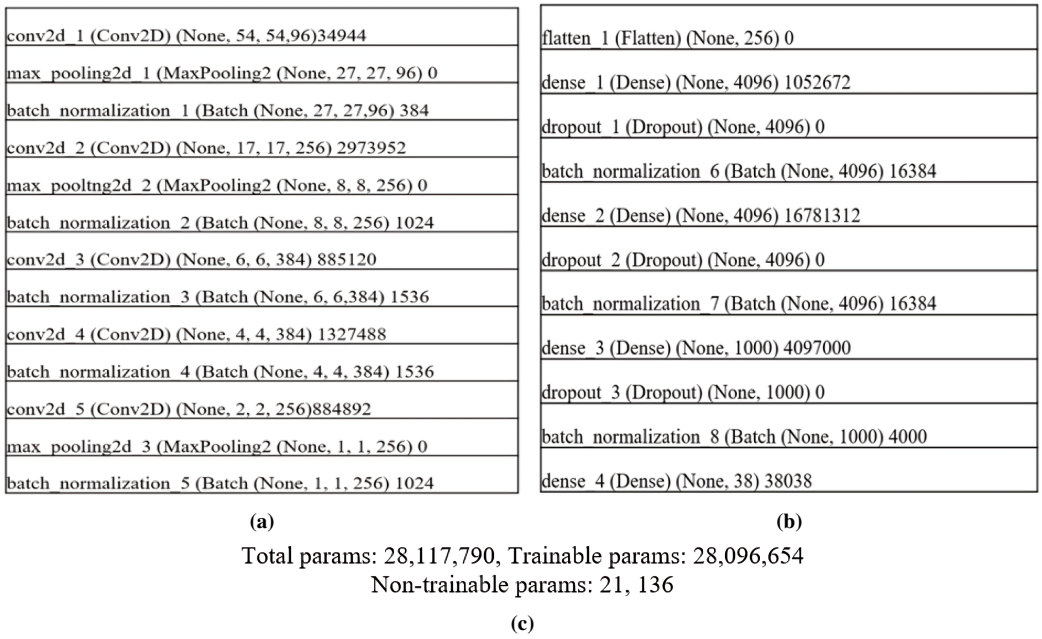


Figure 5. Details of proposed CNN architecture: (a) the structure of convolutional layers, (b) the structure of fully connected layers, and (c) parameter details of proposed CNN architecture.

https://dx.doi.org/10.57647/mjee.2026.2001.06

- obj.data: This file contains the path to the obj.names file, a training file, a validation file, and a backup location for the model (e.g., Google Drive).

A dataset folder containing 'Mask' labelled as a mask, 'Gun' labelled as a gun, and 'NM' labelled as no mask must be created in the darknet/data folder. The file train.txt should contain the path to all the dataset's images and be compiled in the darknet/data folder. From this point, there is an option to download pre-trained weights for Tiny-YOLOv3. In the end, the person who is using the gun is identified as a suspicious person. If the process is stopped here, it will start again and continue where it left off.

4. Experimental results and discussion

The proposed work is executed on a PC with a Windows 10 environment, which contains an Intel Core i79750H CPU with a 2.6 GHz base rate and 4.5 GHz maximum rate and an NVIDIA GeForce RTX 2060 GPU using Python. The proposed method is tested on various datasets such as GitHub [34], Kaggle [35, 36], and own-created datasets. GitHub presents some of the top open data datasets available for use and reuse by researchers, practitioners, and students. The Kaggle dataset provides a place for data scientists to meet and share information to explore research work, build models, participate in competitions, and publish datasets such as images with masks, images without masks, images of guns, images of vehicles, etc. The Kaggle dataset comprises images that are 25,000, including ones with masks and ones without masks. It has 15,000 images with masks and 10,000 without masks. This dataset consists of two classes: people who are wearing a facemask and those who are not. The details about these two classes of the Kaggle dataset are shown in Table 2. Own created dataset includes images from videos of crowded places with people wearing masks, people without masks, and people using guns collected through mobile phones and the videos available on YouTube. The own dataset contains images with guns that have a total size of 1346 images, and mask information has a length of 1043 images. The details about these two classes of the own created dataset are shown in Table 3.

Table 2. Kaggle dataset.

Dataset	No. of Instances
Complete Dataset	25,000
Training Dataset (75%)	18,750
Testing Dataset (25%)	6,250
Facemask having instances	15,000
No Facemask having instances	10,000

Most authors have reported their work on image forensics with the help of datasets such as Kaggle and GitHub.

Table 3. Own-created dataset.

Dataset	No. of Instances
Complete Dataset	2389
Training Dataset (75%)	1792
Testing Dataset (25%)	597

No work on video forensics has been reported till date. Moreover, no prominent work has been done on video forensics for gun and suspicious person identification. Hence, a novel approach for gun, mask, and suspicious activity detection must be developed simultaneously using soft computing techniques. The results obtained using the proposed methodology are compared with those obtained by a customized CNN network based on various evaluation metrics such as accuracy, precision, recall, and computation time [62, 63, 64, 65]. The effectiveness of the proposed method is measured in terms of different evaluation parameters, which are enlisted as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{Computation time} = \text{Time required to perform a computational process} \tag{4}$$

where, TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

4.1 Performance evaluation of the proposed method

The performance of the proposed method is tested for gun and mask detection on various datasets such as GitHub, Kaggle, and own dataset using customized CNN and YOLO architectures. The performance metrics such as accuracy, precision, recall, and F1-score for gun and mask detection are shown in Table 4 and Table 5, respectively.

The performance difference between YOLO and Customized CNN shows that the YOLO architecture is even more successful in detecting objects such as guns and masks. This indicates that the YOLO architecture outperforms customized CNN architectures. One possible reason why the machine translation is not accurate for customized CNN is that initial weights are not randomly initialized, and the second possible reason is the pretraining of the YOLO model on the ImageNet dataset.

The Loss vs Epoch graph typically shows how the training loss (or sometimes validation loss) changes with each epoch during training and this is shown in figure 6 (a).

• For YOLO (Gun Detection and Mask Detection):

Table 4. Performance evaluation metrics for gun detection.

Architectures	Performance evaluation metrics			
	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
YOLO	100	100	100	100
Customized CNN	61.54	57	28.57	44.44

Table 5. Performance evaluation metrics for mask detection.

Architectures	Performance evaluation metrics			
	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
YOLO	100	100	100	100
Customized CNN	61	60	75	61

- Since YOLO achieves perfect performance (100% accuracy, 100% precision, 100% recall, and 100% F1-score), it is likely that its loss would decrease rapidly during the early epochs and converge quickly. This suggests that YOLO is an efficient model with fast convergence and minimal loss over time.
- After reaching low values, the loss curve should flatten out since the model is already performing optimally, implying minimal further improvement.

• For Customized CNN (Gun Detection):

- The Customized CNN has significantly lower performance in all metrics: 61.54% accuracy, 57% precision, 28.57% recall, and 44.44% F1-score.
- The loss curve for this model would likely start higher compared to YOLO, indicating poorer initial performance. During the training, the loss would generally decrease over time, but not as quickly as YOLO, since the model is struggling with overfitting or under fitting.
- The curve might plateau sooner than YOLO's because of its suboptimal performance across different metrics.

• For Customized CNN (Mask Detection):

- Here, the Customized CNN shows slightly better performance than in the gun detection task, with 61% accuracy, 60% precision, 75% recall, and 61% F1-score.
- The loss curve for this model would start high and gradually decrease but still not match YOLO's smooth convergence. It might have more oscillations or a slower descent, indicating that the model is improving but faces difficulties in achieving optimal performance.

The Accuracy vs Epoch graph displays how the model's accuracy improves as the number of epochs increases. This graph typically starts low and increases over time as the model learns and this is shown in figure 6 (b).

• For YOLO (Gun Detection and Mask Detection):

- Since YOLO achieves perfect accuracy (100%) from the beginning, the accuracy graph would immediately reach a flat line at 100% from epoch 1, indicating no further improvement after the initial stages.
- The accuracy curve would show a sharp rise to 100% at the first epoch and stay constant throughout

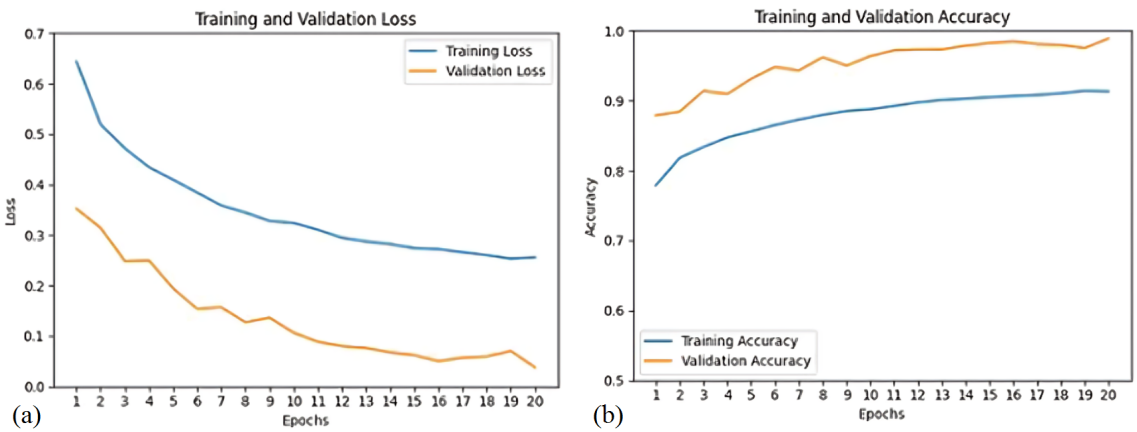


Figure 6. Graphs for (a) Loss vs Epoch and (b) Accuracy vs Epoch.

https://dx.doi.org/10.57647/mjee.2026.2001.06

training.

• For Customized CNN (Gun Detection):

- The Customized CNN has a relatively low accuracy in both gun detection (61.54%) and mask detection (61%).
- For Gun Detection, the accuracy curve would start low and gradually increase with each epoch, but since it's not achieving high performance, the curve would likely level off at around 60% in later epochs.
- For Mask Detection, the accuracy curve might follow a similar pattern, but potentially with a slightly higher plateau around 60-65%, given its better recall (75%) and F1-score (61%).

K-fold cross-validation is a robust technique for evaluating the performance of machine learning models. In this approach, the dataset is divided into K subsets, and the model is trained K times, each time using K-1 subsets for training and the remaining subset for testing. This helps to assess the model's generalizability by ensuring that every data point is used for both training and testing, minimizing overfitting. The performance evaluation metrics-Accuracy, Precision, Recall, and F1-Score—are calculated for each fold, and their averages are used to summarize the model's overall performance and shown in figure 7.

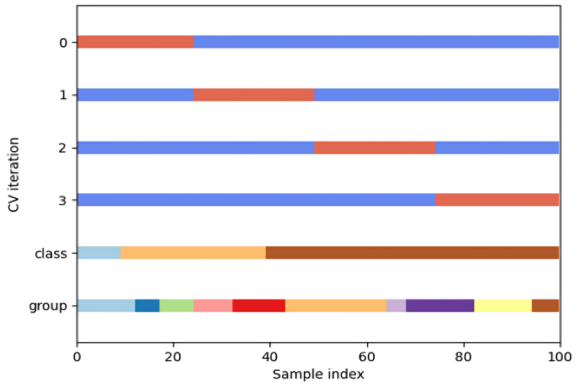


Figure 7. K-fold cross validation of proposed system.

Performance evaluation:

- Gun Detection: YOLO achieves perfect performance across all metrics, while the customized CNN has lower accuracy (61.54%), precision (57%), recall (28.57%), and F1-score (44.44%).
- Mask Detection: YOLO again performs flawlessly, while the customized CNN shows moderate performance with accuracy (61%), precision (60%), recall (75%), and F1-score (61%).

This approach ensures a more reliable assessment of the proposed system's performance across different scenarios.

For gun detection:

• Step 1: Split the Dataset into 4 Folds
The dataset has 100 samples. These samples are randomly divided into 4 folds. Each fold contains 25 samples.

- Fold 1: 25 samples
- Fold 2: 25 samples
- Fold 3: 25 samples
- Fold 4: 25 samples

• Step 2: Training and Testing on Each Fold
For each fold, the model is trained on 75 samples (3-folds) and tested on the remaining 25 samples (1-fold).
1st Iteration:

- Train on Fold 2, Fold 3, Fold 4, test on Fold 1
- Compute performance metrics: accuracy, precision, recall, F1-score

2nd Iteration:

- Train on Fold 1, Fold 3, Fold 4, test on Fold 2
- Compute performance metrics: accuracy, precision, recall, F1-score

3rd Iteration:

- Train on Fold 1, Fold 2, Fold 4, test on Fold 3
- Compute performance metrics: accuracy, precision, recall, F1-score

4th Iteration:

- Train on Fold 1, Fold 2, Fold 3, test on Fold 4
- Compute performance metrics: accuracy, precision, recall, F1-score

• Step 3: Average the Results Across All Folds
After testing on each fold, you will have 4 sets of performance metrics for each model. The final performance for 4-fold cross-validation is the average of the metrics across all folds.
For example: • YOLO's Performance (Gun Detection):

- Accuracy: 100% in each fold → Average Accuracy = 100%
- Precision: 100% in each fold → Average Precision = 100%
- Recall: 100% in each fold → Average Recall = 100%
- F1-Score: 100% in each fold → Average F1-Score = 100%

• Customized CNN's Performance (Gun Detection):

- Accuracy: 61.54% (same as provided, may vary slightly in real cross-validation folds)

https://dx.doi.org/10.57647/mjee.2026.2001.06

- Precision: 57%
- Recall: 28.57%
- F1-Score: 44.44%

So, after 4-fold cross-validation, YOLO would still have perfect performance across all metrics, while the customized CNN's performance will remain lower, reflecting the metrics you provided earlier.

For mask detection:

Let's now apply 4-fold cross-validation to Mask Detection with similar steps.

• Step 1: Split the Dataset into 4 Folds

Again, there are 100 samples for mask detection, and they are randomly divided into 4 folds with 25 samples per fold.

- Fold 1: 25 samples
- Fold 2: 25 samples
- Fold 3: 25 samples
- Fold 4: 25 samples

• Step 2: Training and Testing on Each Fold

For each fold, we train on 75 samples (3 folds) and test on 25 samples (1 fold). This process will be repeated for all 4 folds.

1st Iteration:

- Train on Fold 2, Fold 3, Fold 4, test on Fold 1
- Compute performance metrics: accuracy, precision, recall, F1-score

2nd Iteration:

- Train on Fold 1, Fold 3, Fold 4, test on Fold 2
- Compute performance metrics: accuracy, precision, recall, F1-score

3rd Iteration:

- Train on Fold 1, Fold 2, Fold 4, test on Fold 3
- Compute performance metrics: accuracy, precision, recall, F1-score

4th Iteration:

- Train on Fold 1, Fold 2, Fold 3, test on Fold 4
- Compute performance metrics: accuracy, precision, recall, F1-score

• Step 3: Average the Results Across All Folds

The performance metrics from all 4 iterations will be averaged to get the final result.

For example:

• YOLO's Performance (Mask Detection):

- Accuracy: 100% in each fold → Average Accuracy = 100%
- Precision: 100% in each fold → Average Precision = 100%
- Recall: 100% in each fold → Average Recall = 100%
- F1-Score: 100% in each fold → Average F1-Score = 100%

• Customized CNN's Performance (Mask Detection):

- Accuracy: 61%
- Precision: 60%
- Recall: 75%
- F1-Score: 61%

The confusion matrix in Fig. 8 (a) shows that there are 22 False Positives (FP), 28 True Negatives (TN), 28 True Positives (TP), and 22 False Negatives (FN). This indicates a balanced number of correct and incorrect predictions across both classes. The confusion matrix in Fig. 8 (b) reveals 20 False Positives (FP), 40 True Negatives (TN), 30 True Positives (TP), and 10 False Negatives (FN). This suggests a relatively strong performance with a higher number of correct predictions, especially for the negative class.

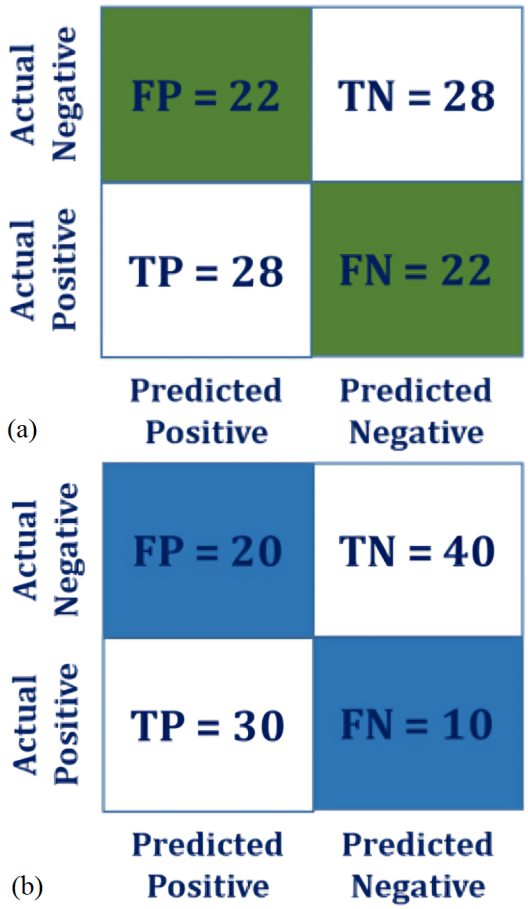


Figure 8. Estimated Confusion Matrix for: (a) Gun Detection and (b) Mask Detection by using Customized CNN).

4.2 Comparative analysis

Comparative analysis is carried out on various public datasets such as GitHub, Kaggle, and own-created datasets with the previously reported methods. Table 6 shows a comparison of the proposed methodology with existing techniques for mask detection.

Table 6 shows that the proposed system achieved 100% accuracy in detecting masks, even in low-quality videos. It is also observed that the NASNetMobile technique gives an accuracy of 97.54% for mask detection on its own created dataset prepared from freely available datasets and real-world images. The accuracy for customized CNN and CNN-LSTM is significantly less than the proposed methodology.

The YOLOv3 architecture demonstrates exceptional novel performance, achieving 100% accuracy, precision, recall, and F1-score for both mask detection and gun detection. In contrast, the Customized CNN model shows significantly lower results, with an accuracy of 61%, precision of 57%, recall of 63%, and an F1-score of 59.85% for mask detection. For gun detection, the Customized CNN yields an accuracy of 61%, precision of 57%, recall of 28.57%, and an F1-score of 38.06%. Additionally, the computational efficiency of the proposed YOLO methodology is outstanding, with a processing time of just 0.031 ms for both mask and gun detection. This is notably faster compared to the Customized CNN architecture, which takes 0.8 ms for mask detection and 0.7 ms for gun detection. Since most of the work is reported for mask detection [23, 24] on different datasets such as AIZOO and ViDMASK. So, comparison analysis cannot be done with other techniques as being the first to report gun detection and suspicious activity detection results on video collected from GitHub, Kaggle, and own created datasets.

4.3 Output results

The primary objective of this proposed work is divided into two segments: The first, which concerns the identification of weapons, is carried out using customized CNN and YOLO. The second segment consists of mask identification in the video. In this realm of video criminology, object recognition can be categorized into two types: Suspicious and non-suspicious activities. The non-suspicious training is again divided into normal (figure 9) and abnormal activity (figure 10). If there is no gun in the frame, it is identified as normal activity. Once the gun is identified, either kept somewhere or held by a person, it is marked as abnormal activity (figure 11). A suspicious movement potentially has a different degree of doubt. Suspicious activity is identified when a person uses the gun (figure 12), whereas non-suspicious activity is when the person is just holding the gun. Most authors have separately reported their work on mask and weapon detection, such as pistols [62, 63, 64]. No work has been reported yet for identifying a suspicious person from a video. The proposed methodology shows that suspicious persons using guns are accurately identified in a zoomed view from video inputs and stored in a separate folder. This data can then be used for criminal identification at the time of proof presented in court. This suspicious activity identification stands true for any background, low-quality, fast activity video, etc.

This particular situation, however, is situational, and the gun, mask, and no mask are being shown as "Gun," "Mask," and "NM", respectively, which roughly translates to having a gun, masked, and not masked. The input and output images obtained from the proposed methodology for the detection of masks with a single person, dual persons, multiple persons in the frame, and normal activity identification are shown in Figs. 9(a)-(d), Figs. 9(e)-(f), Figs. 9(g)-(j), respectively.

As observed from figure 9, it is seen that the proposed system is efficient in identifying masks in various situations, such as a single person with or without a mask (Figs. 9(a)-(d)), the dual person with or without a mask (Figs. 9(e)-(f)), crowded places with persons wearing a mask and persons without a mask (Figs. 9(g)-(j)), low-quality videos (Figs. 9(c)), etc.

The input and output images obtained from the proposed methodology for mask and gun detection with different backgrounds, shapes, and sizes of guns and abnormal activity identification are shown in figure 10.

From figure 10, it is observed that the proposed system identifies guns irrespective of having different shapes (Figs. 10(a)-(c), Figs. 10(e)-(g), Figs. 10(k)), sizes (Figs. 10(a)-(c), Figs. 10(e)-(g), Figs. 10(k)), colors (figs. 10(a)-(b), Figs. 10(h)) and angles of holding the gun (Figs. 10(a)-(k)). It also gives efficient output even with different types of backgrounds (Figure 10(a)-(b), figure 10 (e), figure 10 (g), figure 10 (k)). All these are identified as abnormal activities, as each one is holding a gun. Once the person uses the gun, they are captured in a zoomed view as a suspicious person.

Table 6. Comparison of the proposed method with existing techniques for mask detection.

Sr. No.	Techniques used	Dataset used	Performance evaluation metrics		
			Accuracy (%)	Precision (%)	Recall (%)
1.	YOLO Architecture	GitHub, Kaggle and own dataset	100	100	100
2.	Customized CNN	GitHub, Kaggle and own dataset	61.54	57	28.57
3.	CNN-LSTM [23]	AIZOO	61.00	56	-
4.	NASNetMobile [24]	ViDMASK	97.54	96.28	100

https://dx.doi.org/10.57647/mjee.2026.2001.06

Sr. No.	Input image before mask detection	Output image after mask detection
(a)		
(b)		
(c)		
(d)		
(e)		
(f)		
(g)		
(h)		

https://dx.doi.org/10.57647/mjee.2026.2001.06

Sr. No.	Input image before mask detection	Output image after mask detection
(i)		
(j)		

Figure 9. Images before and after detection of masks with a single person, dual person, multiple persons in the frame, and normal activity identification from video frames.

Sr. No.	Input image before mask and gun detection	Output image after mask and gun detection
(a)		
(b)		
(c)		
(d)		

https://dx.doi.org/10.57647/mjee.2026.2001.06

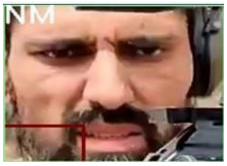
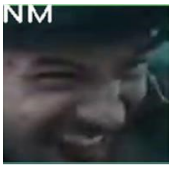
Sr. No.	Input image before mask and gun detection	Output image after mask and gun detection
(e)		
(f)		
(g)		
(h)		
(i)		
(j)		
(k)		

Figure 10. Images before and after mask and gun detection with different backgrounds, shapes, sizes, and abnormal activity identification from video frames.

https://dx.doi.org/10.57647/mjee.2026.2001.06

The input and output images obtained from the proposed methodology for identifying a person from a suspicious activity video are shown in Figure 11.

Figure 11 shows images from the video before and after identifying normal, abnormal, and suspicious activity. Figure 11(a)-(b) and Figs. 11(c)-(d) show normal and abnormal activity detection, respectively. The zoomed view of the suspicious person is captured and stored in a separate folder. The suspicious activity detection can be seen in Figs. 11(e)-(f).

Sr. No.	Input image before suspicious person identification	Output image after suspicious person identification
(a)		Non-suspicious activity
(b)		
(c)		
(d)		
(e)		
(f)		
(g)		

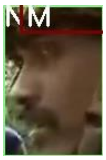
Sr. No.	Input image before suspicious person identification	Output image after suspicious person identification
(h)		
(i)		

Figure 11. shows images from the video before and after identifying normal, abnormal, and suspicious activity. Fig. 11(a)-(b) and Fig. 11(c)-(d) show normal and abnormal activity detection, respectively. The zoomed view of the suspicious person is captured and stored in a separate folder. The suspicious activity detection can be seen in Fig. 11(e)-(f).

Sr. No.	Activity detection	Input image before activity identification	Output image after activity identification	Detection
(a)	Normal activity detection			Mask detection
(b)	Normal activity detection			Non-Mask detection
(c)	Abnormal activity detection			Only holding a gun (Gun detection)
(d)	Abnormal activity detection			Non-Mask and gun detection
(e)	Suspicious activity detection			Non-Mask detection + This person is also using the gun to act as a Suspicious person identified in a zoomed view and stored in a separate folder.
(f)	Suspicious activity detection			Non-Mask detection + This person is also using the gun to act as a Suspicious person identified in a zoomed view and stored in a separate folder.

Figure 12. Images before and after identification of normal, abnormal, and suspicious activity detection from video.

https://dx.doi.org/10.57647/mjee.2026.2001.06

5. Conclusion

This paper proposes a novel framework for identifying normal, abnormal, and suspicious activity within the video. The three components are the gun detection model, mask detection, and suspicious person detection using the YOLO model for each, which surpasses more traditional architecture models such as CNNs [45], VGGs [30], and MobileNetV2s [46]. A framework for digital video forensics is also proposed. This framework comprises two segments: gun identification and mask identification. The YOLO architecture was more successful than the custom CNN architecture because it didn't randomly initiate weights and included a pretraining section. The YOLO architecture can simultaneously recognize objects such as guns and masks, which is one more advantage over old-world architectures. The accuracy, precision, and recall of YOLOv3 architecture are 100%, 100%, and 100% respectively for mask detection and 100%, 100%, and 100% respectively for gun detection as compared to Customized CNN with an accuracy of 61%, the precision of 60%, recall of 75% for mask detection and accuracy of 61.54%, the precision of 57%, recall of 28.57% for gun detection. Using a larger dataset of actual criminal records can further enhance the YOLO architecture's performance. A greater dataset can be created by thorough assortment from different sources; on the off chance that not, delicate registering procedures become helpful to work on the size and extent of the dataset. The proposed framework for identifying normal, abnormal, and suspicious activities within video footage demonstrates significant advancements in digital video forensics, particularly in gun and mask detection. By utilizing the YOLO architecture, which excels over traditional models like CNNs, VGGs, and MobileNetV2, this system achieves 100% accuracy, precision, and recall for both gun and mask detection. The success of YOLO can be attributed to its pre-training phase and the ability to simultaneously detect multiple objects, making it a powerful tool for real-time surveillance and security applications. The findings suggest that incorporating larger datasets, particularly from real-world criminal records, could further enhance the performance of this system, ensuring its continued effectiveness in tackling complex and sensitive data challenges. This framework, leveraging deep learning's capabilities, marks a critical step towards automating and improving routine security tasks, with the potential for wide-reaching applications in various sectors.

Acknowledgment
The result analysis and simulation part has been done by Sunpreet Kaur Nanda, Introduction and discussions have been written by Dr. Deepika Ghai and Dr. Suman Lata Tripathi, the comparative analysis has been made by Dr. Sandeep Kumar and the manuscript has been reviewed by Prashant Ingole. The author gratefully acknowledges Sunpreet Kaur Nanda, Deepika Ghai, Prashant Ingole, Sandeep Kumar, and Suman Lata Tripathi for their work on the original version of this document.

Authors contributions
All authors contributed equally to the conception, design, execution, and writing of this work. All authors read and approved the final manuscript.

Availability of data and materials
The authors declare that the data supporting the findings of this study are available within the paper.

Conflict of interests
The authors assert that they do not have any identifiable conflicting financial interests or personal relationships that might be perceived to influence the work presented in this paper.

References

1. Bengio Y. "**Learning deep architectures for AI.**" Found. Trends Mach. Learn. 2009; 2:1–127. DOI: 10.1561/2200000006
2. Hinton GE, S.Osindero, and Teh YW. "**A fast learning algorithm for deep belief nets.**" Neural Comput. 2006; 18:1527–54. DOI: 10.1162/neco.2006.18.7.1527
3. Deng L. "**Expanding the scope of signal processing.**" IEEE Signal Processing Mag. 2008; 25:2–4. DOI: 10.1109/MSP.2008.920380
4. Tang Y and Eliasmith C. "**Deep networks for robust visual recognition.**" 27th International Conference on Machine Learning 2010 :1055–62. DOI: 10.5555/3104322.3104456
5. Hinton GE. "**A practical guide to training restricted Boltzmann machines.**" Univ. Toronto, Tech. Rep. 2012; 7700. DOI: 10.1007/978-3-642-35289-8_32
6. Kushwaha AK and Wadhe A. "**Design and Implementation of Forensic Framework for Video Forensics.** " Int. J. Curr. Eng. Technol. 2015; 5:1015–8
7. Milani S, Fontani M, Bestagini P, Barni M, Piva A, Tagliasacchi M, and Tubaro S. "**An overview on video forensics.**" APSIPA Transactions on Signal and Information Processing 2012; 1:1–18. DOI: 10.1017/ATSIP.2012.2
8. Militante SV and Dionisio NV. "**Real-Time Facemask Recognition with Alarm System using Deep Learning.**" 11th IEEE Control and System Graduate Research Colloquium (ICSGRC) 2020; 87:106–10. DOI: 10.1109/ICSGRC49013.2020.9232610
9. Damer N, Grebe JH, C.Chen, Boutros F, Kirchbuchner F, and Kuijper A. "**The effect of wearing a mask on face recognition performance: an exploratory study.**" Int. Conf. Biometrics Special Interest Group (BIOSIG) 2020 :1–6. DOI: 10.48550/arXiv.2007.13521

10. Abudarham N, Shkiller L, and Yovel G. "**Critical features for face recognition.** " Cognition 2019; 182:73–83. DOI: 10.1016/j.cognition.2018.09.002
11. Zhi H and Liu S. "**Face recognition based on genetic algorithm.**" J. Vis. Commun. Image Represent. 2019; 58:495–502. DOI: 10.1016/j.jvcir.2018.12.012
12. Nanda SK, Ghai D, Ingole P, and S.Pande. "**Soft Computing Techniques based Digital Video Forensics for Anomaly Detection.**" J. Comput. -Assist. Methods Eng. Sci. 2022; 30:111–30. DOI: 10.24423/cames.447
13. Zhou Z, Tang D, Wang X, Han W, Xiangyu L, and Zhang K. "**Invisible mask: Practical attacks on face recognition with infrared.**" Cryptography and Security 2018. DOI: 10.48550/arXiv.1803.04683
14. Masi I, Wu Y, Hassner T, and Natarajan P. "**Deep face recognition: A survey.**" 31st SIBGRAPI Conf. on Graphics, Patterns, and Images (SIBGRAPI) 2018 :471–8. DOI: 10.1109/SIBGRAPI.2018.00067
15. Mahmood Z, Muhammad N, Bibi N, and Ali T. "**A review on state-of-the-art face recognition approaches.**" Fractals 2017; 25:1750025. DOI: 10.1142/S0218348X17500256.
16. Zhao W, Chellappa R, Phillips PJ, and Rosenfeld A. "**Face recognition: A literature survey.**" ACM Comput. Surv. 2003; 35:399–458. DOI: 10.1145/954339.954342
17. Sun Y, Liang D, Wang X, and Tang X. "**Deepid3: Face recognition with deep neural networks.**" 2013; Computer Vision and Pattern Recognition. DOI: 10.48550/arXiv.1502.00873
18. Kortli Y, Jridi M, Falou AA, and Atri M. "**Face recognition systems: A Survey.**" Sensors 2020; 20:342. DOI: 10.3390/s20020342
19. Wan Q, Panetta K, and Agaian S. "**A video forensic technique for detecting frame integrity using the human visual system-inspired measure.**" IEEE Int. Symp. on Technologies for Homeland Security (HST) 2017 :1–6. DOI: 10.1109/THS.2017.7943466
20. H.Kaur and Choudhary KR. "**Digital forensics: implementation and analysis for Google Android framework.**" Information Fusion for Cyber-Security Analytics 2017 :307–31. DOI: 10.1007/978-3-319-44257-0_13
21. Kaushal M, Khehra BS, and Sharma A. "**Soft Computing based Object Detection and Tracking Approaches: State-of-the-Art Survey.** " Appl. Soft Comput. 2018; 70:423–64. DOI: 10.1016/j.asoc.2018.05.023
22. Muhammad K, Ahmad J, Mehmood I, Rho S, and Baik SW. "**Convolutional Neural Networks based Fire Detection in Surveillance Videos.**" IEEE Access 2018; 6:18174–83. DOI: 10.1109/ACCESS.2018.2812835
23. Bajestani MF, Abadi SSHR, Fard SMD, and Khodadadeh R. "**AAD: Adaptive Anomaly Detection through Traffic Surveillance Videos.**" 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS) 2018 :1–7. DOI: 10.48550/arXiv.1808.10044
24. Conti M, Dehghantanha A, Franke K, and Watson S. "**Internet of Things security and forensics: Challenges and opportunities.** " Sci. Direct. 2018; 78:544–6. DOI: 10.1016/j.future.2017.07.060
25. Chen Q and Sang L. "**Facemask recognition for fraud prevention using the Gaussian mixture model.**" J. Vis. Commun. Image Represent. 2018; 55:795–801. DOI: 10.1016/j.jvcir.2018.08.016
26. Khatun MA, Yousuf MA, Ahmed S, Uddin MZ, Alyami SA, Al-Ashhab S, Akhdar HF, Khan A, Azad A, and Moni MA. "**Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor.**" IEEE Journal of Translational Engineering in Health and Medicine 2022; 10:1–16. DOI: 10.1109/JTEHM.2022.3177710
27. Sarcar ST and Yousuf MA. "**Detecting violent arm movements using CNN-LSTM.**" 5th IEEE International Conference on Electrical Information and Communication Technology (EICT) 2021 :1–6. DOI: 10.1109/EICT54103.2021.9733510.
28. Rezaee K, Rezakhani SM, Khosravi MR, and Moghimi MK. "**A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance.**" Personal and Ubiquitous Computing 2024; 28:135–51. DOI: 10.1007/s00779-021-01586-5
29. Lohiya R and Shah P. "**Video-Based Face Detection and Tracking for Forensic Applications.**" Int. J. Comput. Sci. & Commun. 2016; 7:210–8. DOI: 10.090592/IJCSC.2016.033
30. Ayyappa Y, Neelakanteswara P, Bekkanti A, and CMAK Z. Basha YT nd. "**Automatic Face Mask Recognition System with FCM AND BPNN.** " 5th Int. Conf. on Computing Methodologies and Communication (ICCMC) 2021 :1134–7. DOI: 10.1109/ICCMC51019.2021.9418243
31. Zhang X, Han Y, Xu W, and Wang Q. "**HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture.**" Information Sciences 2021; 557:302–16. DOI: 10.1016/j.ins.2019.05.023
32. Ileberi E, Y.Sun, and Wang Z. "**A machine learning based credit card fraud detection using the GA algorithm for feature selection.** " Journal of Big Data 2022; 9. DOI: ,vol.9,no.24,2022.doi: 10.1186/s40537-022-00573-8

33. Mim TR, Amatullah M, Afreen S, Yousuf MA, Uddin S, Alyami SA, Hasan KF, and Moni MA. "**GRU-INC: An inception-attention based approach using GRU for human activity recognition.**" Expert Systems with Applications 2023; 216:119419. DOI: 10.1016/j.eswa.2022.119419

34. Geng L, Zhang S, Tong J, and Xiao Z. "**Lung segmentation method with dilated convolution based on VGG-16 network.**" Comput. Ass. Surg. 2019; 24:27–33. DOI: 10.1080/24699322.2019.1649071

35. Srivastava S, Kumar P, Chaudhry V, and Singh A. "**Detection of Ovarian Cyst in Ultrasound Images Using Fine-Tuned VGG-16 Deep Learning Network.**" SN Comput. Sci. 2020; 1:1–8. DOI: 10.1007/s42979-020-0109-6

36. Rezaee M, Zhang Y, Mishra R, Tong F, and Tong H. "**Using the VGG-16 network for individual tree species detection with an object-based approach.**" 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS) 2018 :1–7. DOI: 10.1109/PRRS.2018.8486395

37. Islam S, Khan SIA, Abedin MM, Habibullah KM, and Das AK. "**Bird species classification from an image using the VGG-16 network.** " 7th Int. Conf. on Computer and Communications Management 2019 :38–42. DOI: 10.1145/3348445.3348480

38. Kamilaris A and Prenafeta-Boldu FX. "**Deep learning in agriculture: A survey.**" Comput. Electron. Agric. 2018; 147:70–90. DOI: 10.1016/j.compag.2018.02.016

39. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, and Dean J. "**A guide to deep learning in healthcare.**" Nature Med. 2019; 25:24–9. DOI: 10.1038/s41591-018-0316-z

40. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, and Chintala S. "**Pytorch: An imperative style, high-performance deep learning library.**" Machine Learning 2020. DOI: 10.48550/arXiv.1912.01703

41. Zhang Z, Cui P, and Zhu W. "**Deep learning on graphs: A survey.**" IEEE Trans. Knowledge Data Eng. 2022; 34:249–70. DOI: 10.1109/TKDE.2020.2981333

42. Balasubramanian Y, Sivasankaran K, and Krishraj SP. "**Forensic video solution using facial feature-based synoptic Video Footage Record.**" IET Comput. Vis. 2017. DOI: 10.1049/iet-cvi.2015.0238

43. Lillis D, Becker B, ÓSullivan T, and Scanlon M. "**Current Challenges and Future Research Areas for Digital Forensic Investigation.**" Cryptography and Security 2016. DOI: 10.48550/arXiv.1604.03850

44. Galvan F and Battiato S. "**Image/Video Forensics: Theoretical Background, Methods and Best Practices- Part Two: From Analog to Digital World.**" SICUREZZA eGIUSTIZIA S-I/MMXIX 2020 :45–50

45. Primeau E and Primeau MA. "**Video Forensic Services.**" 2020. Available from: https://www.videoforensicexpert.com/contact-us/

46. Krizhevsky A, Sutskever I, and Hinton GE. "**ImageNet Classification with Deep Convolutional Neural Networks.** " Commun. ACM 2017; 60:84–90. DOI: 10.1145/3065386

47. "**Image Category Classification using Deep Learning.**" MathWorks 2020. Available from: https://www.mathworks.com/help/vision/examples/image-category-classification-using-deep-learning.html

48. Sánchez J and Perronnin F. "**High-dimensional signature compression for large-scale image classification.**" IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) 2011 :1665–72. DOI: 10.1109/CVPR.2011.5995504

49. "**Types of Convolutional Neural Networks: LeNet, AlexNet, VGG-16 Net, ResNet and Inception Net.**" CompleteGate 2020. Available from: https://www.completegate.com/2017022864/blog/deep-machine-learning-images-lenet-alexnet-cnn/allpages

50. M.Sabokrou, Fayyaz M, Fathy M, Moayed Z, and Klette R. "**Effect of nanocatalysts on the transesterification reaction of first, second and third generation biodiesel sources - A mini-review.**" Comput. Vis. Image Underst. 2018; 172:88–97. DOI: 10.1016/j.cviu.2018.02.006

51. Jin CB, Li S, and Kim H. "**Real-Time Action Detection in Video Surveillance using Sub-Action Descriptor with Multi-CNN.** " Comput. Vis. Pattern Recognit. 2017 :1–29. DOI: 10.48550/arXiv.1710.03383

52. Vu H, Nguyen T, Travers A, Venkatesh S, and Phung D. "**Energy-Based Localized Anomaly Detection in Video Surveillance.**" In Proc. of Advances in Knowledge and Data Mining, Jeju, South Korea 2017 :1641–653. DOI: 10.48550/arXiv.2105.03270

53. Sinha RK, Pandey R, and Pattnaik R. "**Deep Learning for Computer Vision Tasks: A Review.** " Conf. on Intelligent Computing and Control 2017 :1–5. DOI: 10.48550/arXiv.1804.03928

54. Nasir M, Muhammad K, Lloret J, Sangaiah AK, and Sajjad M. "**Fog Computing Enabled Cost-Effective Distributed Summarization of Surveillance Videos for Smart Cities.**" J. Parallel Distrib. Comput. 2019; 126:161–70. DOI: 10.1016/j.jpdc.2018.11.004

55. Muhammad K, Hamza R, Ahmad J, Lloret J, Wang H, and Baik SW. "**Secure Surveillance Framework for IoT Systems using Probabilistic Image Encryption.**" IEEE Trans. on Industrial Informatics 2018; 14:3679–89. DOI: 10.1109/TII.2018.2791944

56. Munir M, Siddiqui SA, Dengel A, and Ahmed S. "**DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series.**" IEEE Access 2019; 7:1991–2005. DOI: 10.1109/ACCESS.2018.2886457

57. Singh RD and Aggarwal N. "**Video Content Authentication Techniques: A Comprehensive Survey.**" Multimedia Systems 2018; 24:211–40. DOI: 10.1007/s00530-017-0538-9

58. Sultani W, Chen C, and Shah M. "**Real-World Anomaly Detection in Surveillance Videos.** " IEEE/CVF Conference on Computer Vision and Pattern Recognition 2018; 10:6479–88. DOI: 10.1109/CVPR.2018.00678

59. Li S, Choo KKR, Sun Q, Buchanan WJ, and Cao J. "**IoT forensics: Amazon echoes as a use case.**" IEEE Internet of Things Journal 2019; 6:6487–97. Available from: 10.1109/JIOT.2019.2906946

60. Jerian M, Paolino S, Cervelli F, Carrato S, Mattei A, and Garofano L. "**A forensic image processing environment for the investigation of surveillance video.**" Forensic Science International 2011; 167:2007. DOI: 10.1016/j.forsciint.2006.06.048

61. N. Ul H, Fraz MM, Hashmi TS, and Shahzad M. "**Orientation aware weapons detection in visual data: a benchmark dataset.**" Computing 2022; 104:2581–604. DOI: 10.1007/s00607-022-01095-0

62. Nanda SK, Ghai D, and Pande S. "**VGG-16-Based Framework for Identification of Facemask Using Video Forensics.**" Proceedings of Data Analytics and Management. Lecture Notes on Data Engineering and Communications Technologies 2022; 91. DOI: 10.1007/978-981-16-6285-0_54

63. Nanda SK and Ghai D. "**Future of video forensics in IoT.**" Electronic Devices and Circuit Design Challenges and Applications in the Internet of Things 2022 :113–33

64. Nanda SK, Ghai D, and Ingole P. "**Analysis of video forensics system for gun, mask, and anomaly detection using soft computing techniques.**" Journal of Cyber Security and Mobility 2022 :549–74. DOI: 10.13052/jcsm2245-1439.1143

65. Ghai D, Saxena S, Dhingra G, and Tripathi SL. "**A comprehensive review on performance-based comparative analysis, categorization, classification and mapping of text extraction system techniques for images.**" Multimed Tools Appl 2025; 84:2327–2484. DOI: 10.1007/s11042-024-20257-0