
REFERENCES

1. Ahlawat, S., Choudhary, A., Nayyar, A., Singh, S., & Yoon, B. (2020). Improved handwritten digit recognition using convolutional neural networks (CNN). *Sensors*, *20*(12), 3344.
2. Al-Dhamari, A., Sudirman, R., & Mahmood, N. H. (2020). Transfer deep learning along with binary support vector machine for abnormal behavior detection. *IEEE Access*, *8*, 61085–61094.
3. AlMarri, S., Zaheer, M. Z., & Nandakumar, K. (2024). A multi-head approach with shuffled segments for weakly-supervised video anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 132–142).
4. Al-Nawashi, M., Al-Hazaimeh, O. M., & Saraee, M. (2017). A novel framework for intelligent surveillance systems based on abnormal human activity detection in academic environments. *Neural Computing and Applications*, *28*, 565–572.
5. Alshammari, A., & Rawat, D. B. (2019). Intelligent multi-camera video surveillance system for smart city applications. In *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0317–0323). IEEE.
6. 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. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*, 53.
7. Amin, S. U., Abbas, M. S., Kim, B., Jung, Y., & Seo, S. (2024). Enhanced anomaly detection in pandemic surveillance videos: An attention approach with EfficientNet-B0 and CBAM integration. *IEEE Access*, *12*, 162697–162712.
8. Amudha, L., & PushpaLakshmi, R. (2023). Performance analysis of hybrid RR algorithm for anomaly detection in streaming data. *Computer Systems Science and Engineering*, *45*(3), 2299–2312.
9. Asad, M., Jiang, H., Yang, J., Tu, E., & Malik, A. A. (2022). Multi-stream 3D latent feature clustering for abnormality detection in videos. *Applied Intelligence*, *52*(1), 1126–1143.

-
10. Baccouche, A., Garcia-Zapirain, B., Olea, C. C., & Elmaghraby, A. S. (2021). Breast lesions detection and classification via YOLO-based fusion models. *Computers, Materials & Continua*, 69(1), 1317–1335.
 11. Balasundaram, A., & Chellappan, C. (2020). An intelligent video analytics model for abnormal event detection in online surveillance video. *Journal of Real-Time Image Processing*, 17(4), 915–930.
 12. Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*. <https://arxiv.org/abs/2004.10934>
 13. Bouindour, S., Snoussi, H., Hittawe, M., Tazi, N., & Wang, T. (2019). An online and adaptive method for detecting abnormal events in videos using spatio-temporal ConvNet. *Applied Sciences*, 9(4), 757.
 14. Chen, D. Y., Peng, L., Li, W. C., & Wang, Y. D. (2021). Building extraction and number statistics in WUI areas based on UNet structure and ensemble learning. *Remote Sensing*, 13(6), 1172.
 15. Cherian, A. K., & Poovammal, E. (2021). Anomaly detection in real-time surveillance videos using deep learning. In S. Smys, J. M. R. S. Tavares, R. Bestak, & F. Shi (Eds.), *Computational vision and bio-inspired computing* (Vol. 1318, pp. 217–226). Springer, Singapore.
 16. Cucchiara, R., Grana, C., Prati, A., & Vezzani, R. (2005). Computer vision system for in-house video surveillance. *IEEE Proceedings - Vision, Image and Signal Processing*, 152(2), 242–249.
 17. Datta, K., Hossain, I., Choi, S., Saletore, V., Ambert, K., Godinez, W. J., & Zhang, X. (2019). Training multiscale-CNN for large microscopy image classification in one hour. In *International Conference on High Performance Computing* (pp. 463–477). Cham: Springer International Publishing.

18. Dhevanandhini, G., & Yamuna, G. (2024). An optimal intelligent video surveillance system in object detection using hybrid deep learning techniques. *Multimedia Tools and Applications*, 83, 44299–44332.
19. Dilek, E., & Dener, M. (2024). Enhancement of video anomaly detection performance using transfer learning and fine-tuning. *IEEE Access*, 12, 73304–73322.
20. Ding, L., & Goshtasby, A. (2001). On the Canny edge detector. *Pattern Recognition*, 34(3), 721–725.
21. Direkoglu, C. (2020). Abnormal crowd behavior detection using motion information images and convolutional neural networks. *IEEE Access*, 8, 80408–80417.
22. Dos Santos, J. C. M., Carrijo, G. A., de Fátima dos Santos Cardoso, C., & et al. (2020). Fundus image quality enhancement for blood vessel detection via a neural network using CLAHE and Wiener filter. *Research on Biomedical Engineering*, 36(1), 107–119.
23. Doshi, K., & Yilmaz, Y. (2020). Any-shot sequential anomaly detection in surveillance videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 934–935).
24. Dyke, R. M., & Hormann, K. (2023). Histogram equalization using a selective filter. *The Visual Computer*, 39, 6221–6235.
25. Fedorov, A., Nikolskaia, K., & Ivanov, S. (2019). Traffic flow estimation with data from a video surveillance camera. *Journal of Big Data*, 6(73).
26. Fanta, H., Shao, Z., & Ma, L. (2020). SiTGRU: Single-tunneled gated recurrent unit for abnormality detection. *Information Sciences*, 524, 15–32.
27. Feng, J., Liang, Y., & Li, L. (2021). Anomaly detection in videos using two-stream autoencoder with post hoc interpretability. *Computational Intelligence and Neuroscience*, 2021(1), 7367870.
28. Fu, Y., Yang, B., & Ye, O. (2024). Spatiotemporal masked autoencoder with multi-memory and skip connections for video anomaly detection. *Electronics*, 13(2), 353.

29. Ganagavalli, K., & Santhi, V. (2024). YOLO-based anomaly activity detection system for human behavior analysis and crime mitigation. *Signal, Image and Video Processing*, 1–11.
30. Ganokratanaa, T., Aramvith, S., & Sebe, N. (2022). Video anomaly detection using deep residual-spatiotemporal translation network. *Pattern Recognition Letters*, 155, 143–150.
31. Ghahremannezhad, H., Liu, C., & Shi, H. (2022). Ammunition component classification using deep learning. In *Proceedings of the International Conference on Machine Learning and Data Mining (MLDM 2022)*, New York, USA.
32. Gong, D., Liu, L., Le, V., Saha, B., Mansour, M. R., Venkatesh, S., & Van Den Hengel, A. (2019). Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1705–1714).
33. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
34. Guo, B., Liu, M., He, Q., & Jiang, M. (2024). Two-stream spatial-temporal auto-encoder with adversarial training for video anomaly detection. *IEEE Access*, 12, 125881–125889.
35. Gupta, A. K., Sharma, R., & Ojha, R. P. (2024). Video anomaly detection based on HSOE-FAST modified deep neural network. *SN Computer Science*, 5(588).
36. Hafeez, M. A., Javed, S., Madden, M., & Ullah, I. (2023). Unsupervised end-to-end transformer-based approach for video anomaly detection. In *Proceedings of the 38th International Conference on Image and Vision Computing New Zealand (IVCNZ)* (pp. 1–7).
37. Hao, Y., Li, J., Wang, N., Wang, X., & Gao, X. (2022). Spatiotemporal consistency-enhanced network for video anomaly detection. *Pattern Recognition*, 121, 108232.
38. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778).

39. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
40. Huang, C., Wu, Z., Wen, J., Xu, Y., Jiang, Q., & Wang, Y. (2022). Abnormal event detection using deep contrastive learning for an intelligent video surveillance system. *IEEE Transactions on Industrial Informatics*.
41. Huang, H., Zeng, Z., Yao, D., Pei, X., & Zhang, Y. (2022). Spatial-temporal ConvLSTM for vehicle driving intention prediction. *Tsinghua Science and Technology*, 27(3), 599–609.
42. Janaki, R., & Lakshmi, D. (2024). Hybrid model-based early diagnosis of esophageal disorders using convolutional neural network and refined logistic regression. *Journal of Image and Video Processing*, 2024(1), 1–27.
43. Jayachandra, N., & Kamal, A. R. N. B. (2019). Real-time marine video image preprocessing technique. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 6(6), 463–469.
44. Khaire, P., & Kumar, P. (2022). A semi-supervised deep learning-based video anomaly detection framework using RGB-D for surveillance of real-world critical environments. *Forensic Science International: Digital Investigation*, 40, 301347.
45. Khan, A., Khan, S., Hassan, B., & Zheng, Z. (2022). CNN-based smoker classification and detection in smart city application. *Sensors*, 22(3), 892.
46. Khan, S. W., Hafeez, Q., Khalid, M. I., Alroobaea, R., Hussain, S., Iqbal, J., & Ullah, S. S. (2022). Anomaly detection in traffic surveillance videos using deep learning. *Sensors*, 22(17), 6563.
47. Kim, Y., Yu, J. Y., Lee, E., & Kim, Y. G. (2022). Video anomaly detection using Cross U-Net and cascade sliding window. *Journal of King Saud University-Computer and Information Sciences*, 34(6), 3273–3284.
48. Kingma, D. P., & Ba, J. L. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

49. Kommanduri, R., & Ghorai, M. (2023). A supervised approach for efficient video anomaly detection using transfer learning. In P. Maji, T. Huang, N. R. Pal, S. Chaudhury, & R. K. De (Eds.), *Pattern recognition and machine intelligence. PReMI 2023. Lecture Notes in Computer Science* (Vol. 14301). Springer, Cham.
50. Kotkar, V. A., & Sucharita, V. (2023). Fast anomaly detection in video surveillance system using robust spatiotemporal and deep learning methods. *Multimedia Tools and Applications*.
51. Kumar, M., Patel, A. K., & Biswas, M. (2024). Real-time detection of abnormal human activity using deep learning and temporal attention mechanism in video surveillance. *Multimedia Tools and Applications*, 83(1), 55981–55997.
52. Lee, Y.-C., Lee, S.-Y., Kim, B., & Kim, D.-Y. (2024). GLBRF: Group-based lightweight human behavior recognition framework in video camera. *Applied Sciences*, 14(6), 2424.
53. Li, C., & Chen, M. (2024). Dy-MIL: Dynamic multiple-instance learning framework for video anomaly detection. *Multimedia Systems*, 30(11).
54. Li, T., Chen, X., Zhu, F., Zhang, Z., & Yan, H. (2021). Two-stream deep spatial-temporal auto-encoder for surveillance video abnormal event detection. *Neurocomputing*, 439, 256–270.
55. Liu, Q., Xiao, L., Yang, J., & Wei, Z. (2021). CNN-enhanced graph convolutional network with pixel- and super pixel-level feature fusion for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(10), 8657–8671.
56. Liu, T., Zhang, C., Niu, X., & Wang, L. (2022). Spatio-temporal prediction and reconstruction network for video anomaly detection. *PLOS One*, 17(5), e0265564.
57. Liu, W., Pan, Y., & Fan, Y. (2024). A reactive deep learning-based model for quality assessment in airport video surveillance systems. *Electronics*, 13(4), 749.

58. Liu, Y., Liu, J., Zhao, M., Li, S., & Song, L. (2022). Collaborative normality learning framework for weakly supervised video anomaly detection. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 69(5), 2508–2512.
59. Liu, Z., Wu, X.-M., Zheng, D., Lin, K.-Y., & Zheng, W.-S. (2023). Generating anomalies for video anomaly detection with prompt-based feature mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 24500–24510).
60. Liu, Z. M., Fu, H., Ma, H., & Chua, T. S. (2020). Enhancing anomaly detection in surveillance videos with transfer learning from action recognition. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 4664–4668).
61. Lv, H., & Sun, Q. (2024). Video anomaly detection and explanation via large language models. *arXiv preprint arXiv:2401.05702*.
62. Ma, Q. (2021). Abnormal event detection in videos based on deep neural networks. *Scientific Programming*.
63. Majhi, S., Dai, R., Kong, Q., Garattoni, L., Francesca, G., & Bremond, F. (2024). Human-scene network: A novel baseline with self-rectifying loss for weakly supervised video anomaly detection. *Computer Vision and Image Understanding*, 241, 103955.
64. Mandal, B., Okeukwu, A., & Theis, Y. (2021). Masked face recognition using ResNet-50. *arXiv preprint arXiv:2104.08997*.
65. Mei, X., Wei, Y., & Chen, H. (2024). Graph-based domain adversarial learning framework for video anomaly detection domain generalization. *Journal of Supercomputing*, 80, 18977–19002.
66. Mishra, S., & Jabin, S. (2024). Anomaly detection in surveillance videos using deep autoencoder. *International Journal of Information Technology*, 16(1), 1111–1122.
67. Monakhov, V., Thambawita, V., Halvorsen, P., & Riegler, M. A. (2023). GridHTM: Grid-based hierarchical temporal memory for anomaly detection in videos. *Sensors*, 23(4), 2087.

68. Mumcu, F., Doshi, K., & Yilmaz, Y. (2022). Adversarial machine learning attacks against video anomaly detection systems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 206–213).
69. Munteanu, D., Bejan, C., Munteanu, N., Zamfir, C., Vasić, M., Petrea, S.-M., & Cristea, D. (2022). Deep-learning-based system for assisting people with Alzheimer’s disease. *Electronics*, *11*(19), 3229.
70. Nasaruddin, N., Muchtar, K., Afdhal, A., & Dwiyantoro, A. P. J. (2020). Deep anomaly detection through visual attention in surveillance videos. *Journal of Big Data*, *7*(1), 87.
71. Nawaratne, R., Alahakoon, D., De Silva, D., & Yu, X. (2020). Spatiotemporal anomaly detection using deep learning for real-time video surveillance. *IEEE Transactions on Industrial Informatics*, *16*(1), 393–402.
72. Nie, Y., Huang, H., Long, C., Zhang, Q., Maji, P., & Cai, H. (2024). Interleaving one-class and weakly-supervised models with adaptive thresholding for unsupervised video anomaly detection. *arXiv preprint arXiv:2401.13551*.
73. Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359.
74. Pelvan, S. Ö., Can, B., & Ozkan, H. (2023). A hierarchical approach for improved anomaly detection in video surveillance. *IEEE Access*, *11*, 101644–101665.
75. Perera, P., & Patel, V. M. (2019). Learning deep features for one-class classification. *IEEE Transactions on Image Processing*, *28*(11), 5450–5463.
76. Qadeer, K., Rehman, W. U., Sheri, A. M., Park, I., Kim, H. K., & Jeon, M. (2020). A long short-term memory (LSTM) network for hourly estimation of PM2.5 concentration in two cities of South Korea. *Applied Sciences*, *10*(11), 3984.
77. Ragedhaksha, Darshini, Shahil, & Arunnehr, J. (2023). Deep learning-based real-world object detection and improved anomaly detection for surveillance videos. *Materials Today: Proceedings*, *80*(Part 3), 2911–2916.

78. Ramadan, M. K., Youssif, A. A. A., & El-Behaidy, W. H. (2022). Detection and classification of human-carrying baggage using DenseNet-161 and Fit One Cycle. *Big Data and Cognitive Computing*, 6(4), 108.
79. Rendón-Segador, F. J., Álvarez-García, J. A., Enríquez, F., & Deniz, O. (2021). ViolenceNet: Dense multi-head self-attention with bidirectional convolutional LSTM for detecting violence. *Electronics*, 10(13), 1601.
80. Sabeena, B., Sivakumari, S., & Teresa, D. M. (2022). Optimization-based ensemble feature selection algorithm and deep learning classifier for Parkinson's disease. *Journal of Healthcare Engineering*, 2022(7), 1–12.
81. Scherer, D., Muller, A., & Behnke, S. (2010). Evaluation of pooling operations in convolutional architectures for object recognition. In *Proceedings of the International Conference on Artificial Neural Networks*, Thessaloniki, Greece, 15–18.
82. Shin, J., Kaneko, Y., Miah, A. S. M., Hassan, N., & Nishimura, S. (2024). Anomaly detection in weakly supervised videos using multistage graphs and general deep learning-based spatial-temporal feature enhancement. *IEEE Access*.
83. Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28.
84. Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-Net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, 9, 82031–82057.
85. Singh, V., Singh, S., & Gupta, P. (2020). Real-time anomaly recognition through CCTV using neural networks. *Procedia Computer Science*, 173, 254–263.
86. Solanki, S., Shah, Y., Rohit, D., & Ramoliya, D. (2023). Unveiling anomalies in surveillance videos through various transfer learning models. In *Proceedings of the 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)* (pp. 300–306).

87. Song, Q., Zheng, Y.-J., Sheng, W.-G., & Yang, J. (2021). Tridirectional transfer learning for predicting gastric cancer morbidity. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(2), 561–574.
88. Sufian, A., Ghosh, A., Sadiq, A., & Smarandache, F. (2020). A survey on deep transfer learning and edge computing for mitigating the COVID-19 pandemic. *Journal of Systems Architecture*, *108*.
89. Taghinezhad, N., & Yazdi, M. (2023). A new unsupervised video anomaly detection using multi-scale feature memorization and multipath temporal information prediction. *IEEE Access*, *11*, 9295–9310.
90. Thuan, D. (2021). Evolution of YOLO algorithm and YOLOv5: The state-of-the-art object detection algorithm.
91. Tran, T. M., Bui, D. C., Nguyen, T. V., & Nguyen, K. (2024). Transformer-based spatio-temporal unsupervised traffic anomaly detection in aerial videos. *IEEE Transactions on Circuits and Systems for Video Technology*, *34*(9), 8292–8309.
92. Tsakanikas, V., & Dagiuklas, T. (2018). Video surveillance systems—Current status and future trends. *Computers & Electrical Engineering*, *70*, 736–753.
93. Ul Amin, S., Ullah, M., Sajjad, M., Cheikh, F. A., Hijji, M., Hijji, A., & Muhammad, K. (2022). EADN: An efficient deep learning model for anomaly detection in videos. *Mathematics*, *10*(9), 1554.
94. Ullah, W., Ullah, A., Haq, I. U., Muhammad, K., Sajjad, M., & Baik, S. W. (2021). CNN features bi-directional LSTM for real-time anomaly detection in surveillance networks. *Multimedia Tools and Applications*, *80*, 16979–16994.
95. Vu, T., Pham, N., Vo, T., & Bui, A. (2022). Robust anomaly detection in surveillance videos using adversarial learning. *Pattern Recognition Letters*, *144*, 129–135.
96. Wang, J., Ji, G., & Zhao, B. (2024). Video anomaly detection using diverse motion-conditioned adversarial predictive network. *Neural Computing and Applications*, *36*, 18645–18659.

97. Wang, Z., Gu, X., Yan, H., et al. (2024). Domain generalization for video anomaly detection considering diverse anomaly types. *Signal, Image and Video Processing*, 18, 3691–3704.
98. Wu, C., Shao, S., Tunc, C., & Hariri, S. (2020). Video anomaly detection using pre-trained deep convolutional neural nets and context mining. In *Proceedings of the 2020 IEEE International Conference on Big Data (Big Data)* (pp. 3015–3022).
99. Wu, J.-C., Hsieh, H.-Y., Chen, D.-J., Fuh, C.-S., & Liu, T.-L. (2022). Self-supervised sparse representation for video anomaly detection. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIII* (pp. 729–745). Springer-Verlag.
100. Xu, M., Yu, X., Chen, D., Wu, C., & Jiang, Y. (2019). An efficient anomaly detection system for crowded scenes using variational autoencoders. *Applied Sciences*, 9(16), 3337.
101. Yu, J. H., Moon, J. H., & Sohn, K. A. (2023). Attention-guided residual frame learning for video anomaly detection. *Multimedia Tools and Applications*, 82, 12099–12116.
102. Zhang, G., Du, Z., Lu, W., & Meng, X. (2022). Dense pedestrian detection based on YOLO-V4 network reconstruction and CIoU loss optimization. In *Journal of Physics: Conference Series* (Vol. 2171, No. 1, p. 012019). IOP Publishing.
103. Zhang, L., Chen, Z.-B., Ma, X.-X., Shan, X.-Y., & Li, Z.-H. (2023). Research on video anomaly detection using cascaded memory-augmented autoencoder. *Comput. J.*, 34(5), 229–241.
104. Zhang, M., & Gunturk, B. K. (2008). Multiresolution bilateral filtering for image denoising. *IEEE Transactions on Image Processing*, 17(12), 2324–2333.
105. Zhang, Q., Wei, H., Chen, J., Du, X., & Yu, J. (2023). Video anomaly detection based on attention mechanism. *Symmetry*, 15(2), 528.

106. Zhang, Z. (2022). Detecting anomaly events in video based on generative adversarial network. *Computational Intelligence and Neuroscience*, 2022, 8633955.
107. Zhou, Y., Li, B., Wang, J., Rocco, E., & Meng, Q. (2022). Discovering unknowns: Context-enhanced anomaly detection for curiosity-driven autonomous underwater exploration. *Pattern Recognition*, 131, 108860.



Avinashilingam Institute for Home Science and Higher Education for Women

(Deemed to be University Estd. u/s 3 of UGC Act 1956, Category 'A' by MHRD

Re-accredited with A++ Grade by NAAC. CGPA 3.65/4, Category I by UGC

Coimbatore - 641 043, Tamil Nadu, India

Appendix L2

**(Item No 5 of
Check List) Details of Research
Publications**

S.No	Article	Journal	Other Details Vol/No/Page No/ Year	Published in UGC- CARE / Scopus Indexed/ Web of Science
1	Hybrid ResNet-50 and LSTM Approach for Effective Video Anomaly Detections in Intelligent Surveillance Systems	International Journal of Intelligent Systems and Applications in Engineering	11(4), 880-889, 2023	Scopus Indexed
2	Comprehensive survey on video anomaly detection using deep learning techniques.	International Journal of Computational Vision and Robotics	Vol 14 No 4 2024 https://doi.org/10.1504/ijcvr.2024.139544	Scopus Indexed

*Proof of list of Journals from Internet to be attached along with copies of reprints.

Scholar

Sreedevi R. Krishnan
16/7/24

Supervisor

Dr. P. Amudha
24/7/2024

The scholar Miss. Sreedevi R. Krishnan (Reg. No. 19PHECPO05) has published her research paper in the following journals:

1. International Journal of Intelligent Systems and Applications in Engineering - indexed and active in Scopus

Srinivasan Sargu
Checked By: *Sargu*
25/7/24

HoD/Dean of Respective School

2. Theoretical Journal of Computational Vision and Robotics -
indexed and at is in Scopus.

This may be considered.

J. Ali
16.07.2024

Hybrid ResNet-50 and LSTM Approach for Effective Video Anomaly Detection in Intelligent Surveillance Systems

Sreedevi R. Krishnan^{1*}, P. Amudha²

Submitted: 09/05/2023

Revised: 18/07/2023

Accepted: 07/08/2023

Abstract: Modern intelligent surveillance systems have given video anomaly detection much attention. Interior and outdoor monitoring devices are widespread in modern public places and smart cities. Due to the limited modelling capabilities and difficulty in capturing complicated relationships, conventional approaches have difficulty in recognizing video anomalies. This research tries to solve the problem using a hybrid strategy combining ResNet-50 (Residual Network-50) and Long Short-Term Memory (LSTM) algorithms for detecting anomaly video activity. Videos of normal and anomaly actions datasets enhanced the video clip's unique features, reducing the data required for storage and transmission. There are separate frames in each video. According to the proposed procedure, each frame is rescaled into a different pixel size before being fed into the ResNet-50 technique for feature extraction. Following feature extraction, the frames are then fed into the classification layer. The values of the new feature vectors are calculated by adding the original feature vectors acquired by ResNet-50. Finally, the LSTM model classifies the video as normal or abnormal using the information extracted from a sequence of frames. The LSTM assigns a classification to the retrieved images. The hybrid technique ResNet-50 and LSTM obtained 96.48% accuracy using the UCSD Ped 1 dataset. The proposed models outperformed the equivalent deep learning models and showed a noticeably higher performance accuracy.

Keywords: Anomaly detection, Deep Learning, LSTM, Rescale, ResNet-50, Video surveillance.

1. Introduction

In today's society, video surveillance data is quite relevant. Industrial facilities, educational institutions, and businesses are responsible for the extensive surveillance data availability. Similarly, cameras installed in public places like city centres, transportation hubs, and places of worship add to the public record. Several modules are used in the processing of surveillance video, including recognising objects and actions and classifying discovered behaviours into groups like abnormal or normal [1]. People can quickly lose their capacity to focus on the monitoring task since video data is growing, and only the security monitor staff can view the restricted amount of video data, which makes it difficult to spot suspicious behaviour on the screen [2]. Small spaces could only use manual surveillance systems, and security personnel were needed to watch for unusual activity. Manual surveillance, which is costly and requires extensive effort to detect suspicious activity, is not advisable [3]. An intelligent video surveillance system monitors critical locations particularly vulnerable to crime, such as ATMs, banking theft, fire detection, fight detection, etc. It can also play a critical role in security by identifying any

suspicious activity [4].

Anomaly detection, essential to autonomous video surveillance, determines whether the other categories described above are successful [5]. Additionally, it is a challenging assignment because even a human observer would need help. After all, the abnormalities to be found are unknown in advance. Finding behaviours that differ from what is expected and usual behaviour is a general description of anomaly detection [6]. Three significant issues impact anomaly detection in the context of autonomous video surveillance. These issues are due to the computationally expensive and challenging nature of video data processing, space and time complexity, and the non-local temporal variations between video frames [7].

The definition of abnormal behaviour depends heavily on the standards established in the environment under consideration [8]. It is a security challenge, meaning some peculiar behaviour has caused an activity outside the parameters [9]. More and more surveillance cameras are being used for video analysis due to the growing need for security. Automatic anomaly detection may be more practical than labour-intensive anomalous event identification [10]. Anomaly identification is challenging because it is difficult to list negative samples and obtain enough negative samples due to their rarity [11].

Due to the rising demand for security, surveillance cameras are being employed for video analysis more and more. Anomaly detection may be a more practical approach than

¹ Research Scholar, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India - 641043
ORCID ID: 0000-0002-2769-7747

² Professor, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India - 641043
ORCID ID : 0000-0001-7763-8633

* Corresponding Author Email: sreedevirkrishnan@gmail.com

labour-intensive abnormal event identification [12]. Popular techniques involve learning a model from recordings of normal events as training data and looking for abnormal events that deviate from the trained model. Deep learning (DL) and computer vision have been used recently to find the anomaly in the videos [13].

Deep learning has demonstrated its potential in various fields, including acoustics, pictures, and natural language processing. However, because abnormalities significantly differ from one another in various application contexts, it is challenging to create intelligent video anomaly detection systems [14]. Deep learning has shown increasingly astounding results in several previously believed to be computationally intractable tasks, such as face matching, recommendation systems, and anomaly detection. Video anomaly detection employs computer vision, which analyses footage to spot unusual activities or behaviours [15]. It employs advanced algorithms and techniques like motion analysis, object detection, recognition, and scene understanding. This approach enhances accuracy and timeliness in areas like surveillance, industrial monitoring, and critical event detection [16]. A deep learning algorithm's capacity to extract geographical and temporal elements from video footage automatically eliminates the necessity for manual feature engineering. This enables the models to detect anomalies effectively by capturing complex and dynamic video patterns. Also, the analysis of item appearances, motion patterns, and spatial connections in video data is used in computer vision-based video anomaly detection to find abnormalities that deviate from the normal [17]. Some deep learning models need to be more comprehensible and easier to understand. This can be a big problem, especially in critical applications where it is crucial to explain any anomalies found before making a decision [18].

The hybrid ResNet-50 and Long Short-Term Memory (LSTM) technique improves the detection of video anomalies. The combination of ResNet-50 and LSTM creates a resilient framework for video anomaly detection by combining the respective capabilities in spatial feature extraction and temporal modelling. The hybrid model delivers improved accuracy, robustness, and generalization by utilizing optical and temporal cues, making it suitable for various applications, including surveillance, industrial monitoring, and critical event detection.

The following are the primary contributions of the video anomaly detection method:

The area of video anomaly detection has made numerous vital strides because of deep learning and computer vision techniques.

➤ DL technique improves the accuracy of video anomaly detection by knowledge properties and representations from

raw data. In addition to recording intricate patterns for thorough analysis and improved discrimination, Convolutional Neural Networks (CNNs) are excellent at extracting spatial data from video frames. Real-time video anomaly detection is optimized by deep learning models using effective architectures, compression, and hardware acceleration.

➤ Achieve accurate accuracy in video anomaly detection by combining the ResNet-50 and LSTM methods. Deep architecture ResNet-50 strikes a compromise between performance and computational effectiveness. This feature extraction technique is appropriate for detecting real-time or almost real-time video anomalies.

➤ LSTM enables more precise anomaly identification and differentiation between normal and abnormal patterns in complex and dynamic settings. LSTM can be combined with feature extraction techniques like deep CNN to gather spatial and temporal information for more precise and reliable anomaly identification in videos.

In the remaining sessions of this document, the following is how they are organized: A description of previous research on the subject is provided in the second section. The third section provides a complete clarification of the research methodology. The fourth section investigates the simulation of the suggested approach and provides examples of the study's findings. The conclusion summarizes the findings.

2. Literature Review

The recent developments in video anomaly detection from 2019 to the present were reviewed and discussed in this session. The goal of video anomaly detection, which seeks to identify unusual portions in a video sequence, is to identify anomalous segments in a video sequence. A 3D CNN-based encoder and a 2D CNN-based decoder are used in a Hao et al. [19] suggested spatiotemporal consistency-enhanced network. The latent space vector is resampled when training with normal data, but this method may not perform as well when abnormal data is included. Combining a discriminator, a transformed video clip, and the model allows us to assess the input clip's consistency. The results are more spatiotemporally consistent when adversarial training is used. However, applying these networks to applications where spatiotemporal consistency is unimportant could add extra complexity and overhead.

Urbanisation and autonomous industrial environments have led to an increase in the demand for intelligent real-time video monitoring. Using known normalcy training datasets and tackling developing anomalous behaviours are two issues Artificial Intelligence (AI) faces for anomaly identification. Nawaratne et al. [20] proposed that real-time video surveillance can be improved using the Incremental Spatio-Temporal Learner (ISTL), which addresses video

anomaly detection. Experiments demonstrate that ISTL is suited for use in industrial and urban contexts due to its accuracy, resilience, low processing cost, and contextual indicators. Only human intervention is required for the learning model's verification and improvement.

Several challenges hinder the advancement of AI in video surveillance anomaly detection. These challenges stem from the details associated with adapting to evolving anomalous behaviours, the reliance on datasets predominantly consisting of normal instances, and the scarcity of comprehensive evaluation methods. As video monitoring expands, the significance of automated anomaly detection techniques becomes paramount in recognising unusual events. While achieving high accuracy through deep learning methods is a prominent goal, it is worth noting that many studies predominantly focus on accuracy alone. Kim et al. [21] developed a Cross U-Net paradigm that examines accuracy and speed. The structure employs a DL model and a cascade sliding window technique to determine the anomaly score of a frame.

The available computer vision and machine learning algorithms for processing images and videos vary, rely on libraries, and demand powerful hardware. Arunehru et al. [22] aim to create an AI-based system that processes video using a live CCTV camera feed while identifying and analysing occurrences. The system addresses important aspects like object detection, counting estimate, and anomaly detection using a convolutional neural network model while being simple to integrate and utilise with API calls. However, the Smart-Surveillance System is quite a powerful range of updates.

Studies train video anomaly detectors to identify underperforming frames during reconstruction and prediction tasks. Shao et al. [23] proposed a novel method for video anomaly identification utilising a generative network topology. The method learns spatial and temporal contexts through the order links between appearances and frames. According to experiments on four datasets, the technique enhanced detection performance on anomalies with various looks and motions. The approach's precision and recall were increased when it was combined with a prediction technique.

An Alex Net-based model for detecting crowd anomalies in video frames for images was suggested by Khan et al. [24]. The method consists of four convolutional layers, three fully connected layers, and a Rectified Linear Unit (ReLU) that acts as the activation function. The Convolution Layers (CLs) generate characteristics for anomaly detection. The model's Area Under the Curve (AUC) and overall accuracy were assessed on three reference datasets. The model achieved a 98% AUC utilising the Receiver Operating Characteristic Curve (ROC), outperforming baseline trials, according to the results. AlexNet is prone to overfitting like

many deep learning models, especially when trained on limited datasets.

The increased usage of cameras for surveillance and tracking abnormal human behaviour in public and private settings has made it more challenging to identify and decipher abnormal conduct in real-world contexts. Khan et al. [25] put forward the idea of employing CNNs to detect anomalies in traffic videos, leading to the development of a robust continuing prediction system with remarkable accuracy. The study's primary emphasis lies in the realm of traffic accidents. Through the analysis of traffic surveillance footage, the trained CNN model exhibited an impressive 82% accuracy in identifying instances of traffic accidents. This approach holds the potential to contribute to the reduction of accident rates and the improvement of safety protocols in urban environments.

The use of stationary cameras is frequently the focus of visual surveillance systems. Urban missions increasingly use Unmanned Aerial Vehicles (UAVs), which provide a fresh aerial perspective. The absence of a study on anomaly identification in drone videos presents difficulties for automatic anomaly detection by UAVs. One-Class Support Vector Machine (OCSVM), Histogram of Oriented Gradient- 3 Dimensional (HOG3D), and CNN are three feature extraction algorithms are used. Chriki et al. [26] novel approach to anomaly detection for UAV-based surveillance missions are suggested with an AUC of 0.78 at the worst and 0.93 at the best; experiments on a dataset substantiate the efficacy of the offered tactics. The proposed method's shortcoming is the failure to recognise some aberrant frames because of how similar they are to normal ones.

Massive amounts of video data are produced by surveillance systems, making it challenging for security experts to detect odd behaviour, Ul Amin et al. [27] suggested a DL model for anomaly detection in surveillance systems called EADN, which is just moderately complex. The technique combines a shot boundary detection technique to separate video into salient shots, a convolutional neural network to extract spatiotemporal information, and LSTM cells to detect anomalies. Studies using benchmark datasets and comparisons with cutting-edge techniques reveal a considerable improvement in the model's performance.

The amount of video data produced by surveillance systems makes analysis difficult for computer vision specialists. Anomaly detection with deep learning decreases the need for human labour while ensuring public safety. A practical methodology for deep features-based intelligent anomaly detection in surveillance networks is presented by Ullah et al. [28]. The framework captures spatiotemporal information from frames to reliably categorise anomalous/normal occurrences in intricate surveillance scenes. Subsequently, it uses a multilayer Bi-Directional

Long Short-Term Memory (BD-LSTM) model after sending them to a pre-trained CNN model. Studies reveal an improvement in accuracy of between 3.41% and 8.09% when compared to cutting-edge techniques.

Most of the currently available literature review focuses on the numerical difficulties of handling HD video data and finding contextual anomalies in surveillance video streams. To surmount the obstacles, present in video anomaly detection, the proposed approach presents a hybrid solution combining the ResNet-50 and LSTM techniques.

3. Methodology

Conventional techniques have issues in detecting video anomalies because of the limited modelling capacity and effort to extract nonlinear, complex correlations from video data. They have limited modelling capacity and find capturing complex and nonlinear relationships in video data difficult. As a result, they may fail to achieve the high-level semantics required for accurate anomaly detection. A hybrid approach combining ResNet-50 and LSTM methods as an effective solution for leveraging DL in video anomaly detection is proposed to address the abovementioned challenges. Here, characteristics from the video frame are extracted using a hybrid approach that combines ResNet-50 and LSTM approaches for video anomaly identification, and it demonstrated performance in image-based tasks, spatial and temporal context awareness, transfer learning abilities, robustness to changes, scalability, and computational efficiency. ResNet-50 is a viable option for extracting functional characteristics from video frames and enhancing the precision and efficacy of anomaly detection systems because of these benefits. Moreover, when used for classification in video anomaly detection, LSTM offers benefits in temporal dependency capture, sequential anomaly detection, memory retention, robustness to variations, contextual understanding, adaptability to variable-length sequences, and the capacity to use transfer learning. The overall performance of video anomaly detection systems is enhanced with the proposed model, which helps to accurately and effectively classify anomalies in video data.

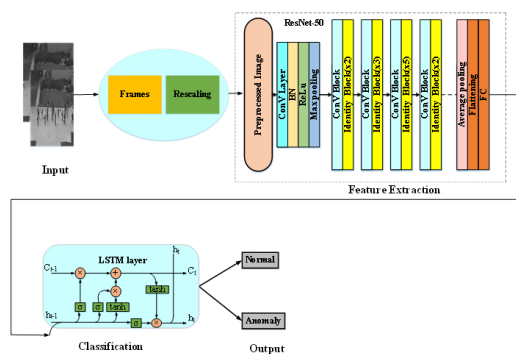


Fig. 1. Architecture diagram for the proposed model

The architecture of the suggested method, which finds video anomalies using LSTM classification, ResNet-50 feature extraction, and rescaling for preprocessing, is shown in Fig. 1. It comprises splitting the video into frames, resizing each frame, normalising pixel values, and utilising a pre-trained ResNet-50 technique. A sequence of feature vectors is created using the ResNet-50 model by extracting features from each frame. Sequential data's temporal dependencies are captured by an LSTM network, and an activation function creates a probability score for categorising video as normal or abnormal. The training and validation sets are built from the labelled video dataset, and the test set is used to assess the model. Following are the various stages of the suggested model.

Video input: The University of California, San Diego (UCSD) Ped1 dataset consists of video clips from UCSD security cameras.

Frame Extraction: After collecting video input, every video frame is resized to 256×256 pixels. The video is composed of multiple frames. The proposed approach is based on extending well-known greyscale textural features to sturdy, intense optical flow fields.

Preprocessing: Preprocessing is done using the rescaling method after the input is received, and it involves separating individual frames from the input video and resizing them to meet the input dimensions of the feature extraction method.

Feature extraction: Residual networks (ResNets) are effective neural models for feature extraction in Deep Neural Networks (DNN). They excel in extracting spatial features from input streams and have demonstrated outstanding performance on various widely used datasets. On many standard datasets, ResNets have displayed excellent performance. There are numerous ResNet variations, including ResNet-18, ResNet-26, ResNet-50, ResNet-101, and ResNet-152. Since videos may contain numerous redundant images, providing the complete video as input for preprocessing is not practical. Second, there are no time annotations available for videos of anomalies. It is helpful in these situations to extract significant features from videos and use them to train the model. To find an abnormality, the foundation of feature extraction is ResNet-50. The anomaly detector model receives the 1538 feature vector as input. The video is categorised as either class 0 (anomaly) or class 1 (normal). The network learns complex features with the aid of the activation function. ReLU is one of the most frequently utilised activation functions. In the feature extraction function, it is employed for video anomalies.

Classification: The suggested system's main objective is to enable visual anomaly detection. An LSTM was used in this detection to categorise a video frame. CNN is involved in the initialisation of the coarse and fine labels. Recurrent

Neural Network (RNN) is used to improve productive behaviour, and in the end, LSTM collaborates with RNN by using memory cells to give instructions to the learning model. In LSTM, internal memory cells are controlled by the input and forget gate network, developed to address the problem of vanishing or exploding gradients in RNN. The forget gate and the input gate, which alter the cell state, are ranked below the cell state. Choosing how much data from the previous memory should be transferred into the next time step is the primary responsibility of the forget gate in an LSTM. The output gate uses the tan h function to construct a vector, much how the input gate chooses how much new data should be fed into the memory cell. These gates enable the LSTM to manage the sequential information's short- and long-term dependencies. The following is the expression for the LSTM:

$$C_t^s = \tanh(\widehat{W}h_g * h_{t-1}^s + \widehat{W}_{xg}^s * x_t^s) \quad (1)$$

$$f_t^s = \sigma(\widehat{W}h_f * h_{t-1}^s + \widehat{W}_{xf}^s * x_t^s) \quad (2)$$

$$i_t^s = \sigma(\widehat{W}h_o * h_{t-1}^s + \widehat{W}_{xi}^s * x_t^s) \quad (3)$$

$$\bar{o}_t^s = \sigma(\widehat{W}h_i * h_{t-1}^s + \widehat{W}_{x\bar{o}}^s * x_t^s) \quad (4)$$

$$C_t^s = f_t^s \odot C_{t-1}^s + \bar{o}_t^s * x_t^s \quad (5)$$

$$h_t^s = \bar{o}_t^s \odot \tanh(C_t^s) \quad (6)$$

The weights \widehat{W}_{h*} throughout the parallel LSTM architecture, including the $\widehat{W}h_g, \widehat{W}h_f, \widehat{W}h_i,$ and $\widehat{W}h_o$ parameters, are used in the equations above to control and monitor the information about the hidden states from the previous time step in addition to $\widehat{W}_{xg}, \widehat{W}_{xf},$ and $\widehat{W}_{x\bar{o}}$, which are applied for the weight matrices for the current input time steps. Input sequence information is shown by the superscript s, time step information is shown by the subscript t, the sigmoid function is utilised σ , and elementwise multiplication is represented by the symbol \odot .

The concept of residual learning was introduced to train ultra-deep CNNs specifically for image recognition tasks. The residual idea is used to explain the sequential information of the top-level layers, and levels are reformulated by locating residual functions with the input layer. The RL function is typically expressed as follows:

$$y = f(\check{X}, \check{W}) + \check{X} \quad (7)$$

The layers' input and subsequent sequential information vectors are called \check{X} and y in Equation (7). The expression $f(\check{X}, \check{W})$ shows the residual knowledge from the connected layers $f(\check{X}, \check{W})$. Residual learning, which builds a sequence using the input given and nonlinear residual, contains the results of these layers. This method establishes a shortcut function across multiple layers for more effective model training, which is one of its main advantages. Additionally,

it aids in preventing the fundamental issue of vanishing gradients brought on by composition with the adaptive residual $f(\check{X}, \check{W})$.

In this study, the normalisation layer in a residual LSTM is used to reduce the dynamic hidden state, normalise the information of the neurons for the LSTM, and shorten the deep RNN training time.

$$\tilde{n}_t = \frac{1}{h} \sum_{i=0}^h (h_t) i \quad (8)$$

$$\delta_t = \sqrt{\frac{1}{h} \sum_{i=0}^h ((h_t)^i - \tilde{n}_t)^2} \quad (9)$$

$$\hat{y}_t = f\left(\frac{\hat{g}}{\delta_t} \odot (h_t - \tilde{n}_t) + b\right) \quad (10)$$

Where b is the trainable weights utilised to rescale the input sequence of the activation function, f, and $(h_t)i$ is the hidden state in each layer of the LSTM of the ith neuron, and the time step is denoted by the subscript. Before the forward connection of each layer of the LSTM, a dropout threshold of 0.5 was applied to lessen overfitting. The baseline study improved the model's effectiveness for captioning video using an encoder and decoder with attention techniques.

A decoder that inputs a video feature matching to the next word based on the words the model had before created was employed by them for word generation. That uses latent correlation between characteristics in different places to account for short-term and long-term dependence. Contrary to the video captioning model, which inputs a feature vector from the video frame sequence to the LSTM and employs a single block of features for sequence learning.

$$k_t = \frac{1}{h} \sum_{i=1}^h \widehat{W}_t^i f_i \quad (11)$$

$$\hat{S}_t = \widehat{W}^T \tan h(\widehat{W}_h h_t + M_h R_f + b_h) \quad (12)$$

$$A_t = \text{softmax}(S_t) \quad (13)$$

The parameters learned for the frame features f_i according to the attention weight \widehat{W}_t^i to return the score S_t . are $\widehat{W}^T, \widehat{W}_h, M_h, b_h$ in the equations. The output possibilities obtained from the Softmax activation layer are shown in A_t at the end. The Softmax layer is used to make the final predictions, and the deep characteristics collected are then used to identify whether the sequence consists of typical events or aberrant activity.

4. Results and Discussion

Thousands of video frames from video datasets were used to train the model, and comprehensive tests were used to assess the effectiveness of the new approach.

4.1. Datasets

The UCSD dataset comprises many video clips previously labelled by the creator as normal or with anomalies. The UCSD Pedestrian 1 (Ped 1) dataset contains 1307 training image frames and 231 testing ones. A fixed camera observing pedestrian walkways was used to gather the UCSD Anomaly Detection Dataset. Non-pedestrian creatures or strange pedestrian motion patterns caused abnormal dealings. Bicyclists, skateboarders, tiny carts, pedestrians crossing walkways, and wheelchair users were frequent outliers. Binary flags per frame and manually created pixel-level binary masks for Peds2 were included in the ground truth annotations, which made it possible to assess how well an algorithm can locate abnormalities.

4.2. Performance Evaluation

This section shows how effectively video anomaly image frames divided into normal or anomalous categories may be classified using the proposed hybrid ResNet-50 and LSTM methods. The training process about the 50 epochs displays the accuracy and loss analyses of the suggested model. The suggested model gains significantly more accuracy and loses utility. It shows an appreciably fast convergence of the suggested model. Fig. 2. (a) and (b) show the evolution of training in accuracy and loss analysis.

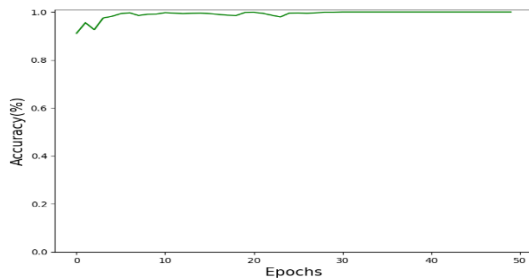


Fig. 2 (a). Training progress of accuracy analysis

The accuracy level of this training progression is 99.67%, which is very high and comparable to the 50th epoch.

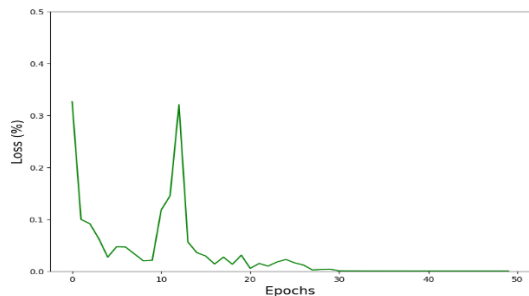


Fig. 2 (b). Training progress of loss analysis

In this training progress, the loss level is very high near the 50th epoch; the loss level of this process is 0.011%.

4.3. Performance Metrics

The following are the performance metrics: Accuracy, Precision, Specificity, Recall, F1-score, and AUC.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \quad (14)$$

The "True Positive Rate," called the Sensitivity or Recall criteria, is the percentage of positive events correctly classified as positive.

The Precision criteria display the percentage of examples from a specific class that were correctly identified compared to all cases from that class, regardless of whether they were correctly or wrongly classified.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (15)$$

The Specificity criteria, called the "True Negative Rate," is used when accurate negative detections are essential. It is the antithesis of the Sensitivity parameter. The percentage of incidents that were classified as negative is shown by this metric.

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (16)$$

A recall is a performance metric widely used in classification tasks to measure how well a model can pick out positive examples from a dataset.

True Positives (TP) refer to the count of events that the technique correctly classified as positive.

False Negatives (FN) is the percentage of illustrations that the model misclassified as positive when they were actually negative.

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (17)$$

The F1-Score criterion compromises between the Accuracy and Precision demands when False Positive and False Negative detections have different costs. This amount where displayed as

$$F1 - score = 2 \times \frac{FN}{TP+FN} \quad (18)$$

The AUC, particularly on the ROC curve, indicates the technique's capacity to distinguish between positive and negative examples; it measures the two-dimensional area beneath the curve. A higher AUC value denotes superior overall performance and discriminative capacity.

$$AUC = \sum \frac{(TPR[i]+TPR[i+1])}{2} * FPR[i + 1] - FPR[i] \quad (19)$$

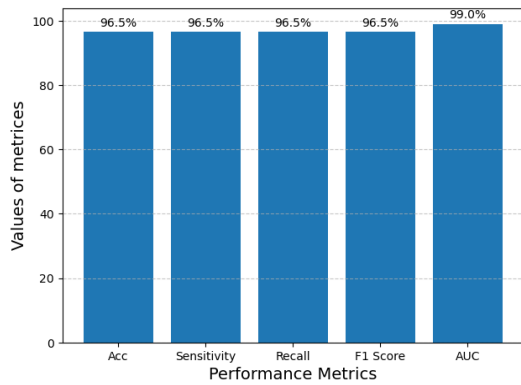


Fig. 3. Performance obtained from the proposed method

Fig. 3, which displays the performance metrics of a proposed video anomaly detection approach, a mix of ResNet-50 and LSTM methodologies, illustrates the overall effectiveness of video anomaly detection. Various performance criteria can be used to assess a hybrid model for video anomaly detection that combines ResNet-50 and LSTM. Typical measurements include Mean Average Precision (MAP), True Positive Rate (TPR), False Positive Rate (FPR), Precision, F1 Score, Accuracy, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These performance metrics aid in determining how well the hybrid ResNet-50 and LSTM model detects anomalies and prevents false alarms. Effective video anomaly detection is indicated by a high TPR, while minimising false alarms is done by a low FPR. Out of all the cases labelled as anomalies, precision determines the percentage of accurately detected anomalies. A higher F1 score indicates better overall performance. AUC-ROC assesses the model's capacity to distinguish between anomalies and regular events at various threshold settings. The average precision across various recall levels is then determined using Mean Average Precision (mAP). Combining these indicators enables a thorough assessment of the model's effectiveness in identifying abnormalities and preventing false alarms.

4.4. Results Obtained

The graph below displays typical (anomalous-free) and abnormal video frames. Fig. 4 interprets images from the video that are normal and abnormal.



Fig. 4. Video Anomaly Detection Normal image frames

Rescaling, which modifies the scale or range of pixel values in video frames, is an essential preprocessing step in detecting video anomalies. The data must be appropriate for the anomaly detection model to learn correctly, which is ensured in this stage. Techniques like normalisation or standardisation can be used to rescale. By dividing each pixel value by the highest value present in the video frames, normalisation modifies each pixel value to fall within a predetermined range, often between 0 and 1. The data is scaled to a similar range, and the mean is removed during standardisation, resulting in pixel values with a mean of 0 and a standard deviation of 1.

The properties of the video data and the precise needs of the anomaly detection task will determine whether to normalise or standardise. While standardisation is ideal for pixel values with a more extensive distribution, normalisation is appropriate for those with a specified maximum range. Rescaling preprocessing must be applied to prevent biases and provide consistency between training and test data. Based on the requirements of the anomaly detection model and the characteristics of the video data, additional preprocessing operations like scaling, cropping, or data augmentation should also be considered.

To extract valuable representations from video frames and apply them in identifying video anomalies, ResNet-50 is a deep convolutional neural network architecture. The procedure entails preprocessing the frames, loading the already trained ResNet-50 model, eliminating the classifier layers, and turning each frame into a collection of high-level characteristics known as feature vectors or embeddings. Commonly used to analyse temporal patterns, temporal aggregation produces features that can be used for anomaly detection. To aggregate temporal data, a variety of methods can be utilised, including averaging, pooling, and the usage of recurrent neural networks. By training a separate anomaly detection model or by employing unsupervised methods like clustering or density estimation, the generated feature vectors can be used for anomaly identification. This method captures rich representations of video frames using ResNet-50's potent feature extraction capabilities, assisting in detecting anomalies based on recognised visual patterns. The images of video anomaly detection after using the suggested strategies are shown in Fig. 5.

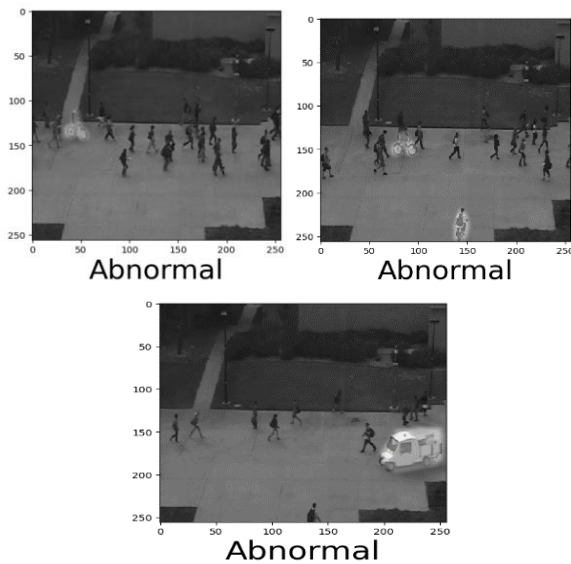


Fig. 5. Video Anomaly Detection Abnormal image frame

4.5. Comparative analysis

This section demonstrates how the proposed methods outperform other models in the hybrid ResNet-50 and LSTM method, allowing the proposed to serve as the method's primary framework. Compared to other networks with fewer parameters, a hybrid of ResNet-50 and LSTM can offer performance that is equivalent to or even better. They contrasted the outcomes of the method with those of recent research studies that have described methods like Sparse Reconstruction (SR) [29], Social Force (SF) [29], Anomaly Net [29], Abnormality Mining and Detection Network (AMDN) [29], Gaussian Mixture Model- Fully Convolutional Network (GMM-FCN) [29], Anomaly Over-exaggeration (AOE) [29], Generative Adversarial Networks (GAN) [29] Dual Spatio- Temporal Network (DSTN) [29], and Deep Regression Spatial- Temporal Network (DR-STN) [29].

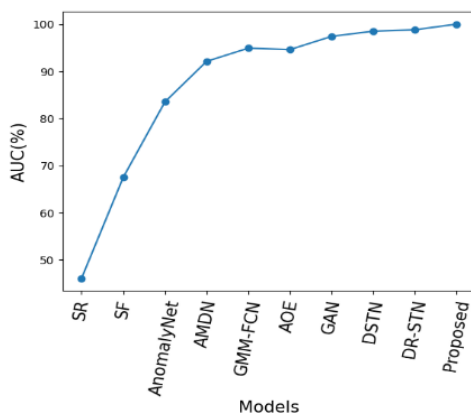


Fig. 6. Comparison of AUC

Fig. 6. displays comparisons of AUC. The AUC of anomaly detection can be increased by using a hybrid of ResNet-50 and LSTM. In terms of AUC, the proposed method performs better than SR, SF, Anomaly Net, AMDN, GMM-FCN, AOE, GAN, DSTN, and DR-STN by 67.5%, 46.1%, 92.1%,

97.4%, 83.5%, 98.5%, 94.9%, 94.6%, and 98.9%, respectively. As a result, conventional methods surpass the unique, revolutionary method, which has an AUC rate of 99.00%.

5. Conclusion

This work proposed a novel hybrid technique for finding anomalous behaviours in the UCSD dataset utilising ResNet-50 and LSTM approaches. The suggested method significantly improves the accuracy of identifying and categorising aberrant actions from the video data. Using ResNet-50, high-level features that reflect discriminative information about the activities may be effectively extracted. The LSTM is then given feature frames to differentiate between the normal and abnormal classes. The best model was proposed, with a 96.49% average classification accuracy. This model has 96.48% sensitivity, 96.48% recall, 96.48% F1-score, and 99.00% AUC values. The research has important implications for numerous practical applications using surveillance systems. ResNet-50 and LSTM-based automated abnormal activity recognition can significantly improve these domains' interactions, safety, and effectiveness. Future studies may examine how to enhance further and optimize the suggested strategy. This can entail researching other ResNet-50 and LSTM architectural variants, investigating cutting-edge data augmentation methods, or assessing the performance on more extensive and varied datasets.

References

- [1] G. Sreenu, & S. Durai, "Intelligent video surveillance: a review through deep learning techniques for crowd analysis," *Journal of Big Data*, vol. 6, no. 1, pp. 1-27, 2019.
- [2] A. Alshammari, & D. B. Rawat, "Intelligent multi-camera video surveillance system for smart city applications," *In 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC) IEEE*, pp. 0317-0323, 2019, January.
- [3] Y. Feng, Y. Yuan, & X. Lu, "Learning deep event models for crowd anomaly detection," *Neurocomputing*, vol. 219, pp. 548-556, 2017.
- [4] P. Khaire, & P. Kumar, "A semi-supervised deep learning-based video anomaly detection framework using RGB-D for surveillance of real-world critical environments," *Forensic Science International: Digital Investigation*, vol. 40, p. 301346, 2022.
- [5] Y. Zhou, B. Li, J. Wang, E. Rocco, & Q. Meng, "Discovering unknowns: Context-enhanced anomaly detection for curiosity-driven autonomous

- underwater exploration,” *Pattern Recognition*, vol. 131, p. 108860, 2022.
- [6] X. Cheng, L. Yuan, Z. Liu, & F. Guo, “Comparative analysis of video anomaly detection algorithms,” *In International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2022) SPIE*, vol. 12288, pp. 431-436, 2022, December.
- [7] R. F. Mansour, J. Escorcia-Gutierrez, M. Gamarra, J. A. Villanueva, & N. Leal, “Intelligent video anomaly detection and classification using faster RCNN with deep reinforcement learning model,” *Image and Vision Computing*, vol. 112, p. 104229, 2021.
- [8] L. A. Saleem, & E. V. Reddy, “A Survey on Deep Learning based Video Surveillance Framework,” *In 2023 International Conference on Computer Communication and Informatics (ICCCI), IEEE*, pp. 1-6, 2023, January.
- [9] K. Singh, S. Rajora, D. K. Vishwakarma, G. Tripathi, S. Kumar, & G. S. Walia, “Crowd anomaly detection using aggregation of ensembles of fine-tuned convnets,” *Neurocomputing*, vol. 371, pp. 188-198, 2020.
- [10] N. L. Lavanya, N. Vijayananda, N. N. Sattigeri, R. K. Nisarga, & N. M. Pooja, “Survey on Abnormal Event Detection and Signalling in Multiple Video Surveillance Scenes Using CNN,” *International Journal of Human Computations & Intelligence*, vol. 2, no. 3, pp. 159-168, 2023.
- [11] J. Ren, F. Xia, Y. Liu, & I. Lee, “Deep video anomaly detection: Opportunities and challenges,” *In 2021 international conference on data mining workshops (ICDMW) IEEE*, pp. 959-966, 2021, December.
- [12] G. Pang, C. Shen, L. Cao, & A. V. D. Hengel, “Deep learning for anomaly detection: A review,” *ACM computing surveys (CSUR)*, vol. 54, no. 2, pp. 1-38, 2021.
- [13] E. Şengönül, R. Samet, Q. Abu Al-Haija, A. Alqahtani, B. Alturki, & A. A. Alsulami, “An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey,” *Applied Sciences*, vol. 13, no. 8, p. 4956, 2023.
- [14] A. Berroukham, K. Housni, M. Lahraichi, & I. Boulfrifi, “Deep learning-based methods for anomaly detection in video surveillance: a review,” *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 1, pp. 314-327, 2023.
- [15] M. Arunachalam, V. Sanghavi, Y. A. Yao, Y. A. Zhou, L. A. Wang, Z. Wen, & F. Mohammad, “Strategies for Optimising End-to-End Artificial Intelligence Pipelines on Intel Xeon Processors,” *arXiv preprint arXiv:2211.00286*, 2022.
- [16] H. T. Duong, V. T. Le, & V. T. Hoang, “Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey,” *Sensors*, vol. 23, no. 11, p. 5024, 2023.
- [17] F. Mumcu, K. Doshi, & Y. Yilmaz, “Adversarial Machine Learning Attacks Against Video Anomaly Detection Systems,” *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 206-213, 2022.
- [18] Z. K. Abbas, & A. A. Al-Ani, “A Comprehensive Review for Video Anomaly Detection on Videos,” *In IEEE 2022 International Conference on Computer Science and Software Engineering (CSASE)*, pp. 1-1, 2022, March.
- [19] Y. Hao, J. Li, N. Wang, X. Wang, & X. Gao, “Spatiotemporal consistency-enhanced network for video anomaly detection,” *Pattern Recognition*, vol. 121, p. 108232, 2022.
- [20] R. Nawaratne, D. Alahakoon, D. De Silva, & X. Yu, “Spatiotemporal anomaly detection using deep learning for real-time video surveillance,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 393-402, 2019.
- [21] Y. Kim, J. Y. Yu, E. Lee, & Y. G. Kim, “Video anomaly detection using Cross U-Net and cascade sliding window,” *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 3273-3284, 2022.
- [22] J. Arunnehru, “Deep learning-based real-world object detection and improved anomaly detection for surveillance videos,” *Materials Today: Proceedings*, vol. 80, pp. 2911-2916, 2023.
- [23] W. Shao, R. Kawakami, & T. Naemura, “Anomaly Detection Using Spatio-Temporal Context Learned by Video Clip Sorting,” *IEICE TRANSACTIONS on Information and Systems*, vol. 105, no. 5, pp. 1094-1102, 2022.
- [24] A. A. Khan, M. A. Nauman, M. Shoaib, R. Jahangir, R. Alroobaea, M. Alsafyani, & C. Wechtaisong, “Crowd Anomaly Detection in Video Frames Using Fine-Tuned AlexNet Model,” *Electronics*, vol. 11, no. 19, p. 3105, 2022.
- [25] S. W. Khan, Q. Hafeez, M. I. Khalid, R. Alroobaea, S. Hussain, J. Iqbal, & S. S. Ullah, “Anomaly

detection in traffic surveillance videos using deep learning,” *Sensors*, vol. 22, no. 17, p. 6563, 2022.

- [26] A. Chriki, H. Touati, H. Snoussi, & F. Kamoun, “Deep learning and handcrafted features for one-class anomaly detection in UAV video,” *Multimedia Tools and Applications*, vol. 80, pp. 2599-2620, 2021.
- [27] S. Ul Amin, M. Ullah, M. Sajjad, F. A. Cheikh, M. Hijji, A. Hijji, & K. Muhammad, “EADN: An Efficient Deep Learning Model for Anomaly Detection in Videos,” *Mathematics*, vol. 10, no. 9, p. 1555, 2022.
- [28] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, & S. W. Baik, “CNN features bi-directional LSTM for real-time anomaly detection in surveillance networks,” *Multimedia tools and applications*, vol. 80, pp. 16979-16995, 2021.
- [29] T. Ganokratanaa, S. Aramvith, & N. Sebe, “Video anomaly detection using deep residual-spatiotemporal translation network,” *Pattern Recognition Letters*, vol. 155, pp. 143-150, 2022.
- [30] Meneses-Claudio, B. ., Perez-Siguas, R. ., Matta-Solis, H. ., Matta-Solis, E. ., Matta-Perez, H. ., Cruzata-Martinez, A. ., Saberbein-Muñoz, J. ., & Salinas-Cruz, M. . (2023). Automatic System for Detecting Pathologies in the Respiratory System for the Care of Patients with Bronchial Asthma Visualized by Computerized Radiography. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(2), 27–34. <https://doi.org/10.17762/ijritcc.v11i2.6107>
- [31] Kwame Boateng, Machine Learning in Cybersecurity: Intrusion Detection and Threat Analysis , *Machine Learning Applications Conference Proceedings*, Vol 3 2023.

Comprehensive survey on video anomaly detection using deep learning techniques

Sreedevi R. Krishnan*

Department of Computer Science and Engineering,
School of Engineering,
Avinashilingam Institute for Home Science and
Higher Education for Women, Coimbatore,
Tamil Nadu, India
Email: sreedevirkrishnan@gmail.com
*Corresponding author

P. Amudha and S. Sivakumari

Department of Computer Science and Engineering,
School of Engineering,
Avinashilingam Institute for Home Science and
Higher Education for Women,
Coimbatore, Tamil Nadu, India
Email: amudha_cse@avinuity.ac.in
Email: sivakumari_cse@avinuity.ac.in

Abstract: The rapid increase in violence and crime leads to the use of video surveillance systems. Handling such huge videos and classifying them as abnormal or not are tedious. Therefore, an automatic anomaly detection method is vital for the real-time detection of anomalous events. Advancements in machine intelligence lead to an automatic anomaly detection system for the timely identification of anomalous events and reducing the after-effects. Recent research uses deep learning techniques for faster and automatic detection of abnormal events from an enormous volume of surveillance videos. Reviewing the video anomaly detection system is very relevant and helps to promote future research in this area. The paper performs a comprehensive study of several video anomaly detection methods using deep learning techniques to detect and predict anomalous events. The paper also surveys various methods used for women's safety. Various methodologies, datasets, and evaluation metrics for detecting video anomalies and comparisons are included.

Keywords: deep learning; CNN; LSTM; GAN; autoencoder; women safety.

Reference to this paper should be made as follows: Krishnan, S.R., Amudha, P. and Sivakumari, S. (2024) 'Comprehensive survey on video anomaly detection using deep learning techniques', *Int. J. Computational Vision and Robotics*, Vol. 14, No. 4, pp.445–466.

Biographical notes: Sreedevi R. Krishnan is a Master of Engineering holder in Computer Science and Engineering and is a research scholar at the Department of Computer Science and Engineering, School of Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India. She has more than 16 years of teaching and two years of

research experience. She is pursuing her PhD at Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India. Her research interest includes data mining, computer vision, machine learning, deep learning, network security and anomaly detection. She is a Life Member of the Indian Society for Technical Education (ISTE). ORCID: <https://orcid.org/0000-0002-2769-7747>.

P. Amudha is working as a Professor in the Department of Computer Science and Engineering, School of Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India. She has more than 20 years of teaching and ten years of research experiences. Her research interests include data mining, machine learning and information security. She has more than 50 publications in national and international journals and conferences. She acts as a reviewer for IEEE, Springer and many reputed journals and conferences. She is a Life Member of ISTE and associate member of CSI and IAENG.

S. Sivakumari is a Professor and Head in the Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India. She has 34 years of teaching and 12 years of research experiences. Her areas of interest include data mining, machine learning, artificial intelligence and soft computing. She has supervised 15 PhD scholars and has more than 70 publications in peer-reviewed journals and prestigious conferences. She is the member of professional bodies like CSI, ISTE and IAENG. She is the reviewer for many IEEE conferences and journals. She has visited abroad for technical deliberations.

1 Introduction

Recently, the violence and crime rate are increasing; if such an anomalous situation is identified or predicted on time, then prompt reporting of such will help avoid the threat and thereby ensure the safety of people. Now, most public places like shops, bus stops, streets, organisations, etc. are equipped with surveillance cameras for the safety of people. The authorities can identify the issues and catch the culprit by analysing these video backups. However, this model can only be utilised for previously occurred events. Therefore, a novel model for anomaly detection on a real-time basis is required.

The primary motivation for selecting this topic is to help ordinary people to be safe by predicting anomalous events and thereby safeguarding them. Video anomaly detection is a much-awaited area of research that can be utilised in future for home security purposes like the safety of older people or even for women and children who need home care. Automatic video anomaly detection can help protect sensitive areas, which helps the militants identify the presence of enemies quickly. This system can also be used in industries, safeguarding fellow workers' industry resources and life. The police department can also identify unusual event patterns automatically and may replace the tedious tasks of the officers regarding the continuous monitoring of significant areas inside cities using the automatic video analysis techniques. Similarly, ATM counters can be safely monitored using these techniques, and much recent research has been conducted based on this area (Viji et al., 2021).

With artificial intelligence (AI) advancement, many tiresome, like surveillance video analysis for anomaly detection operations, can be effectively performed using machine intelligence. It is made possible by the advancements in AI and artificial neural networks, which work using the concept of neurons in the human brain. Like an intrusion detection system in a network system (Amudha et al., 2015), an automatic video anomaly detection system will help identify intrusions that occur under a video surveillance system that is very helpful for identifying intrusions and abnormal events in advance and make necessary precautions.

Many computer vision (CV) methods using deep learning methods are materialised for handling the massive quantity of videos. The deep learning techniques will detect abnormal events and adapt to the newly occurred events in the testing phase (Boundour et al., 2019). In addition, the high-level features can be processed from video data, reducing the need for domain expertise and hardcore feature extraction (Mahapatra, 2018). CV is a field of AI in which computers can see and identify various parameters in the image called features. The CV helps to derive meaning from the given digital image or videos from any visual input device. These innovations can be utilised effectively for the identification of abnormal situations.

The paper intends to provide a survey for identifying the various anomaly detection systems that can provide human safety. It also tries to identify the various safety methods employed for women and child safety. The paper focuses on two significant aspects: video anomaly detection and women's safety. The papers selected for the survey are based on the automatic detection of video anomalies employing deep learning techniques and various safety techniques used for women's safety.

The paper is organised as follows: Section 2 describes methods and recent developments in the video surveillance system used for anomaly detection. Section 3 and Section 4 discuss the datasets and the assessment metrics used for video anomaly detection. Section 5 provides the results and discussions about the survey, and Section 6 concludes the survey.

1.1 The necessity of a smart video surveillance

Based on studies, human observers may miss 60% of targeted issues while monitoring nine or more display units (Sulman et al., 2008). In addition, the human attention span is reduced from 15 to 10 minutes or even more (Bradbury, 2016). Real-time manual recognition of abnormalities in a video surveillance system is tedious, inefficient and costly. Many data will be available in the surveillance system, and manual verification of this big data is unbearable. There exists an in-need for an automatic anomaly detection mechanism.

Based on the study (World Health Organisation, 2021) conducted by WHO in 2018, among 161 countries and areas, nearly 1 in 3 or 30% of ladies are exposed to physical or sexual violence. UNICEF states (UNICEF for Every Child, n.d.) that during the COVID-19 pandemic, violence and exploitation among children and women are increased across countries. A study conducted by UNICEF from 154 countries in 2020 found that violence against children reached 54% more compared to 2019. The advancements in AI can be effectively utilised for detecting violence in public places or organisations by providing real-time anomaly detection.

1.2 Applications of video anomaly detection using deep learning techniques

Deep learning has the fastest advancements in the machine learning research field (Gragnaniello et al., 2020), which can handle a large amount of data and easily detect objects. The result is an incredible impact of using deep learning technologies in many daily applications, especially surveillance security systems, autonomous driving, speech recognition, translations, etc. Gragnaniello et al. (2020) shows that CV deals with many applications and helps with image classification and object detection.

Some of the significant applications of the detection of anomalies in video surveillance are:

- human action recognition
- theft control and detection
- violation of traffic rules
- safety of women in public transport and premises
- child safety in the crowd
- safety of public
- crowd controlling
- COVID protocol violation detection in public places
- presence of facial mask in the crowd
- person identification.

2 Methods and recent developments

A video anomaly detection system generally accepts video as input dataset and is fed to the pre-processing data module, which can remove duplicate and irrelevant characteristics. This module can also convert the input video data into a collection of images. A portion of the video dataset is taken as training dataset, and the remaining is taken as testing dataset. Using deep learning algorithms, the features are extracted and classified into normal and abnormal events, as shown in Figure 1.

Figure 1 Video anomaly detection system

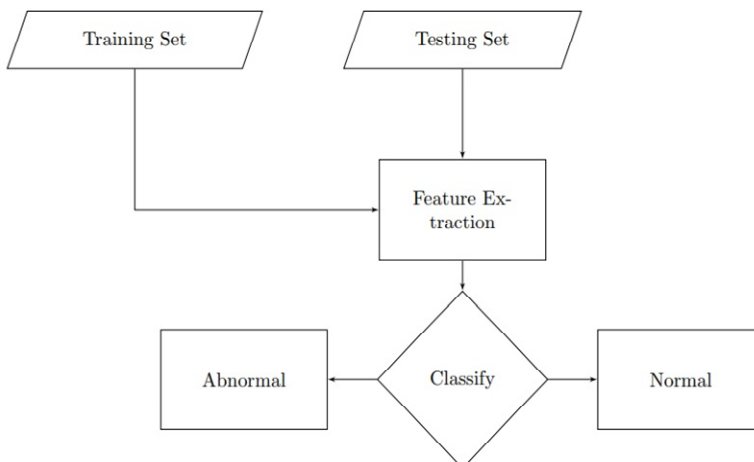


Table 1 Summary of video anomaly recognition using deep learning techniques

<i>Primary focus of the paper</i>	<i>Name of the journal</i>	<i>Merits</i>	<i>Future scope</i>
Spatiotemporal anomaly detection (Nawaratne et al., 2020)	<i>IEEE Transactions on Industrial Informatics</i>	The continuous monitoring of a human observer can be avoided.	It may reduce human involvement in finalising the anomaly or not. False-negative detection.
Abnormal events detection using spatio-temporal ConvNet (Bouindour et al., 2019)	<i>Applied Sciences</i>	High performance in localisation and detection of abnormal events. Simple and fast.	More localisation of abnormal events.
Hybrid deep learning and HOF (Hamdi et al., 2019)	6th International Conference on Control, Decision and Information Technologies (CoDIT)	Good in anomaly detection speed and accuracy. It reduced false negatives.	Need more anomaly localisation.
Anomaly recognition in CCTV using neural networks (Singh et al., 2020)	<i>Procedia Computer Science</i>	The model exhibits an overall accuracy of 97.23% with a reduced over-fitting level.	Need a cost-effective real-time anomaly detection.
Deep anomaly detection using visual attention (Nasaruddin et al., 2020)	<i>J. Big Data</i>	The model can receive an accuracy of 99.25	Need more localised anomaly detection model for the detection of moving objects.
Deep convolutional framework for anomalous behaviour detection (Ko and Sim, 2018)	<i>Engineering Applications of Artificial Intelligence</i>	By observing human behaviours, anomalous events can be identified.	Modify the model to handle all types of inputs rather than only static poses.
Continual learning for detection of anomaly (Doshi and Yilmaz, 2020)	IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)	The model can update the learned feature incrementally within seconds.	Need anomalous labelling for newly learned frames. The identification may be performed using an online object and action recognition.
Ensemble learning using bagging and inception-V3 (Zahid et al., 2020)	IEEE International Conference on Image Processing	Inception-v3 deep learning network removes. Segmentation and improves robustness. Provides 92.06% of accuracy.	Need to improve for fine-grained classification.
Dual discriminator generative adversarial network (Dong et al., 2020)	<i>IEEE Access</i>	The model exhibits good performance concerning the evaluation metrics like EER, AUC and PSNR gap.	The prediction qualities of future frames have to be increased.

Table 1 Summary of video anomaly recognition using deep learning techniques (continued)

<i>Primary focus of the paper</i>	<i>Name of the journal</i>	<i>Merits</i>	<i>Future scope</i>
Bidirectional retrospective generation adversarial network (Yang et al., 2021)	<i>IEEE Access</i>	The model can easily detect anomalous events like fighting, robbery, etc.	The model may modify the BR-GAN to identify smaller objects at a distance.
Deep learning for real-time surveillance (Iqbal et al., 2021)	<i>Security and Communication Networks</i>	The model provides fast learning with feature reduction.	The model may train with a dataset consisting of fire and smoke.
Multi-channel generative framework and supervised learning (Vu et al., 2021)	<i>Sensors</i>	The multi-channel CGAN provides anomaly detection with object-centric anomaly localisation.	SVM classifiers and additional complex datasets can be implemented for more refined results.
Asymptotic bound on false alarm rate for anomaly detection (Doshi and Yilmaz, 2021)	<i>Pattern Recognition</i>	The model performs well with a minimised false alarm.	Timely detection of videos.
Early-anomaly prediction (Emad et al., 2021)	MIUCC – IEEE Access	Future frame prediction can reduce crimes.	Model to predict future frames with an improved level of accuracy.
Deep edge computing (Kim et al., 2021)	<i>IEEE Access</i>	The model consists of modular architecture for detecting video anomalies in real-time.	Implementation of real-time video anomaly detection using a less expensive edge computing device.
Abnormal motion pattern detection using unsupervised method (Acharya and Joshi, 2021)	<i>Engineering and Applied Science Research</i>	The model provides a good level of accuracy by clustering the intervals.	Implement real-world surveillance videos of airports, road traffic systems, interior and exterior videos, etc.
Intelligent surveillance using deep learning (Pawar and Attar, 2020)	<i>International Journal of Computational Vision and Robotics</i>	The multivariate Gaussian probability model achieves significant frame-level anomaly detection.	Implement an anomaly localisation method that detects and localises the anomalies simultaneously.
A temporal smoothed Unet: Ts-Unet (Wang et al., 2021)	<i>Image and Graphics</i>	The model studies the temporal sameness details and attains better anomaly detection.	-

Recently, there has been much advancement in automatic anomaly detection systems using AI. With the advancements in deep learning methods, a more considerable amount of video data can be processed faster. Summary of video anomaly recognition using deep learning techniques are given in Table 1.

Much advancement is evolving in CV using deep learning techniques. The following are the extensive surveys conducted on various models developed between 2018 and 2022. Nawaratne et al. (2020) proposed an incremental spatio-temporal learner (ISTL), which uses real-time video from the surveillance system. The spatio-temporal autoencoder help to acquire the input video's motion structures. The sequence of CNN (Indolia et al., 2018) layers learns spatial, and the sequence of ConvLSTM (Indolia et al., 2018) learns temporal representation. This trained model detects an anomaly, verified and authenticated by individuals. The regular observations are performed with fuzzy aggregation and fed back to ISTL for continuous learning.

In Bouindour et al. (2019), a two-stage method is utilised, which is having feature extraction using a pre-trained 3D residual convolutional network (Tran et al., 2015, 2017) as the first stage and outlier detection using classification as the second stage. The spatio-temporal representations can be obtained using 3D-CNN (Ji et al., 2013) to identify abnormal events. Hamdi et al. (2019) proposed using a pre-trained histogram of optical flow (HOF) and CNN for spatio-temporal feature extraction (Wang and Snoussi, 2013). The one-class support vector machine (SVM) (Wang and Snoussi, 2013) was trained for abnormal and normal event classification. The algorithm provides appreciably speed and accuracy with fewer miss detections and false alarms.

In Singh et al. (2020), the CNN and recurrent neural network (RNN) (Indolia et al., 2018) are used for the identification and classification of anomalous events. The transfer learning (Marcelino, 2018) is applied in the pre-trained inceptionV3 model of Google (Google Cloud, 2021) for object identification. The model is re-trained using various ImageNet-like classes. The RNN is used to extract meaning and can classify the video segments as normal or abnormal events. Nasaruddin et al. (2020) proposed a background (BG) subtraction with a bilateral filter method for motion extraction from the foreground. Using a 3D CNN, the spatio-temporal relation is identified. The BG subtraction is implemented using K-nearest neighbours (KNN) and a mixture of Gaussians (MOG2) (Zivkovic, 2004; Zivkovic and Van Der Heijden, 2006), which helps to eliminate noise and extract the region correctly. The 3D CNN is used for feature extraction and anomaly detection.

Ko and Sim (2018) presents abnormal behaviour detection for an intelligent surveillance system using a deep learning model. The YOLO v2 (Redmon et al., 2016) network helps object detection faster than R-CNN. The posture classification procedure extracts spatial features of abnormal behaviours, and LSTM is used to detect abnormal behaviour. Doshi and Yilmaz (2020) proposed an online anomaly detection model for video surveillance using continual and transfer learning, which will help to decrease the training complication. The extraction of features is achieved by a neural network using transfer learning, and the abnormal event decision is made by continual statistical KNN. The feature detection of real-time video is performed using YOLOv3 (Redmon and Farhadi, 2018), and Euclidean KNN distance is used to identify abnormal data points.

Zahid et al. (2020) proposed an anomaly detection model, which used an inception-v3 deep learning network (Szegedy et al., 2016) for video feature extraction. The anomaly classification is performed using an identical bagging ensemble of a three-layer fully

connected (FC) network. In Dong et al. (2020), the video anomaly detection is performed using a dual discriminator generative adversarial network model (Goodfellow et al., 2020). The model consists of a frame and motion discriminator generator section for a more realistic and continuous frame. The frame discriminator can identify the originality of the frames from the input frames. Furthermore, the motion discriminator identifies whether the optical input flows are actual. During the testing phase of GAN, the quality of the predicted frame is compared with a regular score to identify whether it is abnormal or not. The lower prediction quality frames are considered abnormal ones.

Yang et al. (2021) proposed an anomaly detection model for upcoming frame forecast using a bidirectional retrospective generation adversarial network (BR-GAN). This model consists of a generator, sequence discriminator and frame discriminator. The generator module of the prediction framework consist U-Net (Ronneberger et al., 2015) for bidirectional prediction. The U-Net contain a balanced down-sampling and up-sampling system to mine features from the video input. The forecast frames are given to the input order for the verification review using retrospective prediction. The information like intensity, gradient loss, and adversarial loss of these frames are used for determining whether the model generates real or fake images. The frame discriminator consists of a patchGAN (Isola et al., 2017) matrix, which can determine whether each element in the patch is real or fake. The sequence discriminator is used to identify fake frames based on temporal consistency. Iqbal et al. (2021) proposed a quadcopter surveillance system for input video streams. A known faster RCNN algorithm is used for feature extraction initially. Five models of CNN algorithms were evaluated to identify threats in the paper. A ResNet-50- based faster RCNN was used for high precision and provided an excellent threat detection solution.

Vu et al. (2021) proposed a model for detecting anomalies using ConditionalGAN (CGAN) for prediction and reconstruction. A multi-channel CGAN framework produces feature maps related to motion and appearance. The differences between the feature spaces of images are measured using mean square error (MSE) (Great Learning, 2020) and peak signal-to-noise ratio (PSNR) (Ni, 2020). A binary SVM classifier is used for feature mapping in the final layer of anomaly detection. The anomaly localisation was done using mask R-CNN, which provides a better mapping of errors. Doshi and Yilmaz (2021) proposed an asymptotic bound on false alarm rate, and a unified model for features extraction and anomaly detection, namely multi-objective neural anomaly detector (MONAD) surveillance videos online anomaly detection. The future frame predictor using GAN and YOLOv3 lightweight object detectors is used for feature extraction in MONAD. A non-parametric statistical algorithm is implemented for online anomaly detection from extracted features.

In Emad et al. (2021), the prediction of abnormal behaviour is performed using GAN, which uses a trained dataset and deep learning neural network to classify usual and unusual events. This training information is used for the prediction of future anomalies. Kim et al. (2021) proposed deep neural network architecture using an edge computing environment for real-time video analysis. The model contains two 2D CNN as front-CNN and a shallow 3D CNN as back-CNN. The front-CNN performs feature extraction from the input video, and using three channels, namely red, green and blue (RGB), a three channel feature map (3CFM) is obtained. This result is provided to back-CNN for transfer learning. The training performed using spatial and temporal data of videos using the University of Central Florida (UCF)-Crime Dataset provides an efficient model for recognition of real-time videos in edge computing environments. Acharya and Joshi

(2021) identify the abnormal motion pattern using unsupervised hierarchical agglomerative clustering. The extracted features from the dense trajectories and grouped based on the difference in the intervals. The feature points are clustered using the hierarchical agglomerative clustering technique, detecting abnormal motion patterns.

Pawar and Attar (2020) proposed an anomaly detection model in a surveillance system using a CNN, radial basis function (RBF) and LSTM. A two-dimensional convolutional autoencoder learns the input feature, and using an LSTM network the temporal and statistical correlation is identified. The anomalies are calculated using multivariate Gaussian distribution, and the RBF network classifies nominal and anomalous events. Wang et al. (2021) introduce a temporal-smoothed U-Net for video detection of an anomaly. The model's training is performed for predicting future frames and detecting anomalies using the predicted errors during the testing phase. A temporal-smooth loss is used for better detection of anomalies. The developments in detecting anomalies in CV and deep learning have a vast development and acceptance due to faster and more accurate detection.

Fute et al. (2021) provides a fusion approach for movement detection, which includes a MOG with a neural network. The BG subtraction method is used to speed up the MOG. Shao et al. (2022) proposed an anomaly detection method using auto-encoders and thereby learning the spatio-temporal context among sorted video clips. This method produced better detection of anomalies, and the prediction method can produce improved precision at high recall. Feng et al. (2022) present a detection method using a spatio-temporal autoencoder. This model features perturbation visual interpretation methods for localisation and visualisation of results (Feng et al., 2022). Wang et al. (2022) proposed a deep support vector data description-based video anomaly detection system which can map the standard sample space to the smallest hyperspace and anomalous sample outside the hyperspace. The paper proposed by Wang et al. (2022) used a ConvLSTM and convolutional recurrent autoencoder, which can identify the irregularity of temporal and spatial patterns. This method reconstructs the video clip to a higher reconstruction error named anomalies.

2.1 The existing scenario of women and child safety measures

Many developments are arising for the safety of women and children recently. Most of them were handheld devices that worked based on sensors or internet of things (IoT) technologies. Muskan et al. (2018) proposed an IoT-based wearable women's safety device to create pulse rate and body temperature patterns. The application alerts the authorities if these patterns rise above the threshold. These wearable devices consist of sensors. The ZigBee mesh network helps the device work in areas with internet facilities.

In Ajitha et al. (2020), a GPS-based system consisting of a camera in ladies purse will observe the persons she interacts with, and the picture of the abnormal activities can be provided to concerned authorities or relatives. The GPS area details can also be provided, which will help to locate the area. Yaswanth et al. (2020) proposed an IoT-based self-security wearable device that can share location details and culprit when the model acts in security mode. According to the situation, the fingerprint sensor is used to register as a panic button. The model also helps to input user location and predict the nearby safe place.

With the advancements of GAN, it is possible to create ultrarealistic deep fake videos, which may be used to harass or blackmail women and children, which can lead to false propaganda and may outburst with dire consequences. Suratkar et al. (2020) propose a method to reveal fake videos implemented using transfer learning and CNN. The model is trained to provide a binary classification that can separate real and manipulated videos. The developments based on video analysis and alerting for women's and children's safety are significant because many issues are still arising.

3 Datasets

There are various available datasets for anomaly detection. The University of California San Diego (UCSD) Pedestrian datasets (Ped1 and Ped2), the Chinese University of Hong Kong (CUHK) Avenue dataset, and the ShanghaiTech datasets are extensively used datasets for anomaly detection. The details about the datasets used for identifying the video anomalies used in the survey are specified in Table 2.

- UCSD dataset:* The UCSD Pedestrian 1 (Ped1) dataset (Nawaratne et al., 2020; Yang et al., 2021; Vu et al., 2021; Doshi and Yilmaz, 2021; Acharya and Joshi, 2021; Pawar and Attar, 2020) contain 34 training and 36 testing videos and have 40 unusual events. The unusual events are mainly automobiles which include cars and bicycles. The UCSD Pedestrian 2 dataset (Nawaratne et al., 2020; Bouindour et al., 2019; Hamdi et al., 2019; Dong et al., 2020; Doshi and Yilmaz, 2020, 2021; Yang et al., 2021; Vu et al., 2021; Acharya and Joshi, 2021) consists of 16 training and 12 testing videos and have 12 unusual events similar to Ped1. The significant differences between the two datasets are that they have some perspective distortion due to the camera's direction from to and fro. Figure 2 shows the images of UCSD Ped2, CUHK Avenue and ShanghaiTech datasets used in video anomaly detection.

Figure 2 Datasets with normal and abnormal events (see online version for colours)



Note: The abnormal events are shown in a red box.

Source: Doshi and Yilmaz (2020)

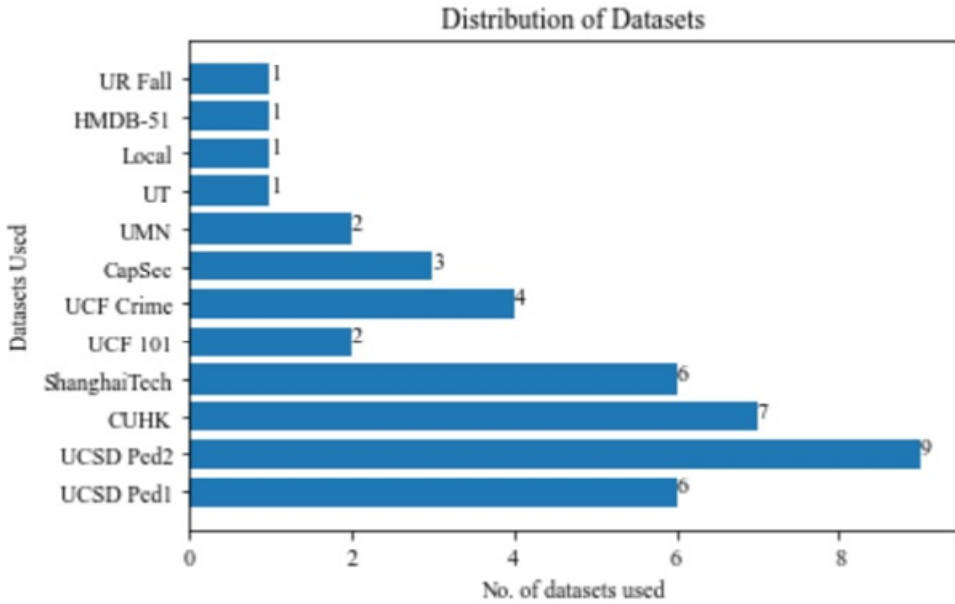
- *CUHK Avenue dataset* (Nawaratne et al., 2020; Dong et al., 2020; Doshi and Yilmaz, 2020, 2021; Yang et al., 2021; Vu et al., 2021; Emad et al., 2021): The CUHK Avenue dataset is a larger dataset that consists of 30,652 frames where 16 are training videos, and 21 are testing videos. The video in the dataset is a campus walkway where individuals walk to and fro a building. The dataset consists of 47 unusual events: running, wandering, and throwing objects. In the video, hikers are walking parallel and towards the surveillance camera.
- *ShanghaiTech dataset* (Dong et al., 2020; Doshi and Yilmaz, 2020, 2021; Yang et al., 2021; Vu et al., 2021; Emad et al., 2021): The dataset consists of street activity which comprises 330 training and 107 testing videos which consist of 130 unusual events covered from 13 various locations. This dataset challenges anomaly detection, including fighting (motion anomalies) and vehicles.
- *UCF-Crime dataset* (Singh et al., 2020; Nasaruddin et al., 2020; Zahid et al., 2020; Kim et al., 2021): The UCF-Crime dataset consists of 800 regular and 810 unusual videos for training. It consists of 150 regular and 140 unusual testing videos. The dataset includes many criminal activities like abuse, stealing, fighting, road accidents, arrests, shoplifting, assault, arson, explosion, shooting and vandalism.
- *UCF-101 dataset* (Bouindour et al., 2019; Kim et al., 2021): It consists of 13,320 videos with 101 classes in five action sets like performing a musical instrument, sports, human-human communication, human-object communication and body-motion.
- *HMDB-51 dataset* (Kim et al., 2021): The human motion database consists of 6,766 videos which contain 51 classes from five action sets. Facial movement with object handling, body movements with object interaction, general facial actions, general body activities, and body actions for human communication. The HMDB-51 dataset is generally used to measure the performance of video action recognition.
- *UMN dataset* (Hamdi et al., 2019; Acharya and Joshi, 2021): The dataset mainly contains three scenes lawn (with 1,450 frames), indoor (with 4,415 frames) and plaza (with 2,145 frames). The two occasions where people walk can be considered normal, and people running can be considered unusual.
- *CapSec dataset* (Bouindour et al., 2019): This is an actual, local case of the laboratory surveillance camera. The dataset represents the daily student and researcher behaviours in the laboratory. The training set consists of regular events, and the testing set consists of unusual events like the falling of people and objects.
- *UT-Interaction-Data* (Ko and Sim, 2018): This is a benchmark dataset that consists of 50 video clips which are having humans with various activities, namely hugging, kicking, handshaking, pointing, pushing and punching. The six activities are considered abnormal events. The dataset helps in the detection of high-level human activities which are complex.

- *U.R. Fall detection dataset* (Kim et al., 2021): The University of Rzeszow Fall detection datasets are benchmark datasets consisting of 70 interior videos, of which 30 are fall events, and 40 are daily life events. The fall videos are taken from two cameras for vertical and horizontal vision. The daily life activities are recorded using a single camera.
- *The available local dataset* (Iqbal et al., 2021): Is obtained from videos taken using surveillance drones in various environments. The videos of persons in various poses were taken with and without weapons, ordinary people and guards. These videos are converted to image frames for the preparation of the dataset.

Table 2 Particular datasets used for identification of video anomalies

<i>Dataset name</i>	<i>Type of dataset</i>	<i>Journal references</i>
UCSD Ped1	Pedestrian walkway	Nawaratne et al. (2020), Yang et al. (2021), Vu et al. (2021), Doshi and Yilmaz (2021), Acharya and Joshi (2021), Pawar and Attar (2020)
UCSD Ped2	Pedestrian walkway	Nawaratne et al. (2020), Bouindour et al. (2019), Hamdi et al. (2019), Dong et al. (2020), Doshi and Yilmaz (2020, 2021), Yang et al. (2021), Vu et al. (2021), Acharya and Joshi (2021)
CUHK Avenue	Street activity	Nawaratne et al. (2020), Dong et al. (2020), Doshi and Yilmaz (2020, 2021), Yang et al. (2021), Vu et al. (2021), Emad et al. (2021)
ShanghaiTech datasets	Street activity and crowd datasets	Dong et al. (2020), Doshi and Yilmaz (2020, 2021), Yang et al. (2021), Vu et al. (2021), Emad et al. (2021)
UCF-Crime Dataset	Crime dataset	Singh et al. (2020), Nasaruddin et al. (2020), Zahid et al. (2020), Kim et al. (2021)
UCF 101	Action record	Bouindour et al. (2019), Kim et al. (2021)
CapSec	Local dataset	Bouindour et al. (2019)
UMN	Lawn, indoor, plaza	Hamdi et al. (2019), Acharya and Joshi (2021)
U.T. Dataset	Interaction dataset	Ko and Sim (2018)
Local Dataset	Videos available using the quadcopter	Iqbal et al. (2021)
HMDB-51 Dataset	Facial actions and body movements	Kim et al. (2021)
U.R. Fall Detection Dataset	Daily living activities and fall	Kim et al. (2021)

The datasets covered in the survey of video anomaly detection and their frequency of use is shown in Figure 3. The UCSD datasets are one of the most popular among them.

Figure 3 Distribution of datasets used in the survey (see online version for colours)

4 Evaluation metric

There are many performance evaluation criteria for video anomaly detection models. Nayak et al. (2021) state that video anomaly detection can be performed based on accuracy-oriented (AO) approach and processing-time oriented (PO) approach. In the AO approach, the video anomaly detection, the performance is measured based on accuracy rate and low false alarm. Mostly complex models with high accuracy may experience high processing time. Generally, research based on generative, spatial, temporal, predictive, and hybrid models fall into the AO category. In the PO approach, the video anomaly detection is minimum frame processing time with a reasonable accuracy level. Most real-time applications modelled for attaining online performance, like lesser computational space and high computational speed, are implemented based on a PO-based approach. The histogram of optical flow orientation and magnitude (HOFM) and lightweight convolutional LSTM autoencoder algorithms of deep learning techniques fall under the PO approach. These models may lack the massive power of DNN.

Nayak et al. (2021) states various evaluation metrics for detecting video anomalies. Since anomaly detection models aim to classify them into normal or abnormal events, a binary classifier is needed for the measurement. A binary classifier's similar and dissimilar score patterns will be considered by setting a decision threshold. If the score of unusual events is more significant than a decided threshold, the output can be positive or negative accordingly. The various performance measures used for the detection of video anomalies are: confusion matrix, area under the curve (AUC), receiver operating characteristic (ROC) curve, equal error rate (EER), mean absolute error (MAE), MSE,

equal error rate of frame level (EERFL), equal error rate of pixel level (EERPL) and PSNR. The assessment metric used in the survey papers is given in Table 3.

- Confusion matrix (Narkhede, 2018)*: The performance can be analysed into true positive (TP), false negative (FN), true negative (TN) and false positive (FP). A matrix is constructed from the performance measures, as shown in Figure 4. The values for the matrixes are provided based on the data samples obtained from the model. The TP condition is given when the prediction of actual values is normal, and the TN is given when actual values are abnormal. The FP is when the model is predicted as usual, but the actual value is abnormal. This situation is also called a type 1 error. When the model is predicted as abnormal and the actual value is normal FN occurs, which is called a type 2 error. It is crucial for an anomaly detection model not to have a type 1 error FP.

Figure 4 Confusion matrix

		Actual Value	
		Normal	Abnormal
Predicted Value	Normal	True Positive	False Positive
	Abnormal	False Negative	True Negative

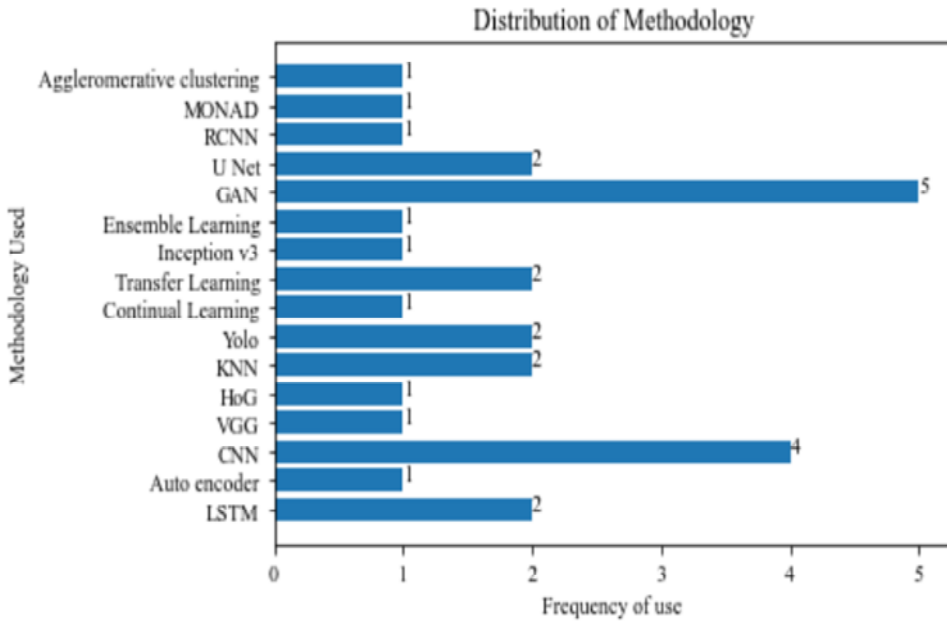
Source: Narkhede (2018)

- Precision (Narkhede, 2018)*: Is a metric for determining the model’s accuracy, which measures the number of actual positives from the predicted positives.
- Recall (Narkhede, 2018)*: Is a metric used for classifying problems that measures the best model. It is considered from the frequency of TPs divided by the total frequency of TPs and FNs.
- Accuracy (Narkhede, 2018)*: This evaluation metric measures the correct prediction from the positive and negative classes. The accuracy measure should be high as possible.
- ROC*: Receiver operating characteristic curve (Nawaratne et al., 2020) plots the true positive rate (TPR) against the false-positive rate according to the decision threshold. The TPR counts correct positive results among the set of positive samples during testing. The false positive rate (FPR) counts incorrect positive results among the negative samples during testing. The higher TPR and lower FPR for a threshold are good metrics.

- *AUC*: The AUC (Nawaratne et al., 2020) is the area under the ROC curve. For optimal, the measures are preferred to have minimised AUC.
- *EER*: The EER (Hamdi et al., 2019) is the value when the false acceptance rate (FP) and false rejection rate (FN) become equal, which is obtained from ROC. The lower the error rate, the higher the model's accuracy. The EERFL and the EERPL (Bouindour et al., 2019) are their variants. The EER rate is obtained from the total FP and FN divided by the number of frames (NF).
- *MSE*: Mean squared error (MSE) (Singh et al., 2020) is the average of the squares of the errors. The lower value of MSE is preferred.
- *PSNR*: PSNR (Yang et al., 2021; Ni, 2020) is an efficient way of accessing the quality of the image. The higher value of PSNR shows that the predicted frame is normal or abnormal.
- *Structural similarity index measure (SSIM)*: SSIM (Emad et al., 2021) is a perceptual metric for identifying images' similarity and prediction of image quality. For a good model, the preferred value of SSIM is high.

Table 3 Details of assessment metrics used in papers

<i>Ref.</i>	<i>Evaluation metric</i>
Nawaratne et al. (2020)	AUC, ROC, EER
Bouindour et al. (2019)	ROC, AUC, EER, EERFL, EERPL
Hamdi et al. (2019)	EER
Singh et al. (2020)	MSE, accuracy
Nasaruddin et al. (2020)	Accuracy
Ko and Sim (2018)	Recall, precision
Doshi and Yilmaz (2020)	AUC
Zahid et al. (2020)	ROC, AUC
Dong et al. (2020)	PSNR
Yang et al. (2021)	PSNR
Iqbal et al. (2021)	Accuracy
Vu et al. (2021)	MSE, PSNR
Doshi and Yilmaz (2021)	ROC, AUC
Emad et al. (2021)	MSE, PSNR, SSIM, AUC, ROC
Kim et al. (2021)	AUC, precision, recall, accuracy
Acharya and Joshi (2021)	EER, ROC
Pawar and Attar (2020)	EER
Wang et al. (2021)	-

Figure 5 Frequency of methodologies used in the survey (see online version for colours)

5 Results and discussion

There are various deep learning approaches for the identification of video anomalies. Figure 5 provides the list of the various methodologies for video anomaly identification and its corresponding frequency of use.

Table 4 summarises the papers in the survey based on datasets, their characteristics, advantages and future enhancements.

Video anomaly detection is an emerging technology in CV and deep learning, and there are many application areas in this field. The survey noted that even future frames could be predicted, and many anomalous situations can be avoided with prompt notification to authorised persons.

Table 5 shows the details of various women's safety measures available. Most papers deal with IoT wearable devices intended to contact the authorities or make emergency calls. Video surveillance systems are prevalent, and the utilisation of the surveillance system for women's safety-oriented anomaly detection is yet to be developed and implemented.

Table 4 Summary of video anomaly detection referred to in the survey

<i>Methodology</i>	<i>Dataset name</i>	<i>Characteristics</i>
ConvLSTM, Autoencoder (Nawarane et al., 2020)	UCSD Ped 1, UCSD Ped 2, CUHK	ISTL model identifies the temporal regularities by conserving the spatial structure of the video stream.
3D residual convolutional network (Bouindour et al., 2019)	UCSD Ped 2, UCF101, CapSec	Avoid training by implementing spatiotemporal feature extraction and incremental classifier.
VGG and histogram of oriented gradients (HOG) (Hamdi et al., 2019)	UCSD Ped 2, UMN	Spatio-temporal feature extraction using CNN and histogram of optical flow (HOF).
CNN and RNN (Singh et al., 2020)	UCF Crime	The model can handle real anomalies and provide maximum accuracy.
The mixture of Gaussians, K nearest neighbours (Nasaruddin et al., 2020)	UCF Crime	The model helps find localised regions for segments, and background (BG) subtraction helps find a region of interest (ROI); these help in more accurate anomaly recognition.
YOLOv2, CNN, LSTM (Ko and Sim, 2018)	U.T. dataset	The model helps in the identification of human behaviour.
YOLO v3, kNN, continual learning and transfer learning (Doshi and Yilmaz, 2020)	UCSD Ped 2, CUHK, ShanghaiTech	The model implements feature extraction of neural networks using continual learning and transfer learning.
Inception-V3 and bagging ensemble of 3 layers F.C. network (Zahid et al., 2020)	UCF Crime	Models a binary classifier using bagging ensemble in inception-V3.
GAN, U-Net (Dong et al., 2020)	ShanghaiTech, UCSD Ped 2, CUHK	First, the normal future frames are predicted, and during testing, the events varying from the predicted frames are considered abnormal.
BR-GAN (Yang et al., 2021)	UCSD Ped 1, UCSD Ped 2, CUHK, ShanghaiTech	The bidirectional and retrospective predictions can fully identify the association between the video frames for more accurate prediction.
RCNN, a quadcopter (Iqbal et al., 2021)	Local Dataset	The quadcopter helps in air surveillance which is more cost-effective and efficient.
ConditionalGAN, SVM (Vu et al., 2021)	UCSD Ped 1, UCSD Ped 2, CUHK, ShanghaiTech	The multi-channel CGANs provide better feature space, and binary SVM is used for anomaly detection.
MONAD (Doshi and Yilmaz, 2021)	UCSD Ped 1, UCSD Ped 2, CUHK, ShanghaiTech	Online video anomaly recognition for Surveillance with asymptotic bounds on false alarm rate and statistical decision making.
Retrospective cycle GAN and multiple-instance learning (MIL) (Emad et al., 2021)	CUHK, ShanghaiTech	The model using GAN can detect anomalies and also predict future events.
3D CNN, 2D CNN and transfer learning (Kim et al., 2021)	UCF 101, UCF Crime, HMDB-51, UR Fall Detection	The model processes 3 channel (RGB) images in an edge-computing environment.
Unsupervised hierarchical agglomerative clustering (Acharya and Joshi, 2021)	UCSD Ped 1, UCSD Ped 2, ShanghaiTech	The moving objects are grouped using hierarchical clustering to identify anomalies in surveillance videos.
CNN, LSTM and multivariate Gaussian (Pawar and Attar, 2020)	UCSD Ped 1	The 2D auto encoder identifies features, and LSTM analyses temporal and statistical correlation. The one-class classification detects anomalies.
Unet, generative model (Wang et al., 2021)	-	The model predicts future frames using regular videos, and the prediction errors are used to detect anomalies.

Table 5 Details of various women's safety measures

<i>Methodology</i>	<i>Characteristics</i>	<i>Merits</i>	<i>Limitations</i>
IoT-pulse rate sensor, temperature sensor, GSM modem, GPS module, ZigBee mesh network Arduino (Muskan et al., 2018)	It uses a wearable safety device that communicates with the emergency call list. The model can work even without an internet facility using ZigBee mesh.	Identifies the abnormal reading of the body like body temperature and pulse rates and can contact the emergency numbers.	It needs to train the model for more accurate detection. If the wearable device falls accidentally into trouble, then detection is impossible.
Emotional VGGnet, IoT, GPS module, SIM API (Ajitha et al., 2020)	The model uses an emotional VGGnet to identify the emotions captured using an input camera hidden in a lady's purse.	The Adam regularisation helps in the more accurate result. GPS module and SIM API help to send SMS to authorities and emergency contact list.	The camera positioning in the purse and the distortions in the image captured may affect the detection.
IoT, Raspberry Pi, finger sensor, KNN (Yaswanth et al., 2020)	The model uses a wearable jacket with a fingerprint sensor as a panic button, which helps capture the culprit's photo.	The KNN machine learning algorithm helps in predicting the nearest safe location.	The model cannot ensure the anomalous situation but provides proof for identifying the culprit.
GAN, CNN, transfer learning (Suratkar et al., 2020)	The model detects fake faces generated using the GAN to prevent women's abuse.	Transfer learning helps carry forward valuable knowledge, which helps in more accurate predictions.	The ResNet model does not perform well in fake face identification.

6 Conclusions

This paper comprehensively surveys numerous video anomaly detection using deep learning approaches. The paper also surveys women's safety and protection methods. The survey analysed various methodologies and the datasets used in video surveillance. The evaluation metrics used for anomaly detection were identified. The paper provides information about the dataset and its features in anomaly detection models. In addition, the paper undergoes a study of evaluation metrics used to detect normal and abnormal actions.

Even though various models are proposed for automatic video anomaly detection, implementing models in real-time is yet to complete. Since the model's performance evaluation varies based on accuracy and time, more research is needed to develop a model that exhibits good accuracy in minimum time.

References

- Acharya, H. and Joshi, B. (2021) 'Abnormal motion pattern detection in video sequences by an unsupervised approach', *Engineering and Applied Science Research*, Vol. 48, No. 5, pp.509–517 [online] <https://doi.org/10.14456/easr.2021.53>.
- Ajitha, P., Chowdary, J.L., Joshika, K., Sivasangari, A. and Gomathi, R.M. (2020) 'Third vision for women using deep learning techniques', *4th International Conference on Computer, Communication and Signal Processing, ICCSP 2020*, pp.18–21 [online] <https://doi.org/10.1109/ICCSP49186.2020.9315196>.
- Amudha, P., Karthik, S. and Sivakumari, S. (2015) 'A hybrid swarm intelligence algorithm for intrusion detection using significant features', *Scientific World Journal*, Vol. 2015 [online] <http://dx.doi.org/10.1155/2015/574589>.
- Bouindour, S., Snoussi, H., Hittawe, M.M., Tazi, N. and Wang, T. (2019) 'An online and adaptive method for detecting abnormal events in videos using spatio-temporal ConvNet', *Applied Sciences (Switzerland)*, Vol. 9, No. 4 [online] <https://doi.org/10.3390/app9040757>.
- Bradbury, N.A. (2016) 'Attention span during lectures: 8 seconds, 10 minutes, or more?', *Advances in Physiology Education*, Vol. 40, No. 4, pp.509–513 [online] <https://doi.org/10.1152/advan.00109.2016>.
- Dong, F., Zhang, Y. and Nie, X. (2020) 'Dual discriminator generative adversarial network for video anomaly detection', *IEEE Access*, Vol. 8, pp.88170–88176, doi: 10.1109/ACCESS.2020.2993373.
- Doshi, K. and Yilmaz, Y. (2020) 'Continual learning for anomaly detection in surveillance videos', *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June, pp.1025–1034 [online] <https://doi.org/10.1109/CVPRW50498.2020.00135>.
- Doshi, K. and Yilmaz, Y. (2021) 'Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate', *Pattern Recognition*, Vol. 114 [online] <https://doi.org/10.1016/j.patcog.2021.107865>.
- Emad, M., Ishack, M., Ahmed, M., Osama, M., Salah, M. and Khoriba, G. (2021) 'Early-anomaly prediction in surveillance cameras for security applications', *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference, MIUCC 2021*, pp.124–128 [online] <https://doi.org/10.1109/MIUCC52538.2021.9447668>.
- Feng, J., Wang, D. and Zhang, L. (2022) 'Crowd anomaly detection via spatial constraints and meaningful perturbation', *ISPRS International Journal of Geo-Information*, Vol. 11, No. 3, p.205, <https://doi.org/10.3390/ijgi11030205>.
- Fute, E.T., Sop Deffo, L.L. and Tonye, E. (2021) 'HADEM-MACS: a hybrid approach for detecting and extracting objects in movement by multimedia autonomous computer systems', *International Journal of Computational Vision and Robotics*, Vol. 11, No. 1, pp.21–40.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2020) 'Generative adversarial networks', *Communications of the ACM*, Vol. 63, No. 11, pp.139–144 [online] <https://doi.org/10.1145/3422622>.
- Google Cloud (2021) *Advanced Guide to Inception v3 on Cloud TPU* [online] <https://cloud.google.com/tpu/docs/inception-v3-advanced> (accessed 24 September 2021).
- Graganiello, D., Bottino, A., Cumani, S. and Kim, W. (2020) 'Special issue on advances in deep learning', *Applied Sciences (Switzerland)*, Vol. 10, No. 9, pp.1–6 [online] <https://doi.org/10.3390/app10093172>.
- Great Learning (2020) *Mean Squared Error – Explained | What is Mean Square Error?* [online] <https://www.mygreatlearning.com/blog/mean-square-error-explained/> (accessed 10 October 2021).
- Hamdi, S., Bouindour, S., Loukil, K., Snoussi, H. and Abid, M. (2019) 'Hybrid deep learning and HOF for anomaly detection', *2019 6th International Conference on Control, Decision and Information Technologies, CoDIT 2019*, June, pp.575–580 [online] <https://doi.org/10.1109/CoDIT.2019.8820352>.

- Indolia, S., Goswami, A.K., Mishra, S.P. and Asopa, P. (2018) 'Conceptual understanding of convolutional neural network – a deep learning approach', *Procedia Computer Science*, Vol. 132, pp.679–688 [online] <https://doi.org/10.1016/j.procs.2018.05.069>.
- Iqbal, M.J., Iqbal, M.M., Ahmad, I., Alassafi, M.O., Alfakeeh, A.S. and Alhomoud, A. (2021) 'Real-time surveillance using deep learning', *Security and Communication Networks*, Vol. 2021 [online] <https://doi.org/10.1155/2021/6184756>.
- Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A. (2017) 'Image-to-image translation with conditional adversarial networks', *Proceedings – 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, January, pp.5967–5976 [online] <https://doi.org/10.1109/CVPR.2017.632>.
- Ji, S., Xu, W., Yang, M. and Yu, K. (2013) '3D convolutional neural networks for human action recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 1, pp.221–231 [online] <https://doi.org/10.1109/TPAMI.2012.59>.
- Kim, J.H., Kim, N. and Won, C.S. (2021) 'Deep edge computing for videos', *IEEE Access*, Vol. 9, pp.123348–123357 [online] <https://doi.org/10.1109/ACCESS.2021.3109904>.
- Ko, K.E. and Sim, K.B. (2018) 'Deep convolutional framework for abnormal behavior detection in a smart surveillance system', *Engineering Applications of Artificial Intelligence*, September 2017, Vol. 67, pp.226–234 [online] <https://doi.org/10.1016/j.engappai.2017.10.001>.
- Mahapatra, S. (2018) 'Why deep learning over traditional machine learning?', *Towards Data Science* [online] <https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063> (accessed on 18 September 2021).
- Marcelino, P. (2018) 'Transfer learning from pre-trained models', *Towards Data Science* [online] <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751> (accessed 24 September 2021).
- Muskan, T.K., Khandelwal, M. and Pandey, P.S. (2018) 'Women safety device designed using IoT and machine learning', *Proceedings – 2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovations, SmartWorld/UIC/ATC/ScalCom/CBDCo*, pp.1204–1210 [online] <https://doi.org/10.1109/SmartWorld.2018.00210>.
- Narkhede, S. (2018) 'Understanding confusion matrix', *Towards Data Science* [online] <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62> (accessed 21 November 2021).
- Nasaruddin, N., Muchtar, K., Afdhal, A. and Dwiyanoro, A.P.J. (2020) 'Deep anomaly detection through visual attention in surveillance videos', *Journal of Big Data*, Vol. 7, No. 1 [online] <https://doi.org/10.1186/s40537-020-00365-y>.
- Nawaratne, R., Alahakoon, D., De Silva, D. and Yu, X. (2020) 'Spatiotemporal anomaly detection using deep learning for real-time video surveillance', *IEEE Transactions on Industrial Informatics*, Vol. 16, No. 1, pp.393–402 [online] <https://doi.org/10.1109/TII.2019.2938527>.
- Nayak, R., Pati, U.C. and Das, S.K. (2021) 'A comprehensive review on deep learning-based methods for video anomaly detection', *Image and Vision Computing*, Vol. 106 [online] <https://doi.org/10.1016/j.imavis.2020.104078>.
- Ni (2020) *Peak Signal-to-Noise Ratio as an Image Quality Metric* [online] <https://www.ni.com/en-in/innovations/white-papers/11/peak-signal-to-noise-ratio-as-an-image-quality-metric.html> (accessed on 10 October 2021).
- Pawar, K. and Attar, V. (2020) 'Deep learning-based intelligent surveillance model for detection of anomalous activities from videos', *International Journal of Computational Vision and Robotics*, Vol. 10, No. 4, pp.289–311 [online] <https://doi.org/10.1504/IJCVR.2020.108152>.
- Redmon, J. and Farhadi, A. (2018) *YOLOv3: An Incremental Improvement* [online] <http://arxiv.org/abs/1804.02767>.

- Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) 'You only look once: unified, real-time object detection', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, December, pp.779–788 [online] <https://doi.org/10.1109/CVPR.2016.91>.
- Ronneberger, O., Fischer, P. and Brox, (2015) 'U-net: convolutional networks for biomedical image segmentation', *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, pp.234–241.
- Shao, W., Kawakami, R. and Naemura, T. (2022) 'Anomaly detection using spatio-temporal context learned by video clip sorting', *IEICE Transactions on Information and Systems*, Vol. 105, No. 5, pp.1094–1102, DOI: 10.1587/transinf.2021EDP7207.
- Singh, V., Singh, S. and Gupta, P. (2020) 'Real-time anomaly recognition through CCTV using neural networks', *Procedia Computer Science*, Vol. 173, pp.254–263 [online] <https://doi.org/10.1016/j.procs.2020.06.030>.
- Sulman, N., Sanocki, T., Goldgof, D. and Kasturi, R. (2008) 'How effective is human video surveillance performance?', *Proceedings – International Conference on Pattern Recognition*, pp.7–9 [online] <https://doi.org/10.1109/icpr.2008.4761655>.
- Suratkar, S., Johnson, E., Variyambat, K., Panchal, M. and Kazi, F. (2020) 'Employing transfer-learning based CNN architectures to enhance the generalizability of deepfake detection', *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020* [online] <https://doi.org/10.1109/ICCCNT49239.2020.9225400>.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016) 'Rethinking the inception architecture for computer vision', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, December, pp.2818–2826 [online] <https://doi.org/10.1109/CVPR.2016.308>.
- Tran, D., Bourdev, L., Fergus, R., Torresani, L. and Paluri, M. (2015) 'Learning spatiotemporal features with 3D convolutional networks', *Proceedings of the IEEE International Conference on Computer Vision 2015 Inter*, pp.4489–4497 [online] <https://doi.org/10.1109/ICCV.2015.510>.
- Tran, D., Ray, J., Shou, Z., Chang, S-F. and Paluri, M. (2017) *ConvNet Architecture Search for Spatiotemporal Feature Learning*, Section 3 [online] <http://arxiv.org/abs/1708.05038>.
- UNICEF for Every Child (n.d.) *Global Annual Results Report 2020: Goal Area 3 Ensuring that Every Child is Protected from Violence and Exploitation* [online] <https://www.unicef.org/reports/global-annual-results-2020-goal-area-3#download> (accessed 17 September 2021).
- Viji, S., Kannan, R. and Jayalashmi, N.Y. (2021) 'Intelligent anomaly detection model for ATM booth surveillance using machine learning algorithm: intelligent ATM surveillance model', *Proceedings – IEEE 2021 International Conference on Computing, Communication, and Intelligent Systems, ICC CIS 2021*, pp.1007–1012, DOI: 10.1109/ICCCIS51004.2021.9397103.
- Vu, T.H., Boonaert, J., Ambellouis, S. and Taleb-Ahmed, A. (2021) 'Multi-channel generative framework and supervised learning for anomaly detection in surveillance videos', *Sensors*, Vol. 21, No. 9, pp.1–16 [online] <https://doi.org/10.3390/s21093179>.
- Wang, B. and Yang, C. (2022) 'Video anomaly detection based on convolutional recurrent autoencoder', *Sensors*, Vol. 22, No. 12 [online] <https://www.mdpi.com/1424-8220/22/12/4647>.
- Wang, B., Yang, C. and Chen, Y. (2022) 'Detection anomaly in video based on deep support vector data description', *Computational Intelligence and Neuroscience*, Vol. 2022, p.5362093, DOI: 10.1155/2022/5362093.
- Wang, T. and Snoussi, H. (2013) 'Histograms of optical flow orientation for abnormal events detection', *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS*, pp.45–52 [online] <https://doi.org/10.1109/PETS.2013.6523794>.

- Wang, Z., Yang, Z., Zhang, Y., Su, N. and Wang, G. (2021) 'Ts-Unet: a temporal smoothed Unet for video anomaly detection', *Image and Graphics. ICIG 2021. Lecture Notes in Computer Science*, Springer, Cham, Vol. 12890 [online] https://doi.org/10.1007/978-3-030-87361-5_64.
- World Health Organisation (2021) *Violence Against Women* [online] <https://www.who.int/news-room/fact-sheets/detail/violence-against-women> (accessed 17 September 2021).
- Yang, Z., Liu, J. and Wu, P. (2021) 'Bidirectional retrospective generation adversarial network for anomaly detection in videos', *IEEE Access*, Vol. 9, pp.107842–107857 [online] <https://doi.org/10.1109/ACCESS.2021.3100678>.
- Yaswanth, B.S., Darshan, R.S., Pavan, H., Srinivasa, D.B. and Murthy, B.T.V. (2020) 'Smart safety and security solution for women using KNN algorithm and IoT', *MPCIT 2020 – Proceedings: IEEE 3rd International Conference on "Multimedia Processing, Communication and Information Technology"*, February, pp.87–92 [online] <https://doi.org/10.1109/MPCIT51588.2020.9350431>.
- Zahid, Y., Tahir, M.A. and Durrani, M.N. (2020) 'Ensemble learning using bagging and Inception-V3 for anomaly detection in surveillance videos', *2020 IEEE International Conference on Image Processing (ICIP)*, pp.588–592, DOI: 10.1109/ICIP40778.2020.9190673.n.d.
- Zivkovic, Z. (2004) 'Improved adaptive Gaussian mixture model for background subtraction', *Proceedings – International Conference on Pattern Recognition*, Vol. 2, pp.28–31 [online] <https://doi.org/10.1109/icpr.2004.1333992>.
- Zivkovic, Z. and Van Der Heijden, F. (2006) 'Efficient adaptive density estimation per image pixel for the task of background subtraction', *Pattern Recognition Letters*, Vol. 27, No. 7, pp.773–780 [online] <https://doi.org/10.1016/j.patrec.2005.11.005>.