International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Exploring Deep Learning Techniques for Abnormality Detection: A Comparative Analysis on UCF Crime Dataset.

Anchal Pathak[1], Ruchi Jayaswal*[2], Smita Mahajan[3]

Submitted: 27/01/2024 **Revised**: 05/03/2024 **Accepted**: 13/03/2024

Abstract*:* Video surveillance systems have been used significantly to enhance the security of a variety of places that are both private and public. The automatic detection of abnormalities in video surveillance is an intriguing research topic. Despite the recent development of multimedia-based anomaly detection algorithms, video surveillance is still inadequate to detect unexpected events such as illegal acts and crimes. In this study, comparison-based deep learning models are utilised to construct an automated framework that can detect anomalies in videos. To identify abnormalities these four models, VGG16 CNN, VGG16 Bi-LSTM, Slow Fast Network, and DenseNet121 have been used. The proposed approaches have been evaluated using the UCF-Crime dataset. This study distinguishes between usual and unusual events, showing that comparison-based models were capable of classifying each abnormal event. Slow Fast Network obtained 99% accuracy on the UCF-Crime dataset, VGG16 CNN achieved 98% accuracy, VGG16 Bi-LSTM scored 96% accuracy, and DenseNet121 achieved 70% accuracy. Furthermore, the proposed models outperformed comparable deep-learning models in terms of performance accuracy. Our comparison analysis paper describes:

- How the comparison-based approach is effective for understanding the deep learning model for Abnormal detection using the UCF crime dataset?
- This paper highlights that Slow Fast Network and VGG16 CNN Deep Learning Models performed better than the other existing models and also helped to classify the Abnormality Detection from the UCF Crime Dataset.

Keywords: *Video Surveillance, Abnormality, Deep Learning, VGG16 CNN, VGG16 BiLSTM, RNN, Slow Fast Network, UCF Crime, DenseNet121.*

1. Introduction

With a lot of recent difficulties that are affecting the public sector, safety as well as security, there is a growing need for video surveillance to monitor public situations. At first glance, it appears that analysing video surveillance footage for extracting significant as well as relevant information data from behavioural patterns, detecting anomalous acts, and giving quick reactions is a simple task for a person. However, due to significant limitations in human capabilities, it is also difficult for a person to monitor multiple signals at the same time. It is also an endeavour that takes time for necessitates to use of several resources such as personnel and work area. As a result, an Automatic identification approach is critical for Detecting anomalous events is one of the various subdomains of behaviour understanding from video surveillance. In video surveillance systems, the identification of anomalies becomes a challenging process that may encounter various issues like abnormal events being infrequent, and it is difficult to gather large a database of such events. This shortage of samples can cause a few challenges to the learning process. An "anomaly" is defined as something that does not follow a specific pattern. As a result, we are incapable to assign a demonstrate to abnormal events [1].

Anomaly identification could be a method utilized to identify anomalous events, which are categorized into a few major indoor categories, including violence, panic, falls, assaults, patients, and the elderly. These classifications are determined by analysing motions, body movements, and interactions between individuals [1]. The most reason of anomaly detection is to progress the security, security, efficiency, and quality of indoor and outdoor spaces by giving real-time observation and early warning frameworks [2]. As indoor and outdoor environments become progressively complex and interconnected, the require for successful anomaly detection components proceeds to extend, making this innovation a basic component of advanced indoor control and mechanization arrangements. Peculiarity location is portion of the broader field of machine learning and manufactured intelligence [1]. It centers on identifying abnormal or anomalous events, behaviour's, or conditions that happen within the environment [1].The task of modelling and processing the abnormal scene outcomes may appear challenging, even impossible. Recognition of abnormalities in video surveillance is a critical task which can be challenging

[1] *Symbiosis Institute of Technology, Pune, India , anchal.pathak.mtech2022@sitpune.edu.in.*
[2]*Symbiosis Institute of Technology, Pune, India, Ruchi.jayaswal@sitpune.edu.in.*
Symbiosis Institute of Technology, Pune, India , smita.mahajan@sitpune.edu.in
** Corresponding Author Email: Ruchi.jayaswal@sitpune.edu.in.*

because abnormalities are being characterized such as behaviours that depart from known patterns. In some situations, such as a gun club, an abnormal behaviour may be a typical activity. Shooting is normal in shooting clubs, even though it is frequently inappropriate behaviour. Alternatively, certain behaviours that are not necessarily abnormal may be anomalous circumstances [2]. A system that stands out is the deep learning-based method for detecting and preventing anomalies. Several Law regulatory agencies across the world are investigating using deep learning (DL) [1] [2] technologies to safeguard public sector safety [2]. The feedback can utilise DL models to automate the process of extracting data from recorded videos when abnormalities are detected. The categorization of abnormal behaviours is separated into unsupervised, supervised, and semi-supervised learning models from the standpoint of DL [2]. The model has been trained on both usual and unusual behaviours in one deep learning technique. On the other side, deep learning models are trained on both normal and abnormal behaviours in the context of supporting multi-model learning. Supervised deep learning has been used in several studies to identify unusual behaviours in videos. Convolutional neural networks (CNNs) [1], recurrent neural networks (RNNs) [1], long-short-term memory (LSTM) [1], gated recurrent units (GRUs) [1], simple recurrent units (SRUs) [1] are some of the DL models uses for anomaly recognition [1].

To identify unusual behaviour from a multiple-learning strategy, this study proposes to use a comparison of the deep learning model. Anomalies include many anomalous behaviours that occur in the real world. The current ponders, however, focuses on unusual behaviours which are found within the UCF Crime dataset [3], which contains violent and unusual behaviour captured on video in a range of public places. Within the proposed technique for anomaly detection is utilized to extricate the high-level of information from video frames, the VGG16 has been implemented as a CNN-BiLSTM [4], Slow Fast Network [5], and Transfer Learning model [6]. The output produced by the suggested method indicates whether the input video has abnormal or normal behaviour. The limitations of video surveillance systems with human monitoring can be reduced, and abnormal activity detection accuracy increased with the help of this provided model. Below is an outline of the proposed method's key contribution:

1. A comparison model of deep learning is used for abnormal detection from video surveillance.
2. UCF Crime dataset [3], which comprises normal scenes collected by surveillance cameras in 14 sorts of abnormal events.
3. In order to properly understand each abnormality type, we sorted two updated datasets from UCF Crime by segregating normal and abnormal events.

2. Methodology

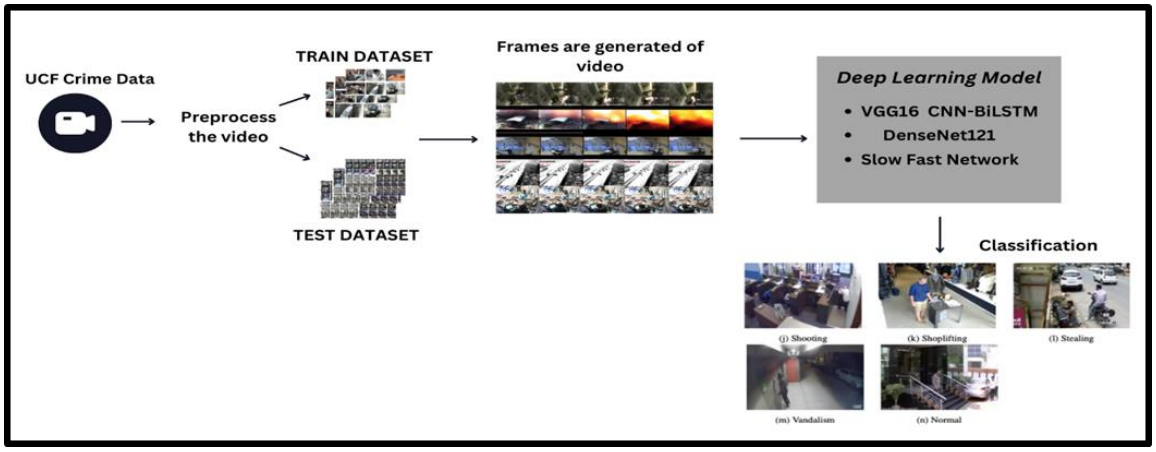

Fig.1. Proposed Methodology of Abnormality Detection.

The proposed framework of an abnormality detection system using videos is shown in Fig 1. To perform an in-depth analysis of three important deep learning architectures: VGG16, Slow Fast Networks, and DenseNet-121, with the goal of evaluating their performance across multiple tasks. The method utilized in this study includes a efficient approach including data preparation, model selection, training, and evaluation. Data preprocessing strategies are reliably connected to all models to standardize input information. Three deep learning structures are chosen as the center of this study: VGG16, Slow Fast Networks, and DenseNet-121. These models are known for a variety of designs, from the classic VGG16 to a committed low-speed network outlined for action recognition, to DenseNet-121 with dense network patterns.

2.1. Video Preprocessing

Video preprocessing can be a key step in working with the UCF Crime Dataset [3], a collection of video sequences that capture a variety of criminal activities in very diverse environments. To preprocess the videos for the UCF crime dataset, we first convert the video clips into individual

frames and sample frames to improve computational efficiency. To improve the generalization of the model, standard image preprocessing techniques such as resizing, normalization, and data augmentation are applied to each frame. To take time aspects into account, you can construct short video sequences by adding successive frames. Data labeling includes classifying outlines or sequences into the suitable crime type. Subsequently, the dataset is separated into training, validation, and test sets to provide suitable coordination between classes [1][2]. We organize preprocessed frames into a organized catalog to encourage data loading and model training. This preprocessing pipeline is designed to plan the UCF crime dataset for video-based deep learning tasks such as action acknowledgment or anomaly detection [2].

2.2. VGG16

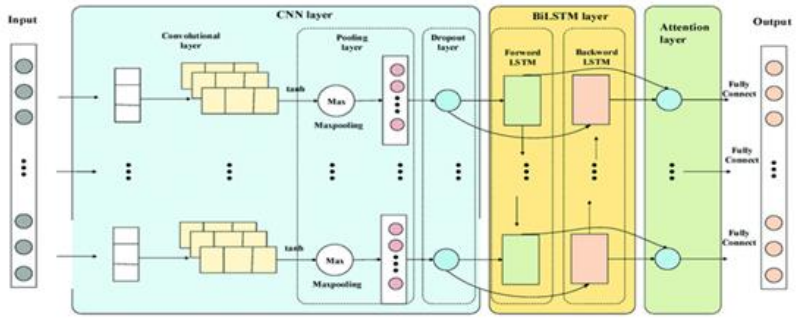

Fig.2. VGG16 Architecture [7]

The VGG16 CNN-BiLSTM architecture [7] proposed for the UCF crime video dataset is designed to effectively capture spatial and temporal features in video sequences to accurately recognize criminal activities. In this design, the VGG16 CNN acts as a spatial feature extractor. Pre-trained VGG16 layers are utilized to extricate high-level visual highlights from individual frames of video clips. These highlights were passed through a Bidirectional Long Short-Term Memory (BiLSTM) network [8]. The BiLSTM component is mindful for modeling temporal elements by considering both forward and backward data flows, which is imperative for understanding the sequential nature of video frames [8]. The network has been trained broadly on the UCF crime dataset and points to learn unique features to recognize different criminal activities over time. During training, the model's weights are upgraded to minimize classification blunder, permitting it to adapt to the complex subtle elements of the dataset. To evaluate the model efficiently, here we separate the UCF crime dataset into training, validation, and test sets. The program trains and learns how to predict particular types of criminal activity inside video sequences. Validation sets are used to optimize hyperparameters and prevent overfitting. The VGG16 CNN-BiLSTM design features collecting both spatial and temporal data, enabling effective recognition of complex criminal activities including unique spatial locations and sequential actions. The resulting working model is expected to be useful in surveillance and security applications, providing a robust solution for detecting criminal activity in video data. The VGG16 CNN-Bi-LSTM demonstrate ordinarily consists of two primary components: the convolutional neural network (CNN) and the bidirectional long short-term memory (Bi-LSTM) network [9]. Each component has its own set of mathematical operations. The VGG16 CNN architecture involves a series of convolutional layers followed by max pooling layers. The mathematical operations in a convolutional layer can be represented as follows:

$$\text{Conv}(Y)=Z*Y+a \quad (1)$$

Where; Z is the weight matrix, Y is the input, and a is the bias term. The convolution is typically followed by an activation function, such as ReLU

$$\text{MaxPool}(Y)=\max(Y) \quad (2)$$

Where, this operation downsamples the spatial dimensions of the input by taking the maximum value in each pooling window.

The Bidirectional Long Short-Term Memory network is a type of RNN that is capable of capturing temporal dependence in the sequential data. The operations in a Bi-LSTM layer are as follows:

$$\text{Forward LSTM: } ht=\text{LSTM}(ht-1, xt) \quad (3)$$

Where: ht is a hidden state at time t obtained by processing the input xt in the forward direction.

$$\text{Backward LSTM: } ht=\text{LSTM}(ht+1, xt) \quad (4)$$

Where: ht is a hidden state at time t obtained by processing the input xt in the backward direction.

$$\text{Output; } ht=[ht; ht] \quad (5)$$

The final hidden state ht is a concatenation of forward and backward hidden states.

The VGG16 CNN and Bi-LSTM are typically combined by

using the features that are extracted by the CNN as input sequences for Bi-LSTM. The output of the Bi-LSTM is then used for making predictions. The mathematical formula for the combined model would involve the forward pass through the VGG16 CNN followed by feeding the features to the Bi-LSTM. The details of the equations would depend on the specific implementation and design choices made in the model architecture.

2.3. DenseNet121

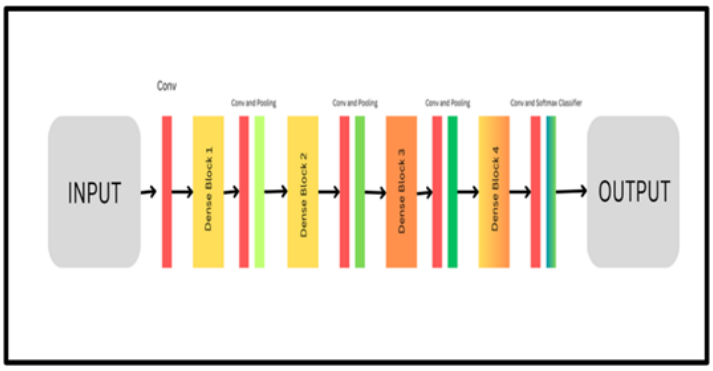

Fig.3. Densenet121 Architecture [41]

DenseNet-121 as shown in Fig.4 short for a dense convolutional network consisting of 121 layers is a deep learning architecture mainly used for image classification tasks. It belongs to the Convolutional Neural Networks (CNN) family and is characterized by dense connectivity patterns that promote feature reuse and gradient flow during training. The architecture consists of multiple densely connected blocks, each containing a sequence of convolutional layers with a fixed number of feature maps. The key innovation in DenseNet121 is the direct connections between layers within the same block, where each layer receives feature maps as input from all the previous layers. This design results in highly efficient parameter usage and enables the network to learn rich and discriminative features [6].

In a normal DenseNet-121 design, the network comprises four densely connected blocks, a transition layer between each block to control the spatial dimensions, and a completely connected layer is taken after by the global average pooling layer for the final classification. The dense connectivity and feature reuse make DenseNet-121 highly effective at learning representations from images, driving to state-of-the-art execution on a wide extend of image classification tasks. DenseNet121is split into Dense Blocks, where the feature map dimensions remain constant inside a block but the number of filters between them changes. Transition Layers are the layers between the blocks that restrain the number of channels to half of the present channels. DenseNet121 requires less parameters than the comparing standard CNN, permitting for include reuse by disposing of duplicate feature maps [10]. As a result, the lth layer receives as input the feature maps of all preceding layers, a0,...,al-1:

$a_l=F_l([a_0,a_1\dots a_{l-1}])$ (6)

Where: [a0,...,al-1] is the concatenation of feature maps, i.e. the output produced in all the layers preceding l (0,...,l-1). The multiple inputs of Fl are concatenated into a single tensor to ease implementation [6].

While going through each wide layer, the size of the feature map increases, with each layer adding 'B' features on top of the global state (existing features). This parameter 'B' is known as the network's growth rate, and it regulates the amount of information included in each layer of the network. If each function Fl generates B feature maps, the lth layer.

$B_1=B_0+B*(L-1)$ (7)

B0 is the number of channels in the input layer, hence the input feature maps. DenseNet121, unlike the existing network structures, can have extremely narrow layers. Although each layer only generates B output feature maps, the number of inputs can be quite large, particularly for the following levels. In order to improve the efficiency and speed of computations, a 1x1 convolution layer can be included as a bottleneck layer before each 3x3 convolution [6].

2.4. Slow Fast Network

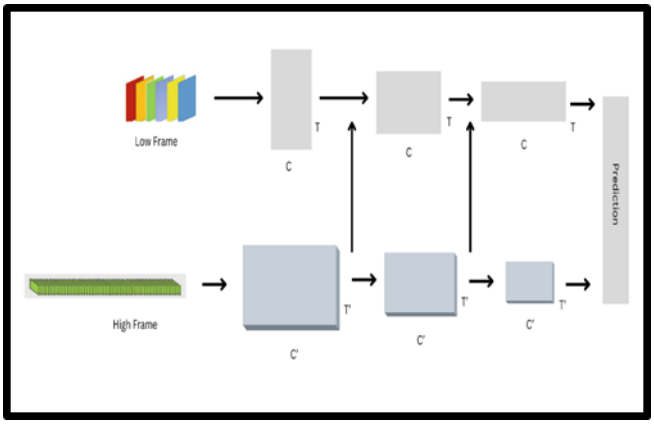

Fig.4. Slow Fast Network Architecture [5]

The Slow Fast Network is a two-pathway deep neural network architecture designed for video analysis, particularly for capturing both spatial and temporal information efficiently. In a Slow Fast network, the Fast pathway is lightweight since it uses a fraction (T', e.g., 1/8) of channels; lateral connections combine them. The Slow pathway encompasses a low outline rate and low temporal resolution, while the Fast pathway features a high frame rate and C'x higher temporal resolution. [5]. Slow Fast networks can be characterized as a single stream design with two distinctive framerates, but we utilize the concept of pathways to replicate the biological Parvo- and Magnocellular counterparts. Our common design incorporates a Slow pathway and a Fast pathway that are connected together by lateral connections to make a Slow Fast network [5]. Adapted to the UCF crime video dataset, the Slow Fast Network gives a robust architecture for accurately capturing data within the frame of spatial and temporal video sequences, improving the detection of criminal activity. This design is characterized by two particular pathways: a "slow" pathway, which is dependable for processing spatial features, and a "fast" pathway, which is dedicated to temporal features. The "slow" path uses deeper, slower CNNs to capture high-level spatial data in individual frames, whereas the "fast" path uses shallower, faster CNNs to capture low-level motion data. These two paths are combined together to form a comprehensive model that takes under consideration both the fine-grained motion subtle elements and high-level spatial setting show within the video data [5]. Spatial data in images is processed slowly. Runs at a reduced frame rate compared to the fast path. It comprises of a few convolutional layers (regularly 2D convolutional layers) that are utilized to capture high-level spatial data in individual data frames. The slow path is designed to capture not only inactive content from video, but too high-level scene context. The temporal information of the video is processed quickly. It captures the fine-grained motion information since it runs at a higher frame rate. The fast pathway, like the slow pathway, features convolutional layers, but it has fewer layers. It captures low-level temporal features in video, such as the motion of objects. The output features from the slow and fast paths are merged to form a complete representation of the video. This fusion can be performed in a variety of ways, including concatenation or weighted combination, allowing the model to include information both spatial and temporal at the same time [5].

3. Results and Discussion

3.1. Dataset

The paper highlights the implementation of the proposed framework for the UCF Crime dataset [3], which consists of violent, criminal, and unusual behaviours captured by video surveillance cameras placed in public areas such as schools, roads, general stores, etc. [1][2]. The selected dataset has been chosen because it was gathered from real-life events that may occur anywhere, at any time. This dataset is made up of video clips that were taken by different surveillance cameras with the intention of detecting abnormal or illegal activities in real-world situations. Videos are taken in both indoor and outdoor environments using various surveillance cameras 1900 video clips are included in the dataset [3]. Furthermore, to the normal events class, these UCF Crime dataset contains lengthy recordings surveillance data feeds covering 13 distinctive classes of anomaly events, including abuse, arrest, arson, ambush, Road accident, Burglaries, Explosions, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism [3]. The training and testing information are orchestrated in experiments in which 75% and 25% are utilized for comparison with other comparable studies. The number of videos from the dataset for each class is given in Table 1.

Table 1. Number of videos for UCF Crime datasets.

Anomalies	Videos	Anomalies	Videos
Abuse	50	Road Accident	50
Arrest	50	Robbery	50
Arson	50	Shooting	50
Assault	50	Shoplifting	50
Burglary	50	Stealing	50
Explosion	50	Vandalism	50
Fighting	50	Normal	50
Total	350	Total	350

3.2. Model Setting

The proposed approach is utilized in studies by VGG16 (CNN, Bi-LSTM), Slow Fast Network, and DenseNet121, which are all accessible within the PyTorch and Keras libraries. A few hyperparameters are utilized to fine-tune the model to attain its optimum execution. The outcomes of experiments using various weight initialization and optimizers are displayed in Table 2. As a result, weights are used to initialize the model while imposing them to optimise it. There's a fixed learning rate of 0.0001 and a settled number of epochs. Various forms of evaluation are employed in experiments. When performing a comparison-based analysis with another CNN [1][3] model from the PyTorch [10] and Keras [2] libraries. Table 2 presents an accuracy comparison among PyTorch, Tensorflow and Keras platforms that were used for simulations.

Table 2. Hyperparameter setting and CNN model Comparison.

Model name	Hyperparameter	Tuning	Epoch	Accuracy
VGG16 CNN	Weight	imagenet	10	98%
	Optimizer	SGD		
VGG16 Bi-LSTM	Weight	imagenet	15	96%
	Optimizer	Adam		
DenseNet121	Weight	imagenet	1	70%
	Optimizer	Adam		
Slow Fast Network	Optimizer	SGD	20	99%

4. Results

Firstly, the proposed is the Comparison-based demonstrate, which is evaluated by calculating its Precision, Accuracy, Recall, and F1 scores. The Accuracy[1], Precision[1], Recall[1], and F1[1] scores for VGG16 CNN+BiLSTM, Slow Fast Network, and DenseNet121 are shown in Table 3. In the below Table 3 the proposed model Slow Fast Network has achieved significant evaluation scores from the UCF-Crime dataset. The most significant accuracy was of Slow Fast Network and VGG16 models and the DenseNet121 gives the AUC score of 70% for the UCF - Crime dataset as compared to another proposed model.

Table 3. Results Model Evaluation of Proposed Methodology

Model Name	Dataset	CLASS	Accuracy	Precision	F1	Recall
VGG16 CNN	UCF CRIME	Abnormal	98%	0.98	0.98	0.99
		Normal		0.99	0.98	0.98
VGG16 BI-LSTM	UCF CRIME	Abnormal	96%	0.96	0.96	0.95
		Normal		0.95	0.96	0.96
Slow Fast Network	UCF CRIME	Abnormal	99%	0.98	0.99	1.00
		Normal		1.00	0.99	0.98
DenseNet121	UCF CRIME	Abnormal and Normal	70%	0.70	0.70	0.71

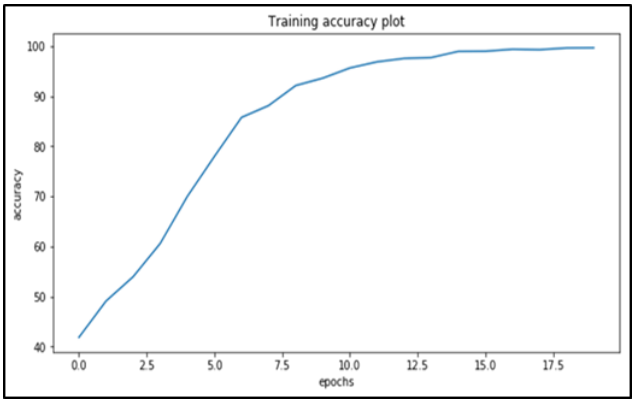

Fig.5. Training Accuracy of Slow Fast Network.

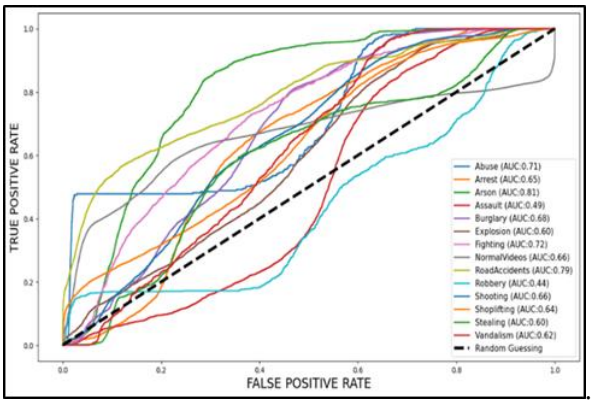

Fig 6. ROC AUC Accuracy for Densenet121.

4.1. Comparison with the existing Models

In this study, the proposed models are assessed for accuracy and compared to other deep learning models. There's constrained research on the detection of anomalies utilized for the UCF-Crime dataset. Table 4 shows the accuracy scores for proposed models and alternative deep learning models for abnormality detection on the UCF Crime dataset. Among the related models are ResNet50 and ConvLSTM [1], ResNet18/34/50 + SRU [2], 2D-CNN and ESN [11], CNN [12], CNN and LSTM [13], ConvGRU-CNN [14]. Anomaly categories are those that include all of the previously described normal events, while normal data is defined as not having any abnormal events. The probability that anomaly events will be appropriately classified is shown by the results of the test classifier. In terms of anomaly detection, Table 4 demonstrates that the proposed VGG16 and Slow Fast Network models performed better than the relevant benchmark models.

Table 4. Models Comparison in terms of accuracy

Reference	*Model*	*Accuracy*
[1]	ResNet50 and ConvLSTM	81.71%
[2]	ResNet18+ SRU	89.08%
	ResNet34 +SRU	90.09%
	ResNet50 + SRU	91.64%
[4]	2D-CNN and ESN	87.55
[6]	CNN	82%
[7]	CNN and LSTM	90.6%
[8]	ConvGRU-CNN	82.22%
Proposed	**VGG16 CNN**	**98%**
Proposed	**VGG16 BiLSTM**	**96%**
Proposed	**Slow Fast Network**	**99%**
Proposed	DenseNet121	70%

5. Conclusion

The present article proposes a unique structure for detecting anomalous behaviour in the UCF-Crime dataset by comparison-based deep learning model i.e. VGG16, Slow Fast Network, DenseNet121. It found several obstacles when putting this idea into practice. We used a dataset that varied in individuals, speeds, and brightness. For example, in certain videos, anomalies occurred, while in others, there were no people seen at all (e.g., car accidents). In addition, we have additional limitations on our dataset to address. The abnormal events might happen in a matter of seconds, and in videos running ten seconds or less, more than eighty percent of the footage demonstrates typical activity. In the UCF Crime dataset, the proposed approach performed better than previous approaches despite all the previously mentioned limitations. It diminishes the original footage of three diverse abnormal events in addition to utilizing all 14 categories of UCF Crime and separates them into two primary categories. With the same background and objects, we have both abnormal and normal incidents. Extricating the foremost vital information from each input video frame, we utilized one of the foremost prevalent VGG16, Slow Fast Network, DenseNet121. Finally, we used a comparison-based proposed model to classify for each video data to determine how well the model classifies the correct category of each input video. In spite of the fact that the experimental findings show that our technique outperforms other existing models, we proposed to improve this classification of all 14 types of abnormalities within the UCF Crime dataset.

Ethics of Statement

I have taken the dataset from the open-source repository, and it is freely available to use.

Credit author statement.

Anchal Pathak : Conceptualization, Implementation, Manuscript writing, Ruchi Jayaswal: Investigation, Supervision, Reviewing, Smita Mahajan :Investigation, Supervision, Reviewing.

Acknowledgements

The authors would like to thank Symbiosis Institute of Technology for providing us platform to do research and also thank to all the co-authors who supported this work.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Vosta, Soheil, and Kin-Choong Yow. 2022. "A CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras" Applied Sciences 12, no. 3: 1021. https://doi.org/10.3390/app12031021.

[2] Maryam Qasim, Elena Verdu, Video anomaly detection system using deep convolutional and recurrent models, Results in Engineering, Volume 18,2023,101026, ISSN2590-1230. https://doi.org/10.1016/j.rineng.2023.101026.

[3] https://www.kaggle.com/datasets/mission-ai/crimeucfdataset.

[4] Koklu M, Cinar I, Taspinar YS. CNN-based bi-directional and directional long-short term memory network for determination of face mask. Biomed Signal Process Control. 2022 Jan; 71:103216. doi: 10.1016/j.bspc.2021.103216. Epub 2021 Oct 9. PMID: 34697552; PMCID: PMC8527867.

[5] https://arxiv.org/pdf/1812.03982.pdf

[6] https://iq.opengenus.org/architecture-of-densenet121/

[7] https://www.researchgate.net/publication/361386340_Prediction_of_Battery_SOH_by_CNNBiLSTM_Network_Fused_with_Attention_Mechanism/citations.

[8] R, C. C K and S. Chaudhari, "Comparative study of CNN, VGG16 with LSTM and VGG16 with Bidirectional LSTM using kitchen activity dataset," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2021, pp. 836-843, doi: 10.1109/I-SMAC52330.2021.9640728.

[9] Halder, R., Chatterjee, R. CNN-BiLSTM Model for Violence Detection in Smart Surveillance. SN COMPUT. SCI. 1, 201 (2020). https://doi.org/10.1007/s42979-020-00207-x

[10] https://arxiv.org/abs/1912.01703

[11] Islam, Muhammad, Abdulsalam S. Dukyil, Saleh Alyahya, and Shabana Habib. 2023. "An IoT Enable Anomaly Detection System for Smart City Surveillance" Sensors 23, no. 4: 2358. https://doi.org/10.3390/s23042358

[12] Khan, Sardar Waqar, Qasim Hafeez, Muhammad Irfan Khalid, Roobaea Alroobaea, Saddam Hussain, Jawaid Iqbal, Jasem Almotiri, and Syed Sajid Ullah. 2022. "Anomaly Detection in Traffic Surveillance Videos Using Deep Learning" Sensors 22, no. 17: 6563. https://doi.org/10.3390/s22176563

[13] Kiprijanovska, Ivana, Hristijan Gjoreski, and Matjaž Gams. 2020. "Detection of Gait Abnormalities for Fall Risk Assessment Using Wrist-Worn Inertial Sensors and Deep Learning" Sensors 20, no. 18: 5373. https://doi.org/10.3390/s20185373

[14] https://reunir.unir.net/bitstream/handle/123456789/14812/ip2023_05_006_0.pdf?sequence=1

[15] Sabokrou, M.; Fathy, M.; Hoseini, M. Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder. Electron. Lett. 2016, 52, 1122–1124.

[16] Yu, J.; Yow, K.C.; Jeon, M. Joint representation learning of appearance and motion for abnormal event detection. Mach. Vision Appl

[17] S. Narynov, Z. Zhumanov, A. Gumar, M. Khassanova and B. Omarov, "Physical Violence Detection in Video Streaming Using Partitioned Skeleton Analysis," 2021 21st International Conference on Control, Automation and Systems (ICCAS), Jeju, Korea, Republic of, 2021, pp. 225-230, doi: 10.23919/ICCAS52745.2021.9649827

[18] M. Cristani, R. Raghavendra, A. Del Bue, and V. Murino, "Human behavior analysis in video surveillance: A social signal processing perspective," Neurocomputing, vol. 100, pp. 86-97, 2013.

[19] Fernando J. Rendón-Segador, Juan A. Álvarez-García, Jose L. Salazar-González, Tatiana Tommasi, CrimeNet: Neural Structured Learning using Vision Transformer for violence detection, Neural Networks, Volume 161,2023, Pages 318-329, ISSN 0893-6080,

[20] https://doi.org/10.1016/j.neunet.2023.01.048.

[21] https://link.springer.com/chapter/10.1007/978-3-642-23678-5_39

[22] Ravindran, V.; Viswanathan, L.; Rangaswamy, S. A novel approach to automatic road-accident detection using machine vision techniques. Int. J. Adv. Comput. Sci. 2016, 7, 235–242.

[23] Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J. Big Data 2021, 8, 53.

[24] Gorokhov, O.; Petrovskiy, M.; Mashechkin, I. Convolutional neural networks for unsupervised anomaly detection in text data. In Proceedings of the 18th International Conference on Intelligent Data Engineering and Automated Learning, Guilin, China, 30 October–1 November 2017; Springer: Cham, Switzerland, 2017; pp. 500–507

[25] Hasan, M.; Choi, J.; Neumann, J.; Roy-Chowdhury, A.K.; Davis, L.S. Learning temporal regularity in video sequences. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 733–742.

[26] Yun, K.; Yoo, Y.; Choi, J.Y. Motion interaction field for detection of abnormal interactions. Mach. Vis. Appl. 2017, 28, 157–171

[27] Ebrahimi Kahou, S.; Michalski, V.; Konda, K.; Memisevic, R.; Pal, C. Recurrent neural networks for emotion recognition in video. In Proceedings of the 2015 17th ACM on International Conference on Multimodal Interaction, Seattle, WA, USA, 9–13 November 2015; pp. 467–474.

[28] Zhou, Fangrong, Gang Wen, Yi Ma, Hao Geng, Ran Huang, Ling Pei, Wenxian Yu, Lei Chu, and Robert Qiu. 2022. "A Comprehensive Survey for Deep-Learning-Based Abnormality Detection in Smart Grids with Multimodal Image Data" Applied Sciences 12, no. 11: 5336. https://doi.org/10.3390/app12115336

[29] Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. ACM Comput. Surv. (CSUR) 2009, 41, 1–58.

[30] Nor, Ahmad Kamal Mohd, Srinivasa Rao Pedapati, Masdi Muhammad, and Víctor Leiva. 2022. "Abnormality Detection and Failure Prediction Using Explainable Bayesian Deep Learning: Methodology and Case Study with Industrial Data" Mathematics 10, no. 4: 554. https://doi.org/10.3390/math10040554

[31] https://link.springer.com/chapter/10.1007/978-981-16-4538-9_33

[32] Qasim Gandapur, Maryam, and Elena Verdú. "ConvGRU-CNN: Spatiotemporal Deep Learning for Real-World Anomaly Detection in Video Surveillance System." International Journal of Interactive Multimedia and Artificial Intelligence (2023): n. pag.

[33] X. Wang, W. Xie and J. Song, "Learning Spatiotemporal Features With 3DCNN and ConvGRU for Video Anomaly Detection," 2018 14th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 2018, pp. 474-479, doi: 10.1109/ICSP.2018.8652354.

[34] https://www.ijimai.org/journal/bibcite/reference/3322

[35] Ullah, W.; Ullah, A.; Hussain, T.; Muhammad, K.; Heidari, A.A.; Del Ser, J.; Baik, S.W.; De Albuquerque, V.H.C. Artificial Intelligence of Things-assisted two-stream neural network for anomaly detection in surveillance Big Video Data. Future Gener. Comput. Syst. 2022, 129, 286–297.

[36] Wu, S.; Moore, B.E.; Shah, M. Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13-18 June 2010; pp. 2054–2060.

[37] Mohammadi, B.; Fathy, M.; Sabokrou, M. Image/video deep anomaly detection: A survey. arXiv

Prepr. 2021, arXiv:2103.01739.

[38] Albattah, W.; Habib, S.; Alsharekh, M.F.; Islam, M.; Albahli, S.; Dewi, D.A. An Overview of the Current Challenges, Trends, and Protocols in the Field of Vehicular Communication. Electronics 2022, 11, 3581.

[39] Li, W.; Mahadevan, V.; Vasconcelos, N. Anomaly detection and localization in crowded scenes. IEEE Trans. Pattern Anal. Mach. Intell. 2013, 36, 18–32.

[40] Sultani, W.; Chen, C.; Shah, M. Real-World Anomaly Detection in Surveillance Videos; IEEE: Piscataway, NJ, USA, 2018; pp. 6479–6488.

[41] Cheng, K.-W.; Chen, Y.-T.; Fang, W.-H. Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation. IEEE Trans. Image Process. 2015, 24, 5288–5301.

[42] https://www.researchgate.net/publication/356197963_Classification_of_Blood_Cells_from_Blood_Cell_Images_Using_Dense_Convolutional_Network.