Analysis of Video Forensics System for Detection of Gun, Mask and Anomaly Using Soft Computing Techniques

Sunpreet Kaur Nanda[1,2,*], Deepika Ghai[1] and Prashant Ingole[3]

[1]*School of Electronics and Electrical Engineering, Lovely Professional University, Punjab – 144 411, India*
[2]*Electronics and Communication Engineering Department, P. R. Pote College of Engineering and Management, Amravati – 444 602, India*
[3]*Department of Information Technology, Prof. Ram Meghe Institute of Technology & Research, Amravati – 444607, India*
E-mail: sunpreetkaurbedi@gmail.com
[*]*Corresponding Author*

Received 25 November 2021; Accepted 15 April 2022;
Publication 07 November 2022

Abstract

The video forensics world is a developing network of experts associated with the computerized video forensics industry. With quickly developing innovation, the video turned out to be the most significant weapon in the battle against individuals who violate the law by catching them in the act. Proof caught on video is viewed as more dependable, more exact, and more persuading than observer declaration alone. But, proof can be effortlessly tempered by utilizing programming. Video forensics examination, tells us about the accuracy of the input video. It has become a challenge for law enforcement agencies to deal with the increasing violence rate which involves the use of masks and weapons. The identification of a person becomes difficult with the use of face masks. The proposed method uses an efficient technique that is YOLO to detect guns, masks and suspicious persons from

Journal of Cyber Security and Mobility, Vol. 11_4, 549–574.
doi: 10.13052/jcsm2245-1439.1143

a video by extracting frames and features. It further compares the obtained frame with the available images in the dataset and generates output with bounding boxes detecting guns, masks and suspicious persons. This paper also examined the domain of video forensics and its outcomes. Experimental results show that the proposed method outperforms the existing techniques tested on different datasets. The precision for YOLO design for guns and masks is 100% and 75% respectively. The precision for customized CNN engineering for guns and face masks is 61.54% and 61.5% respectively. Execution measurements for both models have shown that the YOLO design outperformed the customized CNN with its presentation.

Keywords: Digital forensics, video forensics, tampering, soft computing techniques, YOLO, CNN, suspicious persons.

1 Introduction

The process of gathering and examining the information from past evidence is a part of forensic science. Forensic science is used in widespread domains such as cyber forensics, video forensics, digital forensics, etc. Video forensics includes analysis of video consisting of crime-related activities. With the extensive use of mobile devices and cameras these days, the evidence for crimes performed using masks and guns can be easily caught in a device [1]. Video forensics is the analysis of videos in legal matters using scientific methods of examination and evaluation. The biggest concern these days is regarding security in both crowded and lonely areas. For the detection of guns, masks and suspicious activity performed by a person, we use the soft computing techniques in video forensics, with which one can reduce the efforts and errors of manually doing this task. Video crime scene research covers the assessment disciplines of criminology and software program engineering and carries studies actions in the direction depending upon theories of a selected problem via the usage of computers and digital methods. Moreover, computer-based criminology wishes examinations and informatics to work collectively. Software engineering, groups up with forensic science in several angles. Computers' imagination and perception along crime scene research is a stable blend of facts restoration and examination [2, 3]. The more logical appraisal is available, and the genuine use of such strategies keeps on extending. Following troubles must be defeated in the video-based scientific examination. Recognizing a suspect in a crime scene through pictures or video film is troublesome because of the side face or weapon attributable

to inferior quality in the picture or video film, point of the camera, force in the picture or video film, etc. It would require the improvement of further developed ways to deal with the upgradation of the picture or video film quality [4]. Connections utilizing open archive assets like CCTV video, online picture accounts, or authentic path of these ancient rarities, among subjects in research circumstances, must be set up. As an end result of growing trends within the areas of interpersonal organizations, the Internet of Things (IoT), mobile phones, and so on, are for the law authorization applications, but contraptions take advantage of any open data [5]. Video forensic investigation finds significant proof from current starting points quickly utilizing specific practicality procedures using the ideas of profound learning and man-made brainpower. For criminological examination or for video film-based face or weapon detection, the detection of new faces includes maturing of the faces, markings, legal sketch detection, and close infrared face or weapon detection, powerful methods for evidence assortment like strong detection and the subject distinguishing proof, and so on [6]. It is used in several areas such as recording and securing the scene of crime, preservation and collection of pieces of evidence, identification of criminal or victim, etc. [7, 8].

Individual identity is turning into a developing need step by step in practically all areas. Video forensics would end up being productive for recognizable proof reason in legitimate issues, banks, and so forth. This strategy distinguishes an individual as well as tells us if the information introduced is genuine. Video legal sciences are tending to the worldwide level issues as ID and authentication of that report for individual or lawbreakers. It isn't just needed in a particular area yet it is a worldwide level issue which should be tended to. Strange behavior of any person whenever distinguished at a beginning phase would assist with taking care of a few issues in or before time. As cutting-edge wrongdoing augmentations dramatically, the necessity for video legal sciences criminological inclination in-law approval creates with it. The Indian crime scene investigation association is additionally taking endeavors to improve the productivity of the accessible video criminology strategies. Documentation of confirmation is huge in various pieces of a case, from basically recording the condition of evidence to overhauling nuances that may not be noticeable or evident to the common eye. Consequently, there is a lot of expansion for advancement in video lawful sciences [9]. With every certain edge, these devices have also become the eyewitnesses to essentially all events related to its customers (normal and criminal), so these are to be explored to extract the advantage and careful information that is associated and required [10].

Presently these days, there is an extraordinary change and headway in data innovation. Because of this development, the innovation is effectively open to the greater part of the people [11, 12]. The human dependence on this advancement has extended a ton. Essentially all the data is being taken care of electronically now. Therefore, these devices have become a vital piece of our regular day-to-day existence [13]. These devices have become eyewitnesses to all the events related to its customers, with every sure plot for normal and criminal and are used to look at and eliminate the advantage and exact information that is required. Electronic contraptions are investigated in all cases nowadays in metropolitan areas. PCs, Personal Digital Assistant (PDA), Compact Disk (CD), pen drive, Secure Digital (SD) and MicroSD cards, etc. are largely utilized advanced gadgets [14].

Because of the accessibility of minimal effort advanced and refined camcorders and the accessibility of video sharing sites, for example, YouTube, computerized recordings become the most significant part of day-to-day life [15]. Since recordings can be effortlessly controlled utilizing Available devices, their realness can't be underestimated. Altering a computerized video is not a simple errand, it is testing and tedious assignment in contrast with still pictures, and however video altering programming can be a simple method to control video. Only one out of every odd video fraud is similarly important; the messing with the film of a pop star may matter not exactly the change of film of wrongdoing in advancement [16]. Yet, the alterability of video subverts our good judgment suppositions about its precision and dependability as a portrayal of the real world. As advanced video altering strategies become increasingly modern, it is always important to create apparatuses for recognizing video phony [17, 18]. The video forensics terms such as video forgeries, watermarking and forensic system for video forgeries are explained as follows:

1.1 Video Forgeries

The film business is presumably the most grounded main impetus for the development of video control innovation. With the video altering innovation at present accessible, experts can without much of a stretch eliminate an article from a video arrangement, embed an item from an alternate video source, or even add an article made by PC designs programming. Surely, progressed video control innovation enormously advances our visual experience [19]. In any case, these strategies become progressively accessible to the overall

population, vindictive altering. Even though altering video is generally hard, as of late we have started to experience video fraud. Development in video altering is making a tremendous effect on our general public. Though right now a couple of computerized video falsifications have been uncovered, such occurrences are disintegrating the public trust in the video. In this manner, it is earnest for the scientific network to think of strategies for validating video accounts.

1.2 Watermarking

One answer for video validation is advanced watermarking. There are a few kinds of watermarks. Among them, delicate and semi-delicate watermarks can be utilized to confirm recordings. Delicate watermarking works by embeddings subtle data that will be changed if there is any endeavor to alter the video. Afterward, the installed data can be separated to confirm the genuineness of the video. The semi-delicate watermark works likewise. The thing that matters is that it is less touchy to old-style client adjustments, for example, pressure. The supposition will be that these alterations don't influence the trustworthiness of the video. The significant downside of the watermarking approach is that a watermark should be embedded accurately at the hour of recording.

1.3 Forensic System for Video Forgeries

The legal system is planned to recognize advanced falsifications without the assistance of watermarking (Digital validation), the major supposition behind our procedures is that altering a computerized video may upset certain basic properties of the video and these bothers can be demonstrated and assessed to identify altering. The video forensic system can be isolated into three modules: video analysis, video forensics and video steganalysis. Every procedure center around one explicit type of altering and can't be applied without any help to recognize all video phonies. The utilization of these modules in the blend, gives a promising start to identifying falsification in advanced recordings without watermarks.

The rest of the paper is organized as follows: Section 2 discusses the related work done in video forensics. The proposed methodology of video crime scene investigation is depicted in Section 3. The experimental results and discussion are shown in Section 4. The conclusion is concluded in Section 5.

2 Related Work

A lot of efforts have been put up by researchers for the video forensic system for the detection of various objects such as humans, abnormalities, masks, etc. Hou et al. [1] reviewed methodologies used for human tracking on camera networks. Fully Convolutional Neural (FCNs) organizations and worldly information are utilized and a pre-prepared managed FCN is moved into a solo FCN. This guarantees the recognition of (worldwide) abnormalities in scenes. Other than this, it is an answer for defeating restrictions in preparing tests utilized for learning a total Convolutional Neural Network (CNN). Bhaumik et al. [2] reviewed various soft computing approaches for content-based based video retrieval. The strategy empowers to run a profound learning-based technique at a speed of around 370 fps. Inside and out, the proposed strategy is both quick and exact for peculiarity discovery in video information.

Chen et al. [3] proposed a deep ranking framework that works straightforwardly on the picture pixels instead of hand-created highlights. It learns new portrayals for information in a solo way without the requirement for marks and then reproduces the information to perceive the areas of unusual occasions dependent on the remaking mistakes. All the more significantly, this methodology can be conveyed in both on the web and streaming settings. In this prepared boundaries of the model are driven inline settings while being refreshed steadily with video information showing up in a stream. The algorithm is tested on three benchmark video datasets which show that the proposed strategy can recognize and restrict the anomalies at pixel level with preferable exactness over those of baselines, and accomplish serious execution contrasted and cutting edge draws near. Their technique depends on restricted Boltzmann machines (RBMs) to catch information consistency. Henceforth, they can recognize and restrict sporadic occasions. The framework is prepared straightforwardly on the picture pixels in a solo way. For video real-time, they further present a streaming variant of their strategy that can gradually refresh the boundaries when new video outlines show up. Test results on a few benchmark datasets show that the proposed strategy beats average unaided baselines and accomplishes serious execution compared and best in class technology for peculiarity location. Ultimately, it is noticed that the proposed approach is planned and various RBMs are prepared to catch diverse picture measurements restricted at various locales. Nazare et al. [4] proposed a proficient strategy for identifying peculiarities in recordings. Spatiotemporal engineering for inconsistency recognition in

recordings including swarmed scenes has been proposed. The proposed engineering incorporates two principal parts, one for spatial component portrayal, and one for learning the fleeting advancement of the spatial highlights. They detailed peculiarity recognition as a spatiotemporal grouping exception discovery issue and applied a blend of the spatial component extractor and fleeting sequencer ConvLSTM to handle the issue. By consolidating convolutional method including extractor in both spatial and fleeting space into the encoding-interpreting structure, they fabricate a start to finish teachable model for video irregularity recognition. Notwithstanding the model's capacity to identify irregular occasions and its power to commotion, contingent upon the action's multifaceted nature in the scene, it might create all the more bogus alerts contrasted with different strategies.

Shao et al. [5] gave a profound learning-based oddity location, Defense Logistics Acquisition Directive (DLAD) framework is proposed to improve the acknowledgment issue in video preparing. The framework accomplishes the total location of irregular occasions by including the accompanying critical proposed modules a Background Estimation (BE) Module, an Object Segmentation (OS) Module, a Feature Extraction (FE) Module, and an Activity Recognition (AR) Module. From the start, creators have introduced a BE module that produced a precise foundation wherein a two-stage model is produced to process the foundation assessment. After a top-notch foundation is produced, the OS model is created to separate the article from recordings, and afterward, the object the following cycle is utilized to follow the item through the covering discovery conspire. From the followed objects, the FE module is removed for some valuable highlights, for example, shape, wavelet, and histogram to the anomalous occasion location. For the last advance, the augmented reality module is delegated anomalous or ordinary occasions utilizing the profound learning classifier. Tests are performed on the USCD anomaly detection dataset of unusual exercises, and correlations with the cutting-edge strategies approve the upsides of our calculation. The proposed action acknowledgment framework has been beaten by accomplishing a better EER of 0.75% when contrasted and the current frameworks (20%). Additionally, it shows that the proposed strategy accomplishes an 85% exactness rate in the edge level execution. Revathi et al. [6] investigated a novel unaided profound learning system to recognize irregular occasions in jam-packed scenes. In particular, low-level visual highlights, energy highlights, and movement map highlights are separated depending on spatiotemporal energy estimations. Three convolutional confined Boltzmann machines are prepared to display the mid-level element portrayal of typical examples. At that point,

a multimodal combination plot is used to get familiar with the profound portrayal of group designs. Given the learned profound portrayal, a one-class uphold vector machine model is utilized to identify inconsistent occasions. The proposed technique is assessed utilizing two accessible public datasets and contrasted and best-in-class strategies. The trial results show its serious exhibition for irregularity occasion discovery in video reconnaissance.

Sabokrou et al. [7] showed how object acknowledgment (Faster R-CNN) utilizes the technique to recognize objects' names and their relating area in the video scene as the initial step to execute oddity identification. At that point, the optical stream is utilized to distinguish versatile traffic streams in every part of the edge. Fundamentally, the creators propose an elective strategy for irregular action location utilizing a versatile abnormality recognition system. Additionally, the additional item acknowledgment module (Faster R-CNN) improves the genuine positive with the tradeoff of having inconsequential bogus positive. The proposed strategy has a limitation when there is a solid clamor in the information outlines as the movement conveyance map is assembled dependent on the movement speed of the moving items. Various preparing, tests, and trials have been left for the future because of the absence of time. Future works concern preparing Faster R-CNN over related oddity identification datasets. Jin et al. [8] proposed a profound learning way to deal with identifying certifiable oddities in observation recordings. Because of the intricacy of these practical peculiarities, utilizing just typical information alone may not be ideal for abnormality recognition. Thus the authors endeavored to abuse both ordinary and abnormal recordings. To stay away from work serious fleeting explanations of odd sections in preparing recordings, they gain proficiency with an overall model of peculiarity identification utilizing profound MIL structure with pitifully marked information. To approve the proposed approach, another huge scope irregularity dataset comprising of an assortment of certifiable peculiarities is presented. The test results on this dataset show that the proposed oddity location approach performs altogether in a way that is better than benchmark strategies. Moreover, they exhibit the handiness of the dataset for the undertaking of odd action acknowledgment.

Muhammad et al. [9] present a novel profound learning-based abnormality location approach (DeepAnT) for time arrangement information, which is similarly pertinent to the non-streaming cases. Deeping utilizes unlabeled information to catch and get familiar with the information dispersion that is utilized to conjecture the ordinary conduct of a period arrangement. This module takes a window of time arrangement (utilized as a specific circumstance) and endeavors to foresee the following time stamp. The anticipated

worth is then passed to the oddity locator module, which is liable for labeling the comparing time stamp as ordinary or unusual. The creators have played out a point-by-point assessment of 15 calculations on 10 irregularity location benchmarks, which contain an aggregate of 433 genuine and manufactured time arrangements. Tests show that DeepAnT beats the best-in-class abnormality recognition techniques in the majority of the cases, while performing at par with others. The comparative analysis of existing techniques on video forensics systems is shown in Table 1.

From the literature review, it is concluded that most of the authors have worked on images with techniques such as CNN, KNN, VGG-16, etc for detection of mask, fire, textual matter, etc. The detection of the mask was limited to prevailing Covid-19 conditions. As we are aware that crime is increasing day by day at an alarming rate and criminals generally make use of guns or other weapons. Hence, it becomes important to design a framework that can detect these guns from different videos and help the officials to detect them either at an earlier stage or for the detection of criminals after the crime has been committed. Working on videos is still a challenging task.

3 Proposed Methodology

The proposed method is based on the fact that the detection of masks, guns and suspicious persons is done using an efficient technique called YOLO. This section discusses the customized neural network that is utilized for the present study, and it also, discusses the architecture of YOLO for object detection. We will comprehensively inspect and research significant learning structures for early inconsistency areas in accounts. From the composing review and the writing study, it is seen that both the inception and residual associations have demonstrated commonly magnificent execution with a modestly low computational cost. The general block diagram of the proposed methodology for the video forensic system is shown in Figure 1.

(a) Input Video: – The proposed system requires input video for the detection of masks, guns and anomalous behavior. This video can be in the form of already available video, captured through CCTV, camera, mobile phones, etc. It can also work conveniently with real-time videos.
(b) Frame extraction: – Outline extraction is a helpful resource that completes video content by picking a lot of framework key edges to address video groupings. By far most of the current key edges' extraction strategies are not suitable for video copyright affirmation, as they don't meet

Table 1 Comparison of existing techniques on video forensics system

Author (Year)	Techniques	Performance/ Accuracy/Error	Conclusion
Hou et al. [20] (2014)	Tracking with the deep track, CUDA-PTX, Convolutional Neural Networks	Performance gap 10% with state-of-the-art	It employs CNN architecture and a structural loss function that handles multiple input cues and class-specific tracking.
Bhaumik et al. [21] (2016)	Hybrid soft computing techniques	–	The systems have increased in robustness, efficiency and effectiveness as compared to earlier used traditional approaches. It also helped to reduce user interaction and manual annotation to a great extent.
Chen et al. [22] (2016)	Deep CNN	–	Formulate the person re-identification task as a learning-to-rank problem. Extensive experimental results demonstrate the effectiveness of our proposed approach
Nazare et al. [23] (2016)	Smart Surveillance Framework (SSF) Framework 1. user modules 2. SSF kernel	The performance gain of 17%.	Smart Surveillance Framework (SSF) allows the simultaneous execution of multiple user modules that can be developed independently since they have communication and synchronization through a shared memory, which contributes to scalability and flexibility.
Shao et al. [24] (2016)	Big Data – LDA, KNN, fuzzy clustering	–	By a combination of snapshot images, original surveillance videos and unusual events, valuable clues can be found out much easier, which thus helps the police boost their investigation efficiency.

(*Continued*)

Table 1 Continued

Author (Year)	Techniques	Performance/ Accuracy/Error	Conclusion
Revathi et al. [25] (2017)	Deep learning-based anomy detection (DLAD), Background Estimation (BE) Module, an Object Segmentation (OS) Module, a Feature Extraction (FE)Module, and an Activity Recognition (AR) Module	85% precision rate, EER of 0.75%	OR module is categorized as an abnormal or normal event using a deep learning classifier
Sabokroua et al. [26] (2017)	Deep learning – Fully Convolutional neural network (FCN), AlexNet	EER is 10%, where the best result, in general, is 11%	The proposed method is both fast and accurate for anomaly detection in video.
Jin et al. [27] (2017)	ConvLSTM, Spatiotemporal architecture	EER is 9.5% while the state-of-the-art has 9.9%	Detects Abnormal events but it may produce more false alarms as compared to other events
Muhammad et al. [28] (2018)	CNN, GoogleNet	For Dataset1 accuracy of 85% with 11.67% false alarms. For Dataset2: better than hand-crafted features	This paper improved the flame detection accuracy, but the number of false alarms is still high and more research is required.
Bajestani et al. [29] (2018)	Faster R-CNN, Object Detection	True Positive = 0:56, False Positive = 0:36	This method improves the true positive with the tradeoff of trivial false positive.
Kaushal et al. [30] (2018)	Neural network Fuzzy Logic Neuro-fuzzy Hybrid	–	Various soft computing-based approaches for moving object detection and tracking in videos. The article provides various techniques along with scope, pros, cons and the limitation associated with each of them

(*Continued*)

Table 1 Continued

Author (Year)	Techniques	Performance/ Accuracy/Error	Conclusion
Huang et al. [31] (2018)	Restricted Boltzmann machine (RBM)	–	It improves the average accuracy of multimodal deep representation by 2.65%
Munir et al. [32] (2018)	DeepAnT	Better precision for all the datasets.	Evaluation of DeepAnT on 10 different data sets comprising of 433 time series in total and provided a detailed comparison with 15 state-of-the-art anomaly detection methods.
Singh et al. [33] (2018)	Passive-blind technique Interframe forgery detection	Detection accuracy was 97%, 96.1%, and 93.3% at 9, 6, and 3 Mbps, for MPEG compression.	Presents a repository of information regarding the kinds of tamper attacks a video can suffer from and a comprehensive source of references for the passive-blind techniques proposed for detecting attacks
Sultani et al.[34] (2018)	Deep MIL Ranking Model, binary SVM classifier	False alarm: 1.9%	A new large-scale anomaly dataset consisting of a variety of real-world anomalies is introduced.
Castillo et al. [35] (2018)	DaCoLT – a brightness guided preprocessing approach	Precision of around 73%	This paper proposed a brightness-guided preprocessing approach. The DaCoLT model shows a high potential even in low-quality videos and provides satisfactory results as an automatic alarm system.
Bouindour et al. [36] (2019)	Matlab, Convolutional Neural Networks	EERFL and EERPL of 6.25% and 9.82% respectively	This method is robust, takes into account rare normal events present in the training phase. Besides, it can be incorporated into online CCTV.

(*Continued*)

Table 1 Continued

Author (Year)	Techniques	Performance/ Accuracy/Error	Conclusion
Camerona et al. [37] (2019)	CNN based algorithms GoogLeNet, AlexNet VGG16, ResNet50, SSD	Pre CNN approaches 89% on their custom dataset. Faster R-CNN – Very high YOLO – Low SSD – Medium ResNet – High	This work has highlighted some of the practical challenges of designing a processing system for a CNN-based automated surveillance system using off-the-shelf hardware and open-source algorithms.
Xiao et al. [38] (2019)	Deep learning	–	Proposed a framework for a video-based digital forensics investigation, useful for anti-crime or fast response when crime activities or behaviors are detected
Saleema et al. [39] (2019)	Deep Convolution Neural Networks (CNN), MobileNetV2	EER-0.35 EER-0.52	It is a framework for generating multi-line textual descriptions for video captioning. Feats-rich model extends feature matrix to visual (2-D and 3-D) and facial features. Spatio-temporal characteristics are encompassed by employing deep neural networks.
Sreenu et al. [40] (2019)	Big Data – ImageNet2012, PASCAL VOC, Frames Labeled In Cinema (FLIC), Leeds Sports Pose (LSP)	–	Methods analyzing crowd behavior were discussed.

unequivocal necessities. Keyframe extraction method in a deep neural network is ahead pre-taking care of step in video examination. In this paper, the keyframe extraction method is used to extract frames from the input video. Firstly, an alternative keyframe sequence is obtained which is based on the color traits of the original difference between video frames; then keyframe sequence is achieved according to the

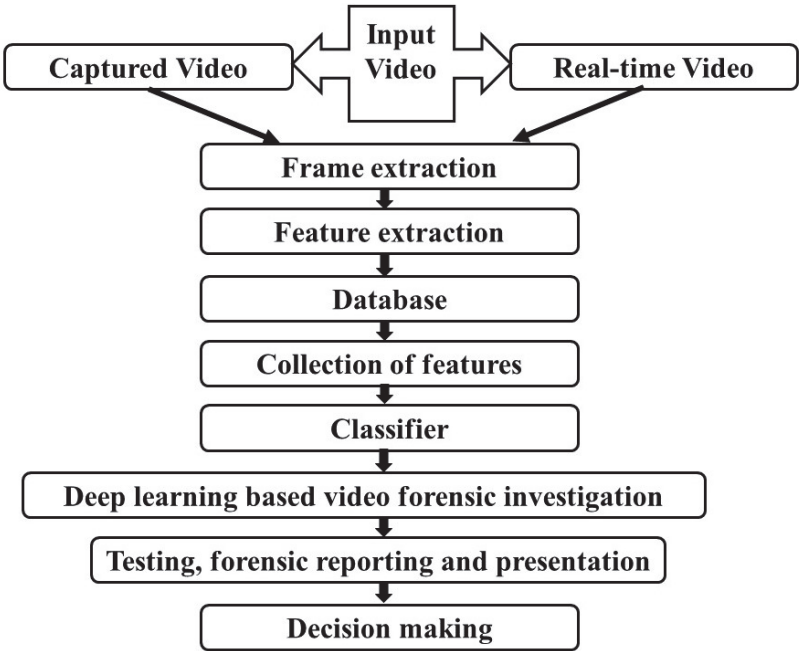

Figure 1 General block diagram of the proposed methodology for video forensic system.

structure characteristic differences between the alternative keyframes sequence, and finally it resolute the number of keyframes to ensure the effectiveness of keyframes. The inspiration driving isolating keyframes is to get extra isolated information from the video in a suitable manner. Each video has its clever traits, for instance, drenching, splendor, contrast, camera point, vibration, dark, region of the movement, number of performers, kind of action, length, and establishment. Considering a tremendous number of elements in each video and treating all accounts likewise accomplish a huge deficiency in keyframe extraction. Therefore, it is critical to see the region of action in constant action video. The area of movement in a film is associated with the point on the screen and camera, to which the examiner's thought is paid, when in doubt. It is seen that thought is paid to the central district generally while recording and watching. Thusly, the outer areas of video diagrams are regularly altered off preceding perceiving the area of interest. Then, the area of movement is shaped as a region in the point of convergence of video diagrams, which makes either the best differentiation or least

likeness between consecutive blueprints, inciting an arrangement for the video to follow the action district. Figuring simply the qualification in regions of action between diagrams generally through the video helps with removing keyframes even more exactly and truly by diminishing the effect of possibly effectively advancing establishments.

(c) Feature extraction: – The Principal Component Analysis (PCA) is used by the proposed method for feature extraction. PCA is a procedure for getting critical elements (in sort of parts) from a tremendous game plan of variables open in an instructive list. PCA is more useful while overseeing 3 or higher-layered data. PCA can be used for inconsistency disclosure and exemption distinguishing proof since they won't be fundamental for the data as it would be seen as upheaval by PCA. Feature extraction intends to diminish the number of features in a dataset by making new features from the current ones.

(d) Database: – After frame extraction and feature extraction a complete database is ready to train and test the model for detection of masks, guns and suspicious persons. This database is used in training new models and/or testing the already trained available models.

(e) Collection of features: – From the extracted database, the extracted features are collected for further use. These features can be used for the detection of various objects like masks etc. using soft computing techniques.

(f) Classifier: – A classifier is a unique instance of a theory. A classifier is a theory or discrete-esteemed capacity that is utilized to allocate class names to specific items. For example, in the email order model, this classifier could be a theory for naming messages as spam or non-spam. In any case, speculation should not be interchangeable with a classifier. Along these lines, we can say that a classifier is an exceptional instance of a theory or model: a classifier is a capacity that assigns a class name to an element. The proposed method uses a deep learning classifier for training and building a database for further identification.

(g) Deep learning enabled video forensic investigation: – Like never before previously, the world in these days encountering expanded digital assaults in every aspect of our day-to-day routines. The present circumstance has made battling cybercrimes a day-by-day battle for the two people and associations. Moreover, this battle has been irritated by the way that the present cybercriminals have felt free to utilize confounded digital assault strategies. A portion of the methods are tiny and subtle and regularly cover in the veneer of true demands and orders.

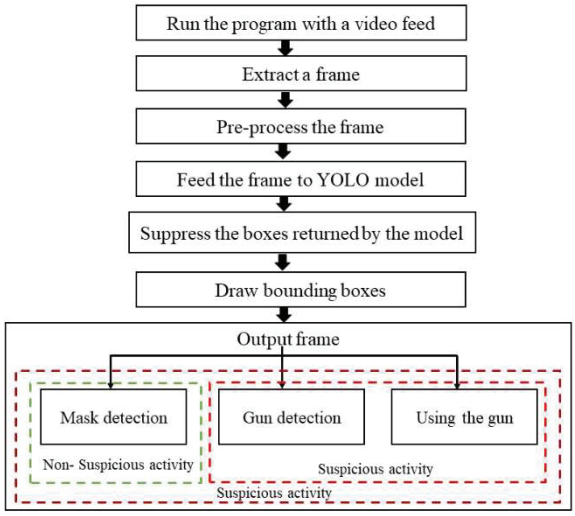

Figure 2 Flowchart of the proposed methodology.

The proposed method trains data using deep learning, by giving input video for training the system for the detection of masks, guns and suspicious persons. Further, it utilizes the same database for identifying the masks, guns and suspicious persons from the input videos.

(h) Testing, Forensic reporting and presentation: – To battle this hazard, particularly after a security episode has occurred, network safety experts, as well as advanced criminological agents, are constrained all of the time to filter through enormous and complex pools of information otherwise called profound figuring out how to divulge Potential Digital Evidence (PDE) that can be utilized to help suits.

(i) Decision Making: – This is the final step involved in the process of analyzing a video in forensics investigation. The final result may then be presented to the law authorities for proving its authentication.

The flowchart of the proposed methodology for the video forensics system is shown in Figure 2.

It predominantly managed the calculation for the proposed technique for recognizing objects like firearms and veils. The proposed calculation can be addressed. The steps involved in the proposed methodology are enlisted as follows:

(a) Run the program with a video feed: Gather images containing persons wearing a mask and without a mask and carrying a gun from the fed

video. Then label image with mask, no mask and gun for YOLO using LABELIMG software and Setup Google Colab with GPU and install TensorFlow GPU.

(b) Extract a frame: Extract the frame using the keyframe extraction method. Download Darknet framework to train object detection model.

(c) Preprocess the frame: Make some changes in the Darknet's "Makefile" to run on GPU.

(d) Feed the frame to YOLO Model: The frame is then fed to the YOLO model. Change yolov3-tiny-training.cfg configuration file to run efficiently for three classes i.e. masks, guns and suspicious activity. The YOLO model has been trained with the parameters mentioned in Table 2 below:

Table 2 Parameters used for training YOLO model

Changed batches	64
Subdivision	16
Maxbatches	6000
Filters	(classes+5) *3 = 24
Classes	3

The Block diagram of the proposed methodology for the detection of guns, masks and suspicious persons is shown in Figure 3 below.

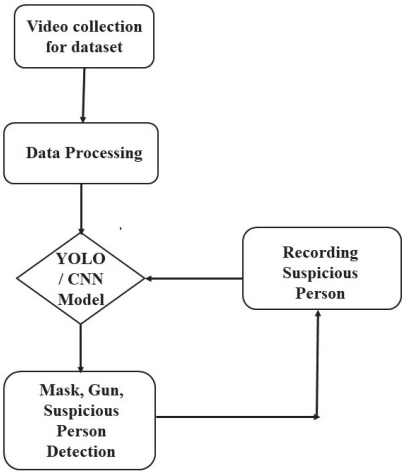

Figure 3 Block diagram of the proposed methodology for detection of guns, masks and suspicious person.

The current methodology, YOLO, is built and taken a stab at PASCAL VOC affirmation datasets as a CNN [41]. There are 24 layers of convolution, followed by two associated layers. By its inspiration, the layers are separated as follows:

- Beginning 20 layers of convolution joined by an ordinary pooling layer and an associated layer is pre-arranged on the popular dataset, ImageNet, with a 1000-class grouping dataset.
- Instructive assortment with objective 224 $\times$ 224 is used for pre-getting ready for request.
- The layers are involved in 1 $\times$ 1 and 3 $\times$ 3 layers of reduction and convolution independently.
- The association is ready for object affirmation in the last four convolutional layers joined by two associated layers.
- Object affirmation needs extra granular information with the objective that the enlightening assortment objective is extended to 448 $\times$ 448.
- The last layer measures class and cutoff points probability.
- The last layer is established through a straight limit, while the past convolutional layers are started through broken ReLU work.
- The data picture is of size 448 $\times$ 448, and the outcome is gauge in the hopping box.

(e) Suppress the boxes returned by the model: Create an obj.names file containing the names of the classes. Create obj. data files contain the path to obj.names, training file, validation file and path to back up the model(in this case it is google drive)

(f) Draw bounding boxes: Draw the bounding boxes with the labels as 'M' for the mask, 'NM' for no mask, 'G' for the gun. Copy and paste the dataset with labels in the darknet/data/obj folder. Create a train.txt file in the darknet/data folder containing the path to all the images of the dataset.

(g) Output frame: Download default and pretrained weights of tiny-yolov3. Start training the model with darknet. Whenever depleted, stop further handling.

4 Experimental Results and Discussion

The current segment for the most part manages the conversation of the dataset, equipment, and programming subtleties used to foster the present proposed methodology. The proposed method is implemented using Personal

Computer (PC), whose working framework is Windows10, the processor is Intel® CoreTM i7-9750H with a base recurrence 2.60GHz and a maximum super recurrence 4.50GHz, and GPU is NVIDIA GeForce RTX 2060 of 6 GB. The dataset used for the proposed methodology is gathered from different open sources and self-produced from different datasets like Github, Kaggle, and the information world. The considered information has a size of 1346 pictures identified with the firearm. Also, the information gathered for covers has been gathered from different open sources like Github and Kaggle. The considered mask dataset has a size of 1043 pictures.

The present proposed work can be arranged into two segments. The primary segment manages the ID of weapons, and the subsequent segment manages the recognizable proof of veils. Both these segments are carried out utilizing customized CNN just as the YOLO organization. In this set of video criminology, object recognition can be characterized into two classifications like suspicious and non-suspicious activity. On account of the identification of mask or gun or both considered suspicious, in any other case, the movement is recognized as non-suspicious action. Suspicious movement may fluctuate in the level of doubt, yet that isn't the mark of conversation in this unique situation. The recognized gun and mask pictures are demonstrated like 'M', 'NM', and 'G' designate 'Masked', 'Not Masked', and 'having Gun' individually. Sample images from the gun dataset and mask dataset are shown in Figures 4 and 5 respectively. The output results obtained from analyzing various videos for identification of masks, identification of guns and identification of a person from abnormal and suspicious activity video are shown in Figures 6, 7 and 8 respectively.

Figure 4 Sample images from the gun dataset.

Figure 5 Sample images from the mask dataset.

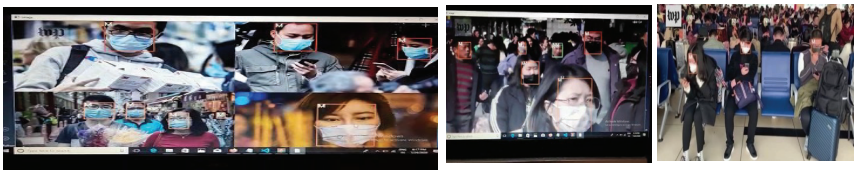

Figure 6 The output images for identification of masks.

Figure 7 The output images of masks and guns identification.

Figure 8 Identification of person from abnormal and suspicious activity video.

The comparative analysis of the proposed method with previously reported video forensics techniques for analysis of video forensics system for detection of gun, mask and anomaly using soft computing techniques are shown in Table 3. The gun dataset involves different types of guns from various angles. Comparison of the three techniques YOLO, CNN and VGG-16 used for the facial mask identification framework and gun detection framework, brought to the inference that:

As can be seen in Table 3, the accuracy of our method is higher than Cameron et al. [18] and Xiao et al methods. [19]. Cameron et al. [18] suggested a VGG-16 algorithm for an automated surveillance system. This method is unable to curtail the processing time curve to maximize the usage of hardware. Xiao et al. [20] recommended an effective technique CNN to improve the quality of closed-circuit television (CCTV) footage for investigation in video forensics. This method lacks detection of the knife in low-quality videos. The accuracy of the proposed method is higher as compared to other existing techniques as it can detect masks and guns in the video with an accuracy of 92.3% and 100% respectively. Thus it is

Table 3 Comparison of the proposed methodology with existing techniques for detection of mask and gun

Author (Year)	Techniques	Datasets	Accuracy (%) (Mask Detection)	Accuracy (%) (Gun Detection)
Cameron et al. [18] (2019)	VGG-16	PASCAL 2012 VOC	89%	95.4%
Xiao et al. [19] (2019)	CNN	Own dataset of CCTV footage	85%	95%
Proposed methodology	Customized CNN	Own dataset with images from Kaggle, Github datasets	61.54%	61.5%
	YOLO		92.3%	100%

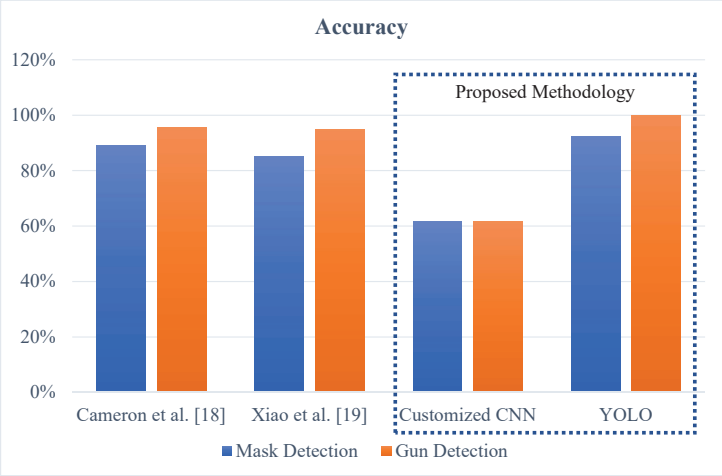

Figure 9 Graphical representation of comparison of the proposed methodology with existing techniques for detection of mask and gun.

concluded that the proposed method shows higher accuracy in comparison to the previously reported methods.

From the above outcomes shown in Figure 9, one can comprehend that the customized CNN isn't contrasted with the YOLO design. It shows that the YOLO design performs better compared to some other altered CNN engineering. YOLO gives an accuracy of 92.31% for mask detection, 100% for gun detection and precision of 75% and 100% for mask and gun detection

respectively. As it has appeared in the past areas, video forensics sciences is currently a hot exploration issue in the sign preparing world opening new issues and examination strings. Notwithstanding a few methods that have been mutated from picture crime scene investigation, video signals present new difficulties in the legal application world because of the sum and the unpredictability of information to be prepared and the wide work of pressure strategies, which may change or eradicate effects had by past sign adjustments.

5 Conclusion

In this paper, a framework is proposed in this article as a part of video forensics for the detection of masks, guns and suspicious persons using soft computing techniques. The proposed framework has three segments: identification of guns, identification of masks and the suspicious person using YOLO architecture. The YOLO architecture strongly surpassed the customized CNN architecture and VGG-16 architecture due to non-initiating the weights randomly, and a pre-training section of the YOLO architecture. YOLO architecture supports the user to identify various objects in the image, in this scenario guns as well as masks. At the same time, that becomes one more advantage to this architecture. Identification of suspicious activity done by the proposed methodology makes it all the more different from existing techniques. The performance of the YOLO architecture can be improved by using a larger dataset in the future.

References

[1] Park, S., Yu, S., Kim, M., Park, K., and Paik, J. (2018). "Dual autoencoder network for retinex-based low-light image enhancement". IEEE Access, 6, 22084–22093.

[2] Fan, W., Wang, K., Cayre, F., and Xiong, Z. (2015). "Median filtered image quality enhancement and anti-forensics via variational deconvolution". IEEE transactions on information forensics and security, 10(5), 1076–1091.

[3] Mandal, S., Deán-Ben, X. L., and Razansky, D. (2016). "Visual quality enhancement in optoacoustic tomography using active contour segmentation priors". IEEE transactions on medical imaging, 35(10), 2209–2217.

[4] Walker, H., and Tough, A. (2015). "Facial Comparison from CCTV footage: The competence and confidence of the jury". Science & Justice, 55(6), 487–498.

[5] Verolme, E., and Mieremet, A. (2017). "Application of forensic image analysis in accident investigations". Forensic science international, 278, 137–147.

[6] Seckiner, D., Mallett, X., Roux, C., Meuwly, D., and Maynard, P. (2018). "Forensic image analysis – CCTV distortion and artifacts". Forensic science international, 285, 77-85.

[7] Li, S., Choo, K. K. R., Sun, Q., Buchanan, W. J., and Cao, J. (2019). "IoT forensics: Amazon echoes as a use case". IEEE Internet of Things Journal, 6(4), 6487–6497.

[8] Jerian, M., Paolino, S., Cervelli, F., Carrato, S., Mattei, A., and Garofano, L. (2007). "A forensic image processing environment for the investigation of surveillance video". Forensic science international, 167(2–3), 207–212.

[9] Nievas, E. B., Suarez, O. D., García, G. B., and Sukthankar, R. (2011). "Violence detection in video using computer vision techniques". In International Conference on Computer Analysis of images and patterns (pp. 332–339).

[10] Gowsikhaa, D., and Abirami, S. (2012). "Suspicious Human Activity Detection from Surveillance Videos". International Journal on Internet & Distributed Computing Systems, 2(2).

[11] Lone, A. H., Badroo, F. A., Chudhary, K. R., and Khalique, A. (2015). "Implementation of forensic analysis procedures for WhatsApp and Viber android applications". International Journal of Computer Applications, 128(12), 26–33.

[12] Kamenicky, J., Bartos, M., Flusser, J., Mahdian, B., Kotera, J., Novozamsky, A., ...and Horinek, J. (2016). "PIZZARO: Forensic analysis and restoration of image and video data". Forensic science international, 264, 153–166.

[13] Senan, M. F. E. M., Abdullah, S. N. H. S., Kharudin, W. M., and Saupi, N. A. M. (2017). "CCTV quality assessment for forensics facial recognition analysis". In 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence (pp. 649–655).

[14] Kaur, H., and Choudhary, K. R. (2017). "Digital forensics: implementation and analysis for google android framework". In Information Fusion for Cyber-Security Analytics (pp. 307–331).

[15] Singh, G., and Singh, K. (2019). "Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation". Multimedia Tools and Applications, 78(9), 11527–11562.

[16] Ramzan, M., Abid, A., Khan, H. U., Awan, S. M., Ismail, A., Ahmed, M., . . . and Mahmood, A. (2019). "A review on state-of-the-art violence detection techniques". IEEE Access, 7, 107560–107575.

[17] Horváth, J., Güera, D., Yarlagadda, S. K., Bestagini, P., Zhu, F. M., Tubaro, S., and Delp, E. J. (2019). "Anomaly-based manipulation detection in satellite images". networks, 29, 21.

[18] Peixoto, B., Lavi, B., Martin, J. P. P., Avila, S., Dias, Z., and Rocha, A. (2019). "Toward subjective violence detection in videos". In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8276–8280).

[19] Khamparia, A., Pande, S., Gupta, D., Khanna, A. and Sangaiah, A.K. (2020), "Multi-level framework for anomaly detection in social networking", Library Hi Tech, Vol. 38 No. 2, pp. 350–366. https://doi.org/10.1108/LHT-01-2019-0023

[20] L. Hou, W. Wan, J.-N. Hwang, R. Muhammad, M. Yang, and K. Han, "Human tracking over camera networks: a review," *EURASIP J. Adv. Signal Process.*, vol. 2017, no. 1, p. 43, Dec. 2017, DOI: 10.1186/s13634-017-0482-z.

[21] H. Bhaumik, S. Bhattacharyya, M. D. Nath, and S. Chakraborty, "Hybrid soft computing approaches to content based video retrieval: A brief review," *Appl. Soft Comput.*, vol. 46, pp. 1008–1029, Sep. 2016, DOI: 10.1016/j.asoc.2016.03.022.

[22] S.-Z. Chen, C.-C. Guo, and J.-H. Lai, "Deep Ranking for Person Re-Identification via Joint Representation Learning," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2353–2367, May 2016, DOI: 10.1109/TIP.2016.2545929.

[23] A. C. Nazare Jr. and W. R. Schwartz, "A scalable and flexible framework for smart video surveillance," *Comput. Vis. Image Underst.*, vol. 144, pp. 258–275, Mar. 2016, DOI: 10.1016/j.cviu.2015.10.014.

[24] Z. Shao, J. Cai, and Z. Wang, "Smart Monitoring Cameras Driven Intelligent Processing to Big Surveillance Video Data," *IEEE Trans. Big Data*, vol. 4, no. 1, pp. 105–116, Mar. 2018, DOI: 10.1109/TBDATA.2017.2715815.

[25] A. R. Revathi and D. Kumar, "An efficient system for anomaly detection using deep learning classifier," *Signal Image Video Process.*, vol. 11, no. 2, pp. 291–299, Feb. 2017, DOI: 10.1007/s11760-016-0935-0.

[26] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayedd, and R. Klette, "Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes," Sep. 2016, Accessed: Feb. 22, 2022. [Online]. Available: https://arxiv.org/abs/1609.00866v2.

[27] C.-B. Jin, S. Li, and H. Kim, "Real-Time Action Detection in Video Surveillance using Sub-Action Descriptor with Multi-CNN," Oct. 2017, Accessed: Feb. 22, 2022. [Online]. Available: https://arxiv.org/abs/1710.03383v1.

[28] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, and S. W. Baik, "Convolutional Neural Networks Based Fire Detection in Surveillance Videos," *IEEE Access*, vol. 6, pp. 18174–18183, 2018, DOI: 10.1109/ACCESS.2018.2812835.

[29] M. Farhadi Bajestani, S. Abadi, S. Fard, and R. Khodadadeh, *AAD: Adaptive Anomaly Detection through traffic surveillance videos*. 2018.

[30] M. Kaushal, B. S. Khehra, and A. Sharma, "Soft Computing based object detection and tracking approaches: State-of-the-Art survey," *Appl. Soft Comput.*, vol. 70, pp. 423–464, Sep. 2018, DOI: 10.1016/j.asoc.2018.05.023.

[31] S. Huang, D. Huang, and X. Zhou, "Learning Multimodal Deep Representations for Crowd Anomaly Event Detection," *Math. Probl. Eng.*, vol. 2018, pp. 1–13, 2018, DOI: 10.1155/2018/6323942.

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

[33] R. D. Singh and N. Aggarwal, "Video content authentication techniques: a comprehensive survey," *Multimed. Syst.*, vol. 24, no. 2, pp. 211–240, Mar. 2018, DOI: 10.1007/s00530-017-0538-9.

[34] W. Sultani, C. Chen, and M. Shah, "Real-World Anomaly Detection in Surveillance Videos," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, pp. 6479–6488. DOI: 10.1109/CVPR.2018.00678.

[35] A. Castillo, S. Tabik, F. Pérez, R. Olmos, and F. Herrera, "Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning," *Neurocomputing*, vol. 330, pp. 151–161, Feb. 2019, DOI: 10.1016/j.neucom.2018.10.076.

[36] S. Bouindour, H. Snoussi, M. Hittawe, N. Tazi, and T. Wang, "An On-Line and Adaptive Method for Detecting Abnormal Events in Videos

Using Spatio-Temporal ConvNet," *Appl. Sci.*, vol. 9, no. 4, p. 757, Feb. 2019, DOI: 10.3390/app9040757.

[37] J. A. D. Cameron, P. Savoie, M. E. Kaye, and E. J. Scheme, "Design considerations for the processing system of a CNN-based automated surveillance system," *Expert Syst. Appl.*, vol. 136, pp. 105–114, Dec. 2019, DOI: 10.1016/j.eswa.2019.06.037.

[38] J. Xiao, S. Li, and Q. Xu, "Video-Based Evidence Analysis and Extraction in Digital Forensic Investigation," *IEEE Access*, vol. 7, pp. 55432–55442, 2019, DOI: 10.1109/ACCESS.2019.2913648.

[39] S. Saleem, A. Dilawari, U. G. Khan, R. Iqbal, S. Wan, and T. Umer, "Stateful human-centered visual captioning system to aid video surveillance," *Comput. Electr. Eng.*, vol. 78, pp. 108–119, Sep. 2019, DOI: 10.1016/j.compeleceng.2019.07.009.

[40] G. Sreenu and M. A. Saleem Durai, "Intelligent video surveillance: a review through deep learning techniques for crowd analysis," *J. Big Data*, vol. 6, no. 1, p. 48, Dec. 2019, DOI: 10.1186/s40537-019-0212-5.

[41] S. K. Nanda, D. Ghai, and S. Pande, "VGG-16-Based Framework for Identification of Facemask Using Video Forensics," in *Proceedings of Data Analytics and Management*, vol. 91, D. Gupta, Z. Polkowski, A. Khanna, S. Bhattacharyya, and O. Castillo, Eds. Singapore: Springer Singapore, 2022, pp. 673–685. DOI: 10.1007/978-981-16-6285-0_54.