Mantik Journal, 9 (2) (2025), ISSN 2685-4236 (Online)

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Prisma-based systematic review of video-based AI applications and challenges in multiple domains

Li Yaoqi[1], Febry Putra Rochim[2]

[1,2] Informatics and Computer Engineering Education, Faculty of Engineering, Universitas Negeri Semarang, Indonesia

A R T I C L E I N F O

Article history:

Received Jun 09, 2025
Revised Jun 26, 2025
Accepted Jul 15, 2025

Keywords:

Artificial intelligence;
Video-based AI;
Deep learning;
PRISMA;
Systematic literature review.

ABSTRACT

Video-based artificial intelligence has emerged as a rapidly growing field, driven by advancements in deep learning and the increasing demand for automation across sectors. This study aims to summarize the trends, applications, and major challenges in the implementation of video-based AI using a PRISMA-based systematic literature review approach. The data synthesized from 17 selected articles indicates that deep learning models such as CNNs, LSTMs, and hybrid architectures have been successfully employed for various tasks including anomaly detection, deepfake classification, long-range surveillance, video compression, and video-based educational assessment. Applications span across security, healthcare, education, and entertainment, with notable improvements in efficiency and accuracy. However, challenges remain concerning privacy, algorithmic bias, and the gap between technological progress and regulatory readiness. Hardware demands and variability in model performance also pose limitations. These findings underscore the importance of interdisciplinary approaches to foster responsible and sustainable innovation in video-based AI. The review offers a comprehensive overview that may serve as a foundation for future research directions and technological development.

Corresponding Author:

Li Yaoqi,
Informatics and Computer Engineering Education,
Universitas Negeri Semarang,
Sekaran, Gunung Pati Semarang 50229. Central Java, Indonesia
Email: liyaoqi@students.unnes.ac.id

1. INTRODUCTION

The development of Artificial Intelligence (AI) technology has experienced remarkable growth in recent years(Lestari et al., 2024; Obaido et al., 2024)particularly with the emergence of deep learning approaches that enable machines to replicate human cognitive abilities such as facial recognition, language comprehension, and image or video interpretation (LeCun et al., 2015). One of the increasingly adopted implementations of AI is video-based AI an intelligent system that utilizes video data as the primary source for automated analysis and decision-making (Sharma et al., 2021) It has been widely applied across various domains including security, education, transportation, entertainment, and healthcare(Bin & Nor, n.d.)

In the security sector, for instance, facial recognition technology integrated into video surveillance systems has been employed to detect suspicious activities in real time(Kaur et al., 2024). Meanwhile, in the education field, video-based AI is used to monitor student engagement in online learning and provide automated feedback. These examples underscore the significant potential of this technology in supporting automation and enhancing efficiency across multiple sectors (Badidi et al., 2023c; Bin & Nor, n.d.; Sharma et al., 2021).

However, despite its promising applications, the widespread adoption of video-based AI remains uneven globally. In many developing countries, challenges such as limited access to high-performance computing hardware, low digital literacy, and insufficient regulatory frameworks hinder the equitable implementation of these technologies (Kaur et al., 2024). Moreover, rapid advancements in AI also present new issues related to privacy, algorithmic bias, system transparency, and governance gaps that have yet to keep pace with technological progress (Alabi, 2025; Zou, 2025).

Numerous studies have explored advancements in video-based AI, ranging from the development of deep learning models to the use of convolutional neural networks (CNNs) for feature extraction from video data (Pattichis et al., 2023; Suresha et al., 2020; Vora et al., 2025). However, much of this research is fragmented and tends to focus on specific technical aspects, thus failing to provide a comprehensive and systematic overview of the field.

Therefore, this study conducts a PRISMA-based systematic literature review to map the current trends, applications, and challenges of video-based AI based on the most recent scientific sources. The review aims not only to provide a holistic understanding of the research landscape but also to identify opportunities for responsible and inclusive innovation, particularly in underrepresented regions. By addressing both technical and socio-economic dimensions, this work serves as a foundation for future research and practical deployment of video-based AI solutions across diverse sectors.

2. RESEARCH METHOD

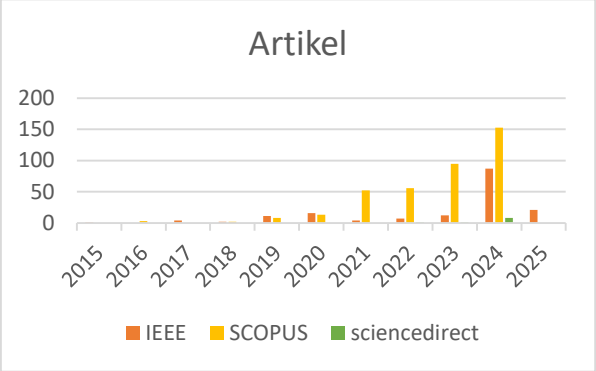
This study adopts a Systematic Literature Review (SLR) approach combined with an exploratory qualitative method. The SLR approach is employed for its capability to identify, evaluate, and synthesize relevant information from existing literature. Meanwhile, the exploratory qualitative method allows for deeper investigation into issues related to video-based AI.

2.1 Research Questions

To provide a clear framework for this literature review, we have identified the following research questions: （a） How has video-based Artificial Intelligence (video-based AI) technology developed in recent years, particularly with the emergence of deep learning approaches that enable machines to replicate human cognitive functions such as facial recognition, language understanding, and the interpretation of images and videos? (b) What are the current applications of video-based AI across various domains such as security, education, transportation, entertainment, and healthcare? (c) What are the key challenges and issues faced in the implementation of video-based AI, including concerns related to privacy, algorithmic bias, system transparency, and regulatory gaps that lag behind technological advancement? (d) How can a systematic literature review contribute to understanding the trends and research directions in video-based AI, as well as identifying development opportunities that could serve as a foundation for researchers and practitioners to create more sustainable and responsible video-based AI innovations?

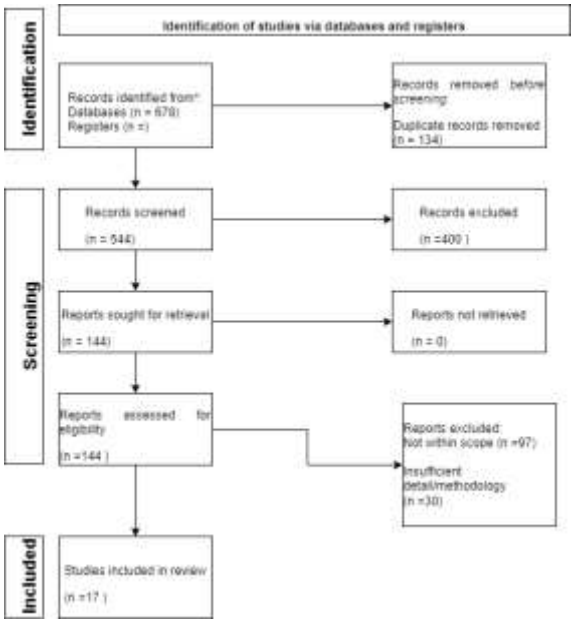
The methodological strategy follows the PRISMA 2020 guidelines, which involve four key stages: identification, screening, eligibility, and inclusion. To minimize selection

and confirmation biases, multiple strategies were implemented. iterature searches were conducted independently by two researchers using Scopus, IEEE Xplore, and ScienceDirect, employing a standardized keyword combination that included domain-related terms (“video-based AI”, “video analysis”), purpose-oriented phrases (“applications”, “real-time processing”), and solution-based terms (“deep learning”, “CNN”, “LSTM”). Duplicate entries were removed using reference management software.



Picture 1. Article range

Based on the identification results over the past ten years, as illustrated in Figure 1, there has been a significant increase in the number of publications in the last five years. Notably, the Scopus database shows a substantial surge beginning in 2021, reaching its peak in 2024 with a total of 153 articles. Similarly, the number of publications from IEEE Xplore has also increased, particularly in 2023 and 2024, with 87 and 21 articles, respectively. In contrast, ScienceDirect contributed only a small number of articles during the observed period one article in both 2022 and 2023, and eight articles in 2024. This trend indicates a sharp rise in interest in the topic of video-based AI in recent years, in line with technological advancements and its growing range of applications across various sectors.



Picture 2. Screening Protocol with PRISMA 2020

Following a systematic selection process based on the PRISMA 2020 guidelines, a total of 678 initial articles were identified from various academic databases. After removing 134 duplicate entries, 544 articles remained for the screening phase. During this stage, an initial review of titles and abstracts was conducted, resulting in the elimination of 400 articles due to irrelevance. Subsequently, 144 articles underwent full-text assessment to determine their eligibility. Of these, 120 articles were excluded, 97 for being outside the scope of the study and 30 for lacking sufficient methodological information. Ultimately, 17 articles were deemed eligible and included in the final review process.

To analyze the data, bibliographic mapping was conducted using VOSviewer software. This tool enables the visualization of conceptual relationships and emerging research trends within the domain of video-based AI. The mapping results revealed strong interconnections among keywords such as *machine learning*, *computer vision*, *real-time*, *implementation*, *framework*, and *challenges*. These findings indicate that research in this field extends beyond technical aspects—such as deep learning and reinforcement learning techniques and addresses practical applications and implementation challenges. Quantitative analysis was employed to examine annual publication trends and the distribution of research domains. Meanwhile, qualitative analysis focused on an in-depth exploration of the challenges highlighted in each article, the technical solutions proposed, and the specific AI techniques applied to address these issues. The combination of quantitative and qualitative approaches was used to provide a comprehensive understanding of the research landscape in video-based AI, including key research focuses, implementation barriers, and future directions for innovation.

Picture 3. Bibliographic mapping related to artificial intelligence

3. RESULTS AND DISCUSSIONS

Table 1. Data Processing

No	Source	Research Title	Finding
1	(Singh et al., 2023)	A Decade of AI and Animation Convergence: A Bibliometric Analysis of Contributions	Bibliometric review of 509 Scopus documents (2013-2023) in AI + animation. Finds rapid growth especially in computer science and engineering with China, USA, Japan, India, and UK leading output. Highlights top-cited works, prolific authors, and emerging themes: deep learning, virtual reality, and HCI. Emphasizes interdisciplinary collaboration and AI's role in enhancing storytelling, simulation, and applications across medicine, education, and scientific visualization.
2	(Singh et al., 2024)	Artificial Intelligence in Gaming: A Bibliometric Analysis of Research Outputs and Trends	Analysis of 950 Scopus publications (2003-2023) on game AI. Identifies IEEE CIG as the main venue and Bulitko, Togelius, Lucas as top authors. The US dominates output, followed by UK and Canada. Core topics include human computer interaction, video games, and deep learning. Notes rising interest in procedural content generation, adaptive gameplay, and interactive

			storytelling.
3	(Goyal et al., 2024)	Cognitive Video Analysis for Anomaly Detection Using Deep Learning"	Deep learning framework tested on 2,000 campus surveillance clips (1,000 normal vs. 1,000 abnormal). Compares ConvLSTM2D (86% acc.), LRCN (90%), and MobileNetV2 (92%). Uses OpenCV frame extraction and spatio-temporal feature fusion. Concludes that combining CNN and RNN architectures yields robust anomaly detection, and suggests future work on transfer learning and real-time IoT integration.
4	(Chittam et al., 2024)	Deep Fake Video Classification with Sequential Input Frames Using Hybrid Deep Learning Model and Bayesian Optimization	Hybrid model: VGG19 feature extractor + stacked LSTM, hyperparameters tuned via Optuna Bayesian optimization. Trained on FaceForensics++ with 20,000 five-frame sequences. Achieved 95.6% accuracy (AUC 0.94) on Deepfakes subset and strong generalization across FaceSwap, Face2Face, and NeuralTextures.
5	(Pinto et al., 2021)	Deep Learning and Multivariate Time Series for Cheat Detection in Video Games	Combines CNNs, GANs, and LSTMs to compress video frames. Incorporates SVD, K-means, and KNN for efficiency. Reduces file size by up to 50% with only ~10% quality loss, enabling adaptive, real-time streaming that can outperform H.264/H.265 in some scenarios.
6	(Khadir et al., 2023)	Deep Learning Empowers Next-Gen Video Compression: Bridging Quality and Efficiency	Two-stage pipeline: motion-based Adaboost detection, then CNN classification via transfer learning (GoogleNet/AlexNet). Achieves ~88–90% accuracy under low resolution, heat haze, and camera shake, suitable for border surveillance.
7	(Wei. H et al., 2018)	Deep Learning-Based Person Detection and Classification for Far Field Video	Two-stage pipeline: motion-based Adaboost detection, then CNN classification via transfer learning (GoogleNet/AlexNet). Achieves ~88–90% accuracy under low resolution, heat haze, and camera shake, suitable for border surveillance.
8	(Sun et al., 2022)	Deep Reinforcement Learning for Video Summarization with Semantic Reward	SV-DSN: unsupervised model fusing visual frames and subtitles with a semantic reward (entropy + topic diversity). Generates more representative summaries for speeches, news, and interviews without labeled data or user input.
9	(Khanam & Roopa, 2025)	Hybrid Deep Learning Models for Anomaly Detection in CCTV Video Surveillance	Hybrid Custom MobileNetV2 + Bi-LSTM on Smart-City CCTV Violence dataset. MobileNet-BiLSTM achieved 94.43% accuracy vs. 90.17% for MobileNetV2 alone, with 14.7 ms/frame inference. Excels in varied lighting and real-time applications.
10	(Badidi et al., 2023a)	Identifying GDPR Privacy Violations Using an Augmented LSTM: Toward an AI-based Violation Alert Systems	Augmented LSTM enriched with Labeled LDA topic features to classify which GDPR article was violated. Evaluated on 750 real incidents; reached 62% accuracy, outperforming TF-IDF and standard LDA baselines.
11	(Badidi et al., 2023b)	Opportunities, Applications, and Challenges of Edge-AI Enabled Video Analytics in Smart Cities: A Systematic Review	This systematic review analyzes the state-of-the-art in Edge-AI video analytics for smart cities. It identifies the benefits of using Edge AI—such as reduced latency, real-time processing, improved privacy, and lower bandwidth usage—compared to traditional cloud-based video analytics. Applications include surveillance, traffic management, healthcare, retail, and education. The study highlights commonly used models (e.g., CNN, YOLO, MobileNet, Mask R-CNN) and privacy-preserving techniques (e.g., encryption, anonymization, federated learning). Challenges include model accuracy, energy consumption, and integration into legacy systems. The paper offers recommendations for future research and emphasizes the transformative potential of Edge-AI video analytics in smart city environments.
12	(Kantarcioglu & Shaon, 2019)	Securing Big Data in the Age of AI	The authors introduce SecureDL, a transparent data-access broker layered atop existing SQL and NoSQL systems (e.g., Hadoop, Spark, HBase) that uniformly enforces access control and privacy policies for AI workloads. SecureDL automatically logs all data requests for governance, sanitizes and redacts sensitive information on-the-fly, detects anomalous or unauthorized access via intrusion-detection techniques, and even generates attribute-based access control policies based on data sensitivity—

			delivering an integrated framework for securing both structured and unstructured big data used in machine learning and AI applications.
13	(Rizwanbasha & Annamalai, 2024a)	Transforming Crime Scene Investigations Through the Integration of Artificial Intelligence in Digital Forensi cs	The study shows that integrating AI into digital forensics slashes the time needed for core tasks—metadata extraction falls from 10 to 2 hours (80% faster), keyword searches drop by 93%, and video analysis time decreases by 73%—while automating feature extraction and report generation reduces manual effort by up to 88% . Accuracy also sees a marked boost: facial recognition rises from 70% to 88% and object detection in videos from 65% to 90% . Moreover, AI enhances the legal admissibility of evidence, with predictive analytics evidence acceptance climbing from 70% to 85%, underscoring AI's role in making crime scene investigations faster, more precise, and more robust in court
14	(Umesiobi et al., 2024a)	Utilising Artificial Intelligence to Augment Physiotherapy Education: A Video-Based Analysis and Grading System for Evaluating Student Therapists' Performance	The authors present an AI-driven system that leverages computer vision (via SlowFast networks) and speech recognition (via OpenAI's Whisper) to objectively analyse and grade physiotherapy students' practical exams. Trained on a richly annotated dataset of video sessions, the tool classifies performance into "Good," "Average," or "Brief" categories, achieving an overall grading accuracy of 89% when compared against expert educator assessments. This multimodal framework not only enhances consistency and fairness in practical skill evaluation but also accelerates feedback delivery, marking a significant step toward integrating AI into healthcare education.
15	(Ermishov & Savchenko, 2024a)	Video-Based Emotional Reaction Intensity Estimation Based on Multimodal Feature Extraction	The authors propose a novel multimodal framework that combines facial embeddings extracted via an EfficientNetB0 backbone (pre-trained on AffectNet), action-unit predictions from the OpenFace toolkit, and audio features generated by the emotion2vec model. These streams are fused through a Transformer Encoder with Multimodal Multi-Head Attention (TEMMA), allowing the system to capture subtle interplays between facial expressions and vocal cues. When evaluated on the ABAW5 Emotional Reaction Intensity Estimation challenge, this architecture achieved third place on the validation leaderboard with an Average Pearson's Correlation Coefficient (APCC) of 0.4038—an improvement of nearly 69% over the baseline model provided by the organizers. Analyzing each modality in isolation revealed that the visual-only pipeline achieved an APCC of 0.3648 (third among all visual submissions), while the best audio-only approach, despite leveraging emotion2vec, was limited to an APCC of 0.1787 due to pervasive background noise in many clips. The consistent superiority of ensemble and fusion strategies over unimodal baselines underscores the critical role of integrating complementary modalities via Transformer-based attention for precise emotion intensity estimation. Finally, by building entirely on open-source tools and publicly available pre-trained models, the authors ensure full reproducibility of their results, and they identify promising directions for future work—such as advanced audio-noise filtering and optimized face-detection routines to accelerate processing without sacrificing accuracy.
16	(Jin, 2024)	Visual Analysis of Artificial Intelligence in Short Video Research Based on Citespace	The study conducts a bibliometric analysis of 509 Scopus-indexed documents published between 2013 and 2023 to explore research at the intersection of artificial intelligence (AI) and animation. It finds that publication output has grown rapidly, particularly within computer science and engineering. China, the United States, Japan, India, and the UK emerge as the top five contributing countries. Highly cited articles and the most prolific authors are identified, and emerging trends—such as deep learning, virtual reality, and human–computer interaction—are highlighted. The findings underscore the importance of interdisciplinary collaboration, showing that AI not only transforms animation technologies but also enriches storytelling and simulation across diverse domains, from medicine and education to scientific visualization. Taken together, the fusion of AI and animation promises to drive creative innovation and

			streamline content production for short-form video in the years ahead.
17	(Zhang et al., 2024)	Examining the Role of Large Language Models in Orthopedics: Systematic Review	This systematic review evaluated 68 studies on the use of large language models (LLMs) such as ChatGPT in orthopedics, categorizing their applications into clinical practice (69%), education (18%), research (12%), and management (1%). Although LLMs showed potential in assisting with diagnosis, documentation, patient education, and orthopedic examinations (e.g., OITE, UKITE), only 8 studies recruited patients and just 1 was an RCT. Performance varied widely: diagnostic accuracy ranged from 55% to 93%, and examination scores (e.g., GPT-4 on OITE) ranged from 45% to 73.6%. The main challenges identified were related to accuracy, readability, reliability, and timeliness of responses. Despite limitations, the review concludes that LLMs may serve as effective "copilots" to enhance efficiency, but are not yet suitable as standalone tools in clinical orthopedic decision-making. Future directions include model fine-tuning, better prompt design, and more rigorous clinical evaluations.

Based on the systematic review of the 17 selected articles, there are significant differences in challenge characteristics between the entertainment sector and sensitive sectors such as healthcare and security. In the entertainment sector, particularly in AI-based gaming and animation domains (Singh et al., 2023, 2024), the main challenges are more related to technical aspects and user experience, such as procedural content development, real-time gameplay adaptation, and the integration of virtual reality and human-computer interaction. Although privacy and data security issues remain a concern, the implications of errors or biases in AI models tend to be less critical compared to sensitive sectors.

In contrast, in the healthcare and security sectors, the challenges faced have much higher ethical, legal, and social dimensions. For example, in the application of video-based AI for digital forensics and crime investigation(Rizwanbasha & Annamalai, 2024b), the accuracy of facial and object identification must be very high due to its direct impact on legal processes and justice. Similarly, in the field of video-based physiotherapy education (Umesiobi et al., 2024b), the precision of automatic evaluation of patient movements is crucial to ensure appropriate medical intervention. On the other hand, the application of AI in emotional reaction analysis(Ermishov & Savchenko, 2024b) presents additional challenges in terms of subjective interpretation and the protection of individuals' psychological data. Therefore, model validation, algorithm transparency, and personal data protection become top priorities in implementation within these sectors.

Moreover, the aspects of sustainability and energy efficiency in the implementation of video-based AI are increasingly relevant to the green digital transformation agenda. Several studies have highlighted the importance of Edge-AI in reducing latency, enhancing privacy, and optimizing bandwidth usage(Badidi et al., 2023a). This technology allows video processing to be performed close to the data source, thereby reducing the need for intensive communication with the cloud, which consumes high amounts of energy. Additionally, some research demonstrates the use of optimized CNN and MobileNet models for video analysis on a smart city scale, which not only improves performance but also contributes to energy savings by reducing the computational load on data centers. In the future, the development of environmentally friendly video-based AI will be key to supporting global initiatives in carbon emission reduction and achieving sustainable development goals (SDGs), especially in the implementation of smart cities and digitally based public services.

4. CONCLUSION

The exploration of video-based Artificial Intelligence highlights a dynamic and interdisciplinary research landscape with significant practical applications. Deep learning models, such as CNNs, LSTMs, and their hybrids, have driven advancements in anomaly detection, video compression, surveillance, deepfake classification, and educational assessments. These innovations have benefited sectors like security, healthcare, education, and entertainment by improving automation, accuracy, and real-time responsiveness. However, challenges persist, including privacy risks, algorithmic bias, computational demands, and the gap between technological progress and regulatory readiness. To address these gaps, strategies such as fostering collaboration between researchers and policymakers, promoting transparency in AI algorithms, and harmonizing global governance frameworks are essential. Furthermore, this study's findings align with international AI ethics policies, such as the EU's AI Act and UNESCO principles, particularly through privacy-preserving techniques and addressing algorithmic bias. Despite these challenges, the continued momentum in research and development suggests a promising future for scalable, sustainable, and socially impactful video-based AI.

REFERENCES

Alabi, M. (2025). *Ethical Challenges in AI: Addressing Bias, Privacy, and Accountability*. https://www.researchgate.net/publication/390582838

Badidi, E., Moumane, K., & Ghazi, F. El. (2023a). Opportunities, Applications, and Challenges of Edge-AI Enabled Video Analytics in Smart Cities: A Systematic Review. In *IEEE Access* (Vol. 11, pp. 80543–80572). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2023.3300658

Badidi, E., Moumane, K., & Ghazi, F. El. (2023b). Opportunities, Applications, and Challenges of Edge-AI Enabled Video Analytics in Smart Cities: A Systematic Review. In *IEEE Access* (Vol. 11, pp. 80543–80572). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2023.3300658

Badidi, E., Moumane, K., & Ghazi, F. E. (2023c). Opportunities, Applications, and Challenges of Edge-AI Enabled Video Analytics in Smart Cities: A Systematic Review. *IEEE Access*, *11*, 80543–80572. https://doi.org/10.1109/ACCESS.2023.3300658

Bin, R., & Nor, M. (n.d.). *AI Applications in Education, Healthcare, and Transportation Trends, Challenges, and Future Directions*.

Chittam, S., Johnson, D., Roy, K., & Yuan, X. (2024). Deep Fake Video Classification with Sequential Input Frames Using Hybrid Deep Learning Model and Bayesian Optimization. *Proceedings - 2024 International Conference on Machine Learning and Applications, ICMLA 2024*, 1432–1438. https://doi.org/10.1109/ICMLA61862.2024.00222

Ermishov, N., & Savchenko, A. V. (2024a). Video-Based Emotional Reaction Intensity Estimation Based on Multimodal Feature Extraction. *RusAutoCon - Proceedings of the International Russian Automation Conference, 2024*, 838–842. https://doi.org/10.1109/RusAutoCon61949.2024.10693960

Ermishov, N., & Savchenko, A. V. (2024b). Video-Based Emotional Reaction Intensity Estimation Based on Multimodal Feature Extraction. *RusAutoCon - Proceedings of the International Russian Automation Conference, 2024*, 838–842. https://doi.org/10.1109/RusAutoCon61949.2024.10693960

Goyal, G., Gupta, R., Vashishtha, S., Singh, C. K., Aqdas, H., & Garg, H. (2024). Cognitive Video Analysis for Anomaly Detection Using Deep Learning. *2024 International Conference on Intelligent Systems for Cybersecurity, ISCS 2024*. https://doi.org/10.1109/ISCS61804.2024.10581360

Jin, Y. (2024). Visual Analysis of Artificial Intelligence in Short Video Research Based on Citespace. *Proceedings - 2024 3rd International Conference on Data Analytics, Computing and Artificial Intelligence, ICDACAI 2024*, 126–130. https://doi.org/10.1109/ICDACAI65086.2024.00031

Kantarcioglu, M., & Shaon, F. (2019). Securing big data in the age of AI. *Proceedings - 1st IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, TPS-ISA 2019*, 218–220. https://doi.org/10.1109/TPS-ISA48467.2019.00035

Kaur, A., Noori Hoshyar, A., Saikrishna, V., Firmin, S., & Xia, F. (2024). Deepfake video detection: challenges and opportunities. *Artificial Intelligence Review*, *57*(6). https://doi.org/10.1007/s10462-024-10810-6

Khadir, M., Hashmi, M. F., & Khadir, M. (2023). Deep Learning Empowers Next-Gen Video Compression: Bridging Quality and Efficiency. *2023 24th International Arab Conference on Information Technology, ACIT 2023*. https://doi.org/10.1109/ACIT58888.2023.10453895

Khanam, M. H., & Roopa, R. (2025). Hybrid Deep Learning Models for Anomaly Detection in CCTV Video Surveillance. *4th International Conference on Sentiment Analysis and Deep Learning, ICSADL 2025 - Proceedings*, 1345–1351. https://doi.org/10.1109/ICSADL65848.2025.10933441

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lestari, I. S., Jumadi, J., & Lukman, N. (2024). IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK DENGAN PRE-TRAINED MODEL MOBILENETV2 UNTUK DETEKSI KOLESTEROL. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, *9*(2), 173–183. https://doi.org/10.36341/rabit.v9i2.4732

Obaido, G., Mienye, I. D., Egbelowo, O. F., Emmanuel, I. D., Ogunleye, A., Ogbuokiri, B., Mienye, P., & Aruleba, K. (2024). Supervised machine learning in drug discovery and development: Algorithms, applications, challenges, and prospects. *Machine Learning with Applications*, *17*, 100576. https://doi.org/10.1016/j.mlwa.2024.100576

Pattichis, M. S., Jatla, V., & Cerna, A. E. U. (2023). *A Review of Machine Learning Methods Applied to Video Analysis Systems*. http://arxiv.org/abs/2312.05352

Pinto, J. P., Pimenta, A., & Novais, P. (2021). Deep Learning and Multivariate Time Series for Cheat Detection in Video Games. *2021 IEEE 8th International Conference on Data Science and Advanced Analytics, DSAA 2021*. https://doi.org/10.1109/DSAA53316.2021.9564219

Rizwanbasha, A., & Annamalai, R. (2024a). Transforming Crime Scene Investigations Through the Integration of Artificial Intelligence in Digital Forensics. *2024 IEEE International Conference on Communication, Computing and Signal Processing, IICCCS 2024*. https://doi.org/10.1109/IICCCS61609.2024.10763868

Rizwanbasha, A., & Annamalai, R. (2024b). Transforming Crime Scene Investigations Through the Integration of Artificial Intelligence in Digital Forensics. *2024 IEEE International Conference on Communication, Computing and Signal Processing, IICCCS 2024*. https://doi.org/10.1109/IICCCS61609.2024.10763868

Sharma, V., Gupta, M., Kumar, A., & Mishra, D. (2021). Video Processing Using Deep Learning Techniques: A Systematic Literature Review. In *IEEE Access* (Vol. 9, pp. 139489–139507). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2021.3118541

Singh, D. K., Kumar, A., Suri, D., Kumar, P., Sharma, R., & Sood, N. (2024). Artificial Intelligence in Gaming: A Bibliometric Analysis of Research Outputs and Trends. *2024 International Conference on Integration of Emerging Technologies for the Digital World, ICIETDW 2024*. https://doi.org/10.1109/ICIETDW61607.2024.10939117

Singh, D. K., Kumar, P., Sharma, M., & Arora, P. (2023). A Decade of AI and Animation Convergence: A Bibliometric Analysis of Contributions. *2023 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0, IDICAIEI 2023*. https://doi.org/10.1109/IDICAIEI58380.2023.10406710

Sun, H., Zhu, X., & Zhou, C. (2022). Deep Reinforcement Learning for Video Summarization with Semantic Reward. *Proceedings - 2022 IEEE 22nd International Conference on Software Quality, Reliability and Security Companion, QRS-C 2022*, 754–755. https://doi.org/10.1109/QRS-C57518.2022.00119

Suresha, M., Kuppa, S., & Raghukumar, D. S. (2020). A study on deep learning spatiotemporal models and feature extraction techniques for video understanding. *International Journal of Multimedia Information Retrieval*, *9*(2), 81–101. https://doi.org/10.1007/s13735-019-00190-x

Umesiobi, V., Alsmadi, H., Hammond, V., Kandasamy, G., & Al Kafri, A. S. (2024a). Utilising Artificial Intelligence to Augment Physiotherapy Education: A Video-Based Analysis and Grading System for Evaluating Student Therapists' Performance. *Proceedings - International*

Conference on Developments in ESystems Engineering, DeSE, 219–224. https://doi.org/10.1109/DeSE63988.2024.10911886

Umesiobi, V., Alsmadi, H., Hammond, V., Kandasamy, G., & Al Kafri, A. S. (2024b). Utilising Artificial Intelligence to Augment Physiotherapy Education: A Video-Based Analysis and Grading System for Evaluating Student Therapists' Performance. *Proceedings - International Conference on Developments in ESystems Engineering, DeSE*, 219–224. https://doi.org/10.1109/DeSE63988.2024.10911886

Vora, D., Kadam, P., Mohite, D. D., Kumar, N., Kumar, N., Radhakrishnan, P., & Bhagwat, S. (2025). AI-driven video summarization for optimizing content retrieval and management through deep learning techniques. *Scientific Reports*, *15*(1), 4058. https://doi.org/10.10438/s41598-025-87824-9

Wei. H, Laszewki. M, & Kehtarnavas. N. (2018). *Deep Learning-Based Person Detection and Classification for Far Field Video Surveillance*. IEEE.

Zhang, C., Liu, S., Zhou, X., Zhou, S., Tian, Y., Wang, S., Xu, N., & Li, W. (2024). Examining the Role of Large Language Models in Orthopedics: Systematic Review. In *Journal of medical Internet research* (Vol. 26, p. e59607). https://doi.org/10.2196/59607

Zou, R. (2025). Research on Ethical Issues, Data Privacy Protection, Algorithmic Bias, and Regulatory Policy of Artificial Intelligence Technology in Digital Transformation. In X. Li, C. Yuan, & L. Vartiak (Eds.), *Proceedings of the 8th International Conference on Economic Management and Green Development* (pp. 234–242). Springer Nature Singapore.