Harnessing Deep Learning Models for Enhanced Violence Recognition in Modern Surveillance Systems

Anand[1], Prof. Vikas Kamle[2]

[1]Research Scholar, [2]Assistant professor
[1,2]Department of Computer science & engineering
[1,2]Patel College of Science & Technology, Indore, India

Abstract
In the realm of video content analysis, distinguishing between 'Violence' and 'NonViolence' poses a significant challenge due to the dynamic and complex nature of video data. Traditional models like ResNet50 and MobileNetV2 have shown substantial capabilities in image classification tasks but struggle with the temporal aspects inherent in video. To address this, we introduce a hybrid deep learning approach that combines the spatial feature extraction strengths of InceptionV3 with the temporal pattern recognition prowess of LSTM networks. Our proposed method involves rigorous preprocessing of data, including noise reduction and feature extraction, followed by a systematic model training process. The performance evaluation of our approach has revealed a remarkable accuracy of 99.86% and a validation accuracy of 92.48%, outperforming the other evaluated models by a significant margin. These results not only validate the efficacy of the hybrid model in video classification tasks but also suggest its potential for broader applications in real-world scenarios that require nuanced content discernment.

Keywords: Violence Recognition, Deep Learning Models, Automated Surveillance, Real-Time Detection, Convolutional Neural Networks (CNNs)

Introduction

Automatic video classification is a rapidly expanding topic in artificial intelligence and machine learning, driven by its significant influence on security, entertainment, and social media. Deep learning has enabled the extraction and acquisition of intricate features from unprocessed data, exceeding earlier methods in computer vision. The fundamental aspect of video content classification is the categorization of video components into pre-established classes. Nevertheless, this strategy is challenging. Videos are characterized by their high-dimensional nature, including both spatial and temporal data. Spatial content encompasses entities, circumstances, and individuals contained inside frames. Videos record the motion and actions that occur over a period of time, giving a sense of the surrounding circumstances. Conventional machine learning techniques were unsuccessful in this aspect, since they required manually crafted features and extensive domain knowledge to get satisfactory outcomes. Convolutional Neural Networks revolutionized the field. These networks possess the ability to autonomously acquire spatial feature hierarchies from picture data, making them very effective in classifying images. Convolutional Neural Networks (CNNs) such as ResNet50 and MobileNetV2 have exceptional performance in analyzing still images, but they face challenges when it comes to classifying videos. The primary cause is their neglect of temporal dependencies, namely the changes that occur from frame to frame. The level of complexity or challenge.

The complexity of human behaviors, such as violent and non-violent contacts, leads to an increase in difficulty when identifying them. The model must accurately analyze each individual frame and comprehend the sequence of frames in order to comprehend complex behaviors. Identifying basic behaviors is straightforward, but automated systems have difficulties due to contextual ambiguity. Long Short-Term Memory networks (LSTMs) are a kind of recurrent neural networks (RNNs) that are designed to identify patterns in data streams. LSTMs are effective for sequential prediction due to their ability to retain and use prior data. LSTMs have the ability to analyze the temporal features of videos in order to understand how the material evolves over time. Due to their inability to process high-dimensional spatial data, LSTMs are not well-suited for video classification. As a result, there have been the development of hybrid models that integrate CNN feature extraction with LSTM sequence modeling. These models use Convolutional Neural Networks (CNNs) to extract comprehensive data at the frame level, and Long Short-Term Memory (LSTM) networks to analyze the sequence of this data in order to understand video content. The hybrid strategy we use combines the use of InceptionV3, a robust convolutional neural network renowned for its ability to identify images, with LSTM networks to effectively capture temporal dynamics. This model utilizes the InceptionV3 architecture to analyze video frames and extract features like as shapes, patterns, and objects. Over time, LSTM layers acquire comprehension of the video's event sequence based on the spatial context of these properties. This approach requires meticulous data preparation. The preparation step involves cleaning video frames, reducing noise, and normalizing data to satisfy the requirements of the model. The model trains on labeled examples after preprocessing. Labelling is essential for the model to get accurate categorization information.

Efficiently training the model's complex architecture requires the optimization of computer resources. To do this, one must carefully choose appropriate hyperparameters, design an effective network architecture, and ensure that the training data accurately reflects the diverse and complicated nature of the model's real-world applications. Metrics evaluate the precision and ability of the model to generalize once it has been trained. Loss measures quantify the degree of alignment between the model's predictions and the actual labels, whereas accuracy metrics indicate the fraction of correct predictions. These datasets are used for training and validation, with the validation dataset serving as a proxy for evaluating the model's performance on unseen data. In our tests, our hybrid model demonstrated superior performance compared to CNNs and LSTMs. The InceptionV3 and LSTM combination has shown exceptional training accuracy and impressive performance on the validation set, indicating that it has successfully learned the fundamental patterns without suffering from overfitting to the training data.

Hybrid deep learning models enhance the process of categorizing video footage. By using Long Short-Term Memory (LSTM) networks, we may overcome the limitations of Convolutional Neural Networks (CNNs) in analyzing temporal information. This enables us to develop systems that effectively understand video data. These models have the potential to enhance automated video analysis, allowing for more sophisticated, efficient, and precise classification of information across many industries and social platforms.

Review Of Literature

We conducted a content analysis specifically targeting online violent entertainment. This research does a comparison between amateur and professional material using a dataset of 2,520 videos from YouTube. The films are categorized based on their popularity, user ratings, and randomly selected samples. The violent frequencies and context of these YouTube video categories were compared to research on television violence, taking into account many factors such as the attributes of the offender and victim, the logic for the violence, and its effects. YouTube is far more secure than television. In contrast to the depiction of violence on television, this occurrence resulted in tangible real-life repercussions and took place in a more unsettling environment. Post hoc comparisons revealed that various video producers and genres depicted violent material in distinct ways.

Given the substantial rise in babies and toddlers' utilization of YouTube, it is essential to scrutinize the content they consume on the site. The YouTube channel indexes were developed by parents and instructors, making it difficult to discern the content that toddlers and babies are seeing. The research focused on analyzing the cognitive, emotional, and social development of newborns using YouTube videos. The study analyzed the cognitive, emotional, and social aspects individually to assess language impairment, verbal and physical aggression, emotional expression, empathy, emotional regulation, representation of prosocial and antisocial behavior, and prosocial communication. As a result, there was a low score for physical violence and a high score for mental distress. Instead of engaging in harmful language, verbal aggression, or physical violence, it is more advantageous for early infant development to exhibit emotional expressiveness, emotional awareness, and prosocial behavior.

Physical and online security must be meticulously handled. Surveillance cameras are an effective method for ensuring the safety of potentially hazardous areas. Unlike humans, they provide continuous monitoring and have the ability to capture extensive volumes of video footage. This footage can be promptly reviewed to identify any abnormalities and implement any remedial actions. The use of automated video and photo identification is crucial for addressing contemporary security concerns. By using machine learning (ML) or deep learning (DL) techniques, it is possible to promptly identify and respond to dubious actions shown in photos and videos. Action recognition, a crucial aspect of violence detection, requires both data and effort. My art only recognizes war, riots, stone flinging, and overall hatred.

The use of computer vision technology and the automation of surveillance systems have significantly enhanced video analysis, leading to notable improvements in both public and industrial security. This is particularly accurate in the domains of human activity recognition, behavioral analysis, and violence detection. Despite recent advancements, real-time surveillance systems still struggle to identify and evaluate violent incidents due to their constrained processing capabilities. Our IIoT-enabled VD-Net (Violence Detection Network) AI platform can identify aggressive behavior in both public and private areas. The model utilizes lightweight ST-TCN blocks and bottleneck layers to specifically address crucial features of the input sequence. Classifiers distinguish between violent and peaceful activities based on the features they have learned. If our system detects any kind of violence, it should promptly inform the relevant authorities. By thoroughly evaluating our method using both surveillance and non-surveillance datasets, we have shown that it improves the state-of-the-art accuracy by 1-4 percentage points.

Amidst the COVID-19 epidemic, there has been a significant increase in the occurrence of violent dating. This essay examines the impact of the normalization of dating violence on social media on young individuals. This research aims to explore how young individuals perceive online dating violence by analyzing the "pretend to punch your girlfriend" trend on TikTok. How does their emotional reaction indicate a change in viewpoint on the equality of relationships and the presence of violence in dating? This study used a mixed-method approach to examine how the "feeling rules" of young individuals are influenced by violent dating situations. Ultimately, this research ends by presenting strategies for platform design aimed at mitigating instances of violent dating on social media.

The interaction between the government and those who oppose it is a frequently discussed subject in rational choice theories and extensive study. This study uses computational approaches to analyze ethnographic interviews in order to get a comprehensive understanding of dissident ideology. Research indicates that dissidents tend to respond with violence when confronted with violent persecution, and with peaceful means when faced with nonviolent persecution. The majority of individuals choose government intervention or security above engaging in protests. These results provide more proof of the reciprocal relationship between states and dissidents, as well as the cognitive parallels between political dissent and collaboration, which is often linked to tit-for-tat dynamics. Furthermore, these findings demonstrate that governmental repression is the primary element that inspires resistance, rather than social contagion or perceived relative hardship. Heuristic thinking tendencies suggest that dissidents may be more open to government reform and collaboration than oppressive governments portray.

Human violence detection has several applications, one of which is in video surveillance systems. The capability to detect violent situations in real-time has the potential to decrease crime rates and prevent loss of life. In most cases, accuracy is considered more important than efficacy and practicality when it comes to conceptions and investigations. Our system is very effective and precise, since it can detect hostile human behavior in real-time. The proposed model consists of three modules: the Spatial Motion Extractor (SME) for extracting regions at the frame level, the Short Temporal Extractor (STE) for extracting temporal features of fast movements at the frame level, and the Global Temporal Extractor (GTE) for extracting long-lasting temporal features at the frame level and fine-tuning the model. We conducted a real-time assessment of the plan to evaluate its effectiveness and efficiency. The method's superiority is shown by the results obtained from the RWF-2000, Movies, and Hockey datasets. The VioPeru dataset was generated by using Peruvian video surveillance camera recordings of both aggressive and non-aggressive incidents, with the aim of assessing the real-time capabilities of the model. Our model produced optimal results while using this dataset.

Video Violence Detection (VVD) is essential in video surveillance, but it gets difficult in crowded environments because of the variety and intricacy of violent episodes. Violent activities are often characterized by a fast, chaotic, and disorganized progression. We provide a unique approach called Angle-level Co-occurrence for analyzing violent video clips. This method utilizes a matrix that effectively captures the relevant aspects of these clips. Our video volume model has a rank-3 tensor that includes a fiber confined to a single plane. ALCM captures the distribution of fiber pairs that have certain similarities in one plane of the rank 3 tensor. It measures the co-occurrence of two distinct quantized angle values between fibers and their neighboring fibers. To construct a comprehensive TOP-ALCM that fully characterizes volume violence, we calculate three ALCMs for three mutually perpendicular planes. In addition, we provide a conventional VVD framework that utilizes CNN directly, as well as a DL-based framework that utilizes TOP-ALCM properties such as entropy, homogeneity, and energy for classification. The TOP-ALCM outperforms state-of-the-art VVD algorithms in experimental settings.

This study provides a comprehensive analysis of the latest instances of violence against minors (individuals under the age of 17) on primetime television. We used same sampling methodology and codebook as the first National Television Violence Study to classify 765 primetime TV episodes across 21 broadcast and cable networks throughout the 2016/2017 period. We segregated television targeted towards children and programming targeted towards adults in order to analyze and contrast the frequency and nature of violent content. The proportion of juvenile programming that contains violent content has significantly decreased over the last two decades, while it remains greater than that of adult programs. The portrayal of violence in children's television is idealized, but it is presented in a more sanitized and trivialized manner, similar to how it is shown in adult programming. Considerations while instructing children about hostility.

We conducted a comprehensive analysis of 540 news items to determine the extent to which the media's depiction of violence aligns with the scientific understanding of its impact on viewers' aggressiveness. Over the last three decades, a correlation has been shown between violence and the presence of violent news content. The tone has rapidly returned to a neutral state after the year 2000. Possible factors contributing to this change include the journalist's gender, the quantity of independent sources used, and the kind of media being examined (e.g., television vs video games). Evaluation of the possible impacts on readers resulting from this news piece.

Perceptions of victims and demographic groupings vary among the public based on the manner in which crimes are shown by the media. News media has the capacity to sustain detrimental stereotypes and minimize acts of violence against transgender individuals, thus exacerbating the significant obstacles already encountered by this community. This study analyzes 316 news pieces that reported on 27 transgender deaths in the US in 2016. It focuses on the portrayal of transgender identities, both positive and negative, the use of language that either supports or undermines these identities, and the systematic organization of transphobia. We go into many subjects, outcomes, and possible paths for further investigation.

The academic interest in monitoring systems is seeing a significant surge. Security cameras mounted in public locations such as schools, hospitals, companies, and roads may be used to predict incidents, surveil online activities, analyze data with specific objectives, and detect intrusions by capturing significant movements and events. This study introduced deep learning architectures to enable real-time identification of violent crime situations using video surveillance. The goal is to gather video footage from live surveillance systems at crime scenes and analyze the characteristics using Deep Recurrent Neural Networks (DRNN) and spatio-temporal (ST) classification. Once the raw material was converted into video frames, the characteristics were retrieved and arranged. It has the ability to detect hostile and violent behavior in real-time and identify anomalies that deviate from normal patterns. We use the extensive UCF Crime anomaly dataset for both the training and testing of our approach. The recommended approach demonstrates strong performance on real-time datasets, with an accuracy of 98%, precision of 96%, recall of 80%, and an F-1 score of 78%.

Content filtering guarantees that individuals of all age groups may browse the internet without any apprehension. Human moderators manually identify material as either violent or nonviolent throughout the moderation process. Content moderators are exposed to graphic violence, which may have a profound emotional and psychological impact on them. This study presents a machine learning algorithm that can accurately classify films as violent or non-violent based on their violent content. It utilizes both auditory and visual information. The first stage involves the segregation of audio and visual components. An audio classifier is capable of discerning between sounds that are characterized by high volume and those that are characterized by low volume. A video classifier assesses the level of violence in a video by relying on the outcomes of an audio classifier that categorizes the sounds as calm.

R. Yadav (2018), This paper presents a recommendation system for e-commerce that utilizes client profiles to provide personalized product recommendations. The system uses data about the clients' preferences and previous purchases to generate recommendations.

V. Prakaulya (2017) The paper proposes a time series decomposition model for forecasting railway passenger numbers. The model decomposes the time series data into different components, such as trend and seasonality, and uses them to make predictions about future passenger numbers.
D. Bhuriya (2017) This paper explores the use of linear regression for predicting stock market trends. The authors investigate the relationship between stock market variables and use regression analysis to make predictions about future stock price.
R. Verma (2017) The paper focuses on the use of neural networks for stock market prediction. The authors train neural networks using historical stock market data and use them to predict future stock prices.
Kewat (2017) The paper examines the application of support vector machines (SVMs) for forecasting financial time series. The authors train SVM models using historical financial data and evaluate their performance in predicting future values.
A. Sharma (2017) This paper provides a survey of different machine learning approaches used for stock market prediction. The authors review various techniques, including regression, neural networks, and support vector machines, and discuss their effectiveness in predicting stock prices.
S. Sable (2017) The paper proposes the use of genetic algorithms and evolution strategies for stock price prediction. The authors employ these optimization techniques to optimize the parameters of a prediction model and improve its accuracy.
A. Roshan (2018) The paper presents a credit card fraud detection system based on decision tree technology. The authors utilize decision trees to classify credit card transactions as either fraudulent or legitimate based on various features and patterns.
H. Soni (2018) This paper explores the use of machine learning techniques to identify patients with rare diseases from electronic health records. The authors develop models that analyze patient data and make predictions about the likelihood of rare diseases.
A. Saxena (2020) The paper proposes a glaucoma detection system based on convolutional neural networks (CNNs). The authors train CNN models using eye images and use them to classify images as either normal or indicative of glaucoma.
B. Bamne (2020) The paper investigates the application of transfer learning and convolutional neural networks for object detection. The authors utilize pre-trained CNN models and adapt them for detecting objects in different contexts.
Gupta, P. (2022) The paper presents an AIoT-based device that enables real-time object recognition for visually impaired individuals. The system combines object recognition algorithms with voice conversion technology to provide auditory feedback to users.
A. Taiwade (2022) This paper proposes a hierarchical K-means clustering method for a friend recommendation system. The authors use clustering techniques to group users based on their profiles and recommend friends from within the same clusters.
R. Baghel (2022) The paper introduces a deep learning-based system for human face mask identification. The authors utilize deep learning algorithms and OpenCV techniques to detect and classify faces as either wearing or not wearing masks.
M. Ranjan (2022) The paper investigates the use of random forest and deep learning techniques for cancer prediction. The authors develop models using these methods and evaluate their performance in predicting cancer cases [29].
Singh, Upendra (2022) The paper presents a system for activity detection and people counting using the Mask-RCNN architecture combined with bidirectional ConvLSTM. The authors use this system to analyze video data and detect different activities and count the number of people involved.
Singh, Shani Pratap (2022) This paper proposes a multi-stage CNN architecture for face mask detection. The authors develop a system that can detect whether a person is wearing a face mask or not using deep learning techniques.
U. Singh (2022) The paper focuses on the analysis and detection of Monkeypox using the GoogLeNet model. The authors utilize the GoogLeNet model to classify images and identify cases of Monkeypox.

Proposed Method

3.1 Proposed Architecture

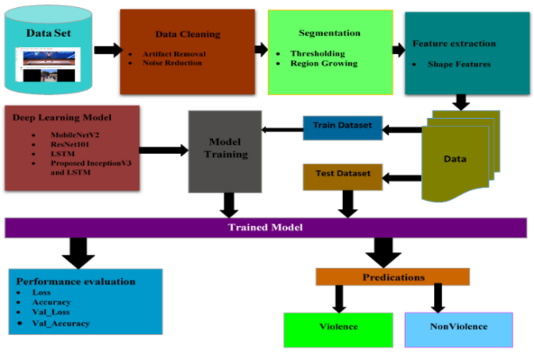

Figure 1. Proposed architecture for the classification of video material into "Violence" and "NonViolence" categories.

The diagram in Figure 1 depicts a workflow for a machine learning algorithm designed to categorize video footage as either "Violence" or "NonViolence". Firstly, a dataset is prepared by performing data cleaning procedures such as removing artifacts and reducing noise. Subsequently, the data is subjected to segmentation, using techniques such as thresholding and region growth, which is then followed by the extraction of form characteristics. Using the provided data, several deep learning models, such as MobileNetV2, ResNet101, LSTM, and a suggested InceptionV3 with LSTM model, are trained using a predefined training dataset. Following the training phase, the models undergo evaluation using a separate dataset to determine their performance based on measures such as loss, accuracy, validation loss, and validation accuracy. The result of this procedure is a trained model capable of making predictions to classify fresh video footage as either "Violence" or "NonViolence." This workflow exemplifies a thorough strategy for addressing the job of video content categorization by following a sequence of organized processes, starting with the initial data preparation and concluding with the final prediction task.

3.2. Pseudocode for predict video content as 'Violence' or 'NonViolence' using InceptionV3 and LSTM

Load Models and Set Threshold:

- Load the InceptionV3 and LSTM models.
- Set the threshold for classifying videos as 'Violence' or 'NonViolence.'

Extract and Process Video Frames:

- Extract frames from the video, resize them, and normalize the pixel values.

Generate Video Features:

- Use InceptionV3 to extract features from the frames.
- Combine these features into a single vector for the video.

Sequence and Predict:

- Prepare the feature vector for LSTM input.
- Run the sequence through LSTM to get a violence probability score.

Classify and Output Results:

- Classify the video based on the score and threshold.
- Output the classification and, if needed, the probability scores.

Implementation And Result Discussion

4.1 Dataset: The dataset is divided into two separate directories: 'NonViolence' and 'Violence'. The 'NonViolence' category has a collection of 1,000 films that portray a range of real-life scenarios, including dining, athletic activities, and singing, without any incidences of violence. In contrast, the 'Violence' category has 1,000 films, each showcasing intense acts of violence in different situations. The precise classification guarantees that the dataset can efficiently train and assess machine learning models assigned with the duty of differentiating between violent material and non-violent actions.

4.2 Illustrative Example

Figure 2. Displays a series of frames from a film where the word "Viole" is superimposed in a prominent red font, partly obstructing the picture

Figure 2 depicts a series of frames from a film where the word "Viole" is superimposed in a prominent red font, partly obstructing the view of the scene. The frames portray an open-air environment, either a street or a park, with individuals discernible in the distance. The red text's arrangement across the frames suggests that the action is centered upon a person or item moving across the picture, signifying movement or a specific point of interest. Considering the same backdrop and illumination in each shot, it is probable that the series represents a brief period of time. The inclusion of the superimposed writing implies that it might be a component of a video analysis or a tracking demonstration, perhaps within the realm of motion detection or video editing.

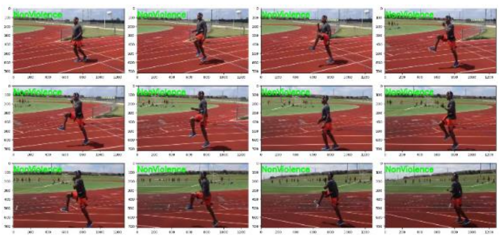

Figure 3. A succession of video frames, where each frame is superimposed with the phrase "NonViolence" in a green typeface.

The image 3 displays a succession of frames from a movie, where each frame has the phrase "NonViolence" superimposed on it in a green font. The footage is recorded outside on a track, showing a guy engaged in a workout, maybe doing high-knee running exercises. The persistent superimposition of the term "NonViolence" on every frame implies that the phrase may be a tag or a classification result generated by a video analysis or machine learning system that is classifying the film's content. The combination of the words, the outside environment, the man's clothing, and his actions combined create the perception of a film relating to sports or exercise. This video has been examined and concluded to be non-violent in nature.

4.3 Comprative Result

Figure 4. Comparison of model losses

The figure 4 titled "Model Loss Comparison" visually compares the loss values of four different models: ResNet50, LSTM, MobileNetV2, and a proposed InceptionV3 combined with LSTM model. The proposed InceptionV3 and LSTM model shows the lowest loss, followed by MobileNetV2, LSTM, and ResNet50, indicating that the InceptionV3 and LSTM model performs better in minimizing loss compared to the others.

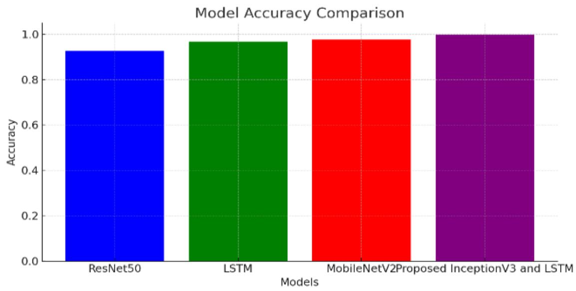

Figure 5. Comparison of model accuracy

The figure 5 titled "Model Accuracy Comparison" compares the accuracy of four models: ResNet50, LSTM, MobileNetV2, and a proposed InceptionV3 combined with LSTM model. All models show very high accuracy, with the proposed InceptionV3 and LSTM model achieving the highest accuracy, slightly outperforming the others. This indicates that the proposed model is slightly more effective in correctly classifying data compared to ResNet50, LSTM, and MobileNetV2.

Conclusion

Deep learning algorithms for video material categorization have shown their ability to recognize complicated patterns and forecast outcomes. ResNet50, LSTM, MobileNetV2, and a suggested hybrid of InceptionV3 and LSTM performed differently, with the hybrid model surpassing others in training and validation measures. Its validation result shows that its better accuracy and reduced loss metrics demonstrate its resilience in learning from training data and generalization to unseen data. We prepare the dataset for learning by cleaning, segmenting, and extracting features. The hybrid model, which uses InceptionV3's spatial pattern recognition and LSTM's sequential data processing, captures both frame details and the video sequence's context. Modern AI systems' deep learning approach from data pretreatment to prediction shows their sophistication. This approach allows the sophisticated categorization of video footage as 'Violence' or 'NonViolence,' demonstrating the potential for these models in surveillance, content management, and media research.

Reference

1. Weaver, Andrew J., Asta Zelenkauskaite, and Lelia Samson. "The (non) violent world of YouTube: Content trends in web video." Journal of Communication 62, no. 6 (2012): 1065-1083.
2. Choi, Yun Jung, and Changsook Kim. "A content analysis of cognitive, emotional, and social development in popular kid's YouTube." International Journal of Behavioral Development (2024): 01650254241239964.
3. Jain, Mahaveer, and Mukesh Kumar. "A Review of Violence Detection Techniques." In 2024 2nd International Conference on Computer, Communication and Control (IC4), pp. 1-6. IEEE, 2024.
4. Khan, Mustaqeem, Abdulmotaleb El Saddik, Wail Gueaieb, Giulia De Masi, and Fakhri Karray. "VD-Net: An Edge Vision-Based Surveillance System for Violence Detection." IEEE Access 12 (2024): 43796-43808.
5. Maddocks, Sophie, and Fallon Parfaite. "“Watch me pretend to punch my girlfriend”: exploring youth responses to viral dating violence." Feminist Media Studies 24, no. 1 (2024): 103-118.
6. Dornschneider-Elkink, Stephanie, and Nick Henderson. "Repression and dissent: How tit-for-tat leads to violent and nonviolent resistance." Journal of Conflict Resolution 68, no. 4 (2024): 756-785.
7. Huillcen Baca, Herwin Alayn, Flor de Luz Palomino Valdivia, and Juan Carlos Gutierrez Caceres. "Efficient human violence recognition for surveillance in real time." Sensors 24, no. 2 (2024): 668.
8. Hu, Xing, Zhe Fan, Linhua Jiang, Jiawei Xu, Guoqiang Li, Wenming Chen, Xinhua Zeng, Genke Yang, and Dawei Zhang. "TOP-ALCM: A novel video analysis method for violence detection in crowded scenes." Information Sciences 606 (2022): 313-327.
9. Martins, Nicole, and Karyn Riddle. "Reassessing the risks: An updated content analysis of violence on US children's primetime television." Journal of Children and Media 16, no. 3 (2022): 368-386.
10. Ferguson, Christopher J., Anastasiia Gryshyna, Jung Soo Kim, Emma Knowles, Zainab Nadeem, Izabela Cardozo, Carolin Esser, Victoria Trebbi, and Emily Willis. "Video games, frustration,

violence, and virtual reality: Two studies." British journal of social psychology 61, no. 1 (2022): 83-99.
11. Osborn, Max. "US news coverage of transgender victims of fatal violence: An exploratory content analysis." Violence against women 28, no. 9 (2022): 2033-2056.
12. Sahay, Kishan Bhushan, Bhuvaneswari Balachander, B. Jagadeesh, G. Anand Kumar, Ravi Kumar, and L. Rama Parvathy. "A real time crime scene intelligent video surveillance systems in violence detection framework using deep learning techniques." Computers and Electrical Engineering 103 (2022): 108319.
13. Rishab, K. S., P. Mayuravarsha, Yashwal S. Kanchan, M. R. Pranav, and Roopa Ravish. "Detection of Violent Content in Videos using Audio Visual Features." In 2023 International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems (ICAECIS), pp. 600-605. IEEE, 2023.
14. Cheng, Ming, Kunjing Cai, and Ming Li. "RWF-2000: an open large scale video database for violence detection." In 2020 25th International Conference on Pattern Recognition (ICPR), pp. 4183-4190. IEEE, 2021.
15. R. Yadav, A. Choorasiya, U. Singh, P. Khare, and P. Pahade, “A Recommendation System for E-Commerce Base on Client Profile,” in Proceedings of the 2nd International Conference on Trends in Electronics and Informatics, ICOEI 2018, 2018, doi: 10.1109/ICOEI.2018.8553930.
16. V. Prakaulya, R. Sharma, U. Singh, and R. Itare, “Railway passenger forecasting using time series decomposition model,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212725.
17. D. Bhuriya, G. Kaushal, A. Sharma, and U. Singh, “Stock market predication using a linear regression,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212716.
18. R. Verma, P. Choure, and U. Singh, “Neural networks through stock market data prediction,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212717.
19. Kewat, R. Sharma, U. Singh, and R. Itare, “Support vector machines through financial time series forecasting,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212859.
20. Sharma, D. Bhuriya, and U. Singh, “Survey of stock market prediction using machine learning approach,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212715.
21. S. Sable, A. Porwal, and U. Singh, “Stock price prediction using genetic algorithms and evolution strategies,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212724.
22. Roshan, A. Vyas, and U. Singh, “Credit Card Fraud Detection Using Choice Tree Technology,” in Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018, 2018, doi: 10.1109/ICECA.2018.8474734.
23. H. Soni, A. Vyas, and U. Singh, “Identify Rare Disease Patients from Electronic Health Records through Machine Learning Approach,” in Proceedings of the International Conference on Inventive Research in Computing Applications, ICIRCA 2018, 2018, doi: 10.1109/ICIRCA.2018.8597203.
24. Saxena, A. Vyas, L. Parashar and U. Singh, "A Glaucoma Detection using Convolutional Neural Network," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 815-820, doi: 10.1109/ICESC48915.2020.9155930.
25. Bamne, N. Shrivastava, L. Parashar and U. Singh, "Transfer learning-based Object Detection by using Convolutional Neural Networks," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 328-332, doi: 10.1109/ICESC48915.2020.9156060.
26. Gupta, P., Shukla, M., Arya, N., Singh, U., Mishra, K. (2022). Let the Blind See: An AIIoT-Based Device for Real-Time Object Recognition with the Voice Conversion. In: Al-Turjman, F., Nayyar, A.

(eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80928-7_8

27. Taiwade, N. Gupta, R. Tiwari, S. Kumar and U. Singh, "Hierarchical K-Means Clustering Method for Friend Recommendation System," 2022 International Conference on Inventive Computation Technologies (ICICT), Nepal, 2022, pp. 89-95, doi: 10.1109/ICICT54344.2022.9850852.
28. R. Baghel, P. Pahadiya and U. Singh, "Human Face Mask Identification using Deep Learning with OpenCV Techniques," 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2022, pp. 1051-1057, doi: 10.1109/ICCES54183.2022.9835884.
29. M. Ranjan, A. Shukla, K. Soni, S. Varma, M. Kuliha and U. Singh, "Cancer Prediction Using Random Forest and Deep Learning Techniques," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Indore, India, 2022, pp. 227-231, doi: 10.1109/CSNT54456.2022.9787608.
30. Singh, Upendra, Gupta, Puja, and Shukla, Mukul. ‘Activity Detection and Counting People Using Mask-RCNN with Bidirectional ConvLSTM’. 1 Jan. 2022 : 6505 – 6520.
31. Singh, Shani Pratap and Shukla, Jayesh and Sharma, Shaili and Daga, khushhal and Bhalavi, Brahman Singh and Singh, Upendra, Face Mask Detection using Multi-Stage CNN Architecture (July 10, 2021). Proceedings of the International Conference on IoT Based Control Networks & Intelligent Systems - ICICNIS 2021, Available at SSRN: https://ssrn.com/abstract=3884022 or http://dx.doi.org/10.2139/ssrn.3884022
32. U. Singh and L. S. Songare, "Analysis and Detection of Monkeypox using the GoogLeNet Model," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022.