International Journal of Infrastructure Research and Management
Vol. 11 (2), December 2023, pp. 85 - 96

EXPLORATORY REVIEW OF DEEP LEARNING IN HUMAN ACTION RECOGNITION: ADVANCING CURRENT MECHANISMS

Liang Zhantu [12], Jin Yuanrong[1], Li Zhenxiang[1], Sun xuekai[1] & Tadiwa Elisha Nyamasvisva[1]
[1]*Infrastructure University Kuala Lumpur, MALAYSIA*
[2] *Dongguan University of Technology, China*

ABSTRACT
Human action recognition, integral to areas like surveillance, robotics, and human-computer interaction, is a challenging endeavor. This paper presents an in-depth literature review of the application of deep learning technology in human behavior recognition. We aim to shed light on the potential of employing deep neural networks, notably Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long-Short-Term Memories (LSTMs), to capture intricate spatiotemporal patterns inherent to human actions. An exposition on traditional methods of human action recognition is provided, underscoring their constraints in dealing with variability, complexity, and imperfect data. Grounded on this foundation, the paper discusses how deep learning models might circumvent these challenges. Through comparative analysis with existing methods, we highlight potential advantages of deep learning models in recognition accuracy, computational efficiency, and resilience to data variations. The essence of this review is not only to elucidate the capabilities of deep learning techniques in human action analysis but also to offer insights beneficial for researchers and professionals aiming to design precise and efficient human action recognition frameworks. The central contribution of this work lies in its synthesis of current methodologies, emphasizing the promise and potential of deep learning in advancing the field.

Keywords:
human action recognition, deep learning, convolutional neural networks, recurrent neural networks, spatiotemporal patterns

INTRODUCTION

Human action recognition has emerged as a critical research area in recent years due to its wide-ranging applications in fields such as sports analysis (Lai et al., 2022; Meng et al., 2020; Qiu et al., 2022), healthcare monitoring (Fridriksdottir & Bonomi, 2020; Lin et al., 2022; Paul et al., 2021), robotics (Gong et al., 2022; Luo et al., 2022; Moon & Seo, 2022), security (Arabi et al., 2022; Nyamasvisva & Arabi, 2022.) and video surveillance (Elharrouss et al., 2021; Khan et al., 2020; Majid et al., 2022; Nasir et al., 2022). The ability to recognize and understand human action patterns is essential for enabling advancements in these applications accurately and efficiently. Traditional approaches to human action recognition, including rule-based systems (Diao et al., 2022; Yang et al., 2021), template matching (Afza et al., 2021; Shahzad & Jalal, 2021), Hidden Markov Models (HMMs) (Elangovan, 2021; Ghadi et al., 2022), and Dynamic Time Warping (DTW) (Ram et al., 2023; Trelinski & Kwolek, 2021), frequently rely on hand-crafted features and constrained classification algorithms, which may not correctly reflect the complexity and variety of human action. However, with the advent of deep learning techniques, such as CNNs (Du & Mukaidani, 2022; Slade et al., 2022), RNNs (Loh et al., 2022; Zhang et al., 2022) and LSTM (li & Cao, 2023; X. Shen & Ding, 2022), significant progress has been made in the field of computer vision, leading to remarkable improvements in various tasks, including image recognition, object detection, and natural language processing.

The goal of this research is to overcome the drawbacks of conventional methods and create systems that are more reliable and accurate by utilizing the power of deep learning to improve human action recognition algorithms. It is anticipated that by taking advantage of deep learning models' capacity to automatically learn discriminative features from raw data, these models will be able to capture complex action patterns, adapt to varying environmental conditions, and provide accurate recognition results even in the presence of noise or erroneous action data. The study subject, research motivation, problem statement, and conclusion are all introduced in this chapter.

RESEARCH MOTIVATION

The motivation for doing this research derives from the growing requirement for powerful and accurate human action recognition systems across multiple domains. Traditional approaches frequently rely on hand-crafted features and rule-based algorithms, which cannot capture the intricate intricacies and nuances of human activity (Sharma et al., 2022; X. Wang et al., 2023). As a result, the performance and applicability of these approaches in real-world circumstances is limited (Islam et al., 2022). Deep learning approaches have shown improved skills in dealing with complicated data representations and learning hierarchical features, making them well-suited for applications like as picture identification and natural language processing (Gaafar et al., 2022). By applying deep learning to human action recognition, there is the potential to revolutionize the field by achieving significant improvements in accuracy, efficiency, and adaptability.

Furthermore, the introduction of low-cost, high-quality action capture equipment, such as inertial sensors, depth cameras, and wearable devices, has supplied researchers with a wealth of data to study human behavior. Deep learning algorithms applied to these datasets have the potential to extract meaningful representations and patterns, leading to improved understanding and recognition of human action (C. Shen et al., 2022). Deep learning algorithms are becoming more popular and feasible as processing power and large-scale computer resources improve (Lauriola et al., 2022).

STATEMENT OF THE PROBLEM

The problem addressed by this research is the urgent need to develop advanced methods using deep learning techniques to overcome the limitations of traditional human action recognition systems. These systems often struggle to accurately and efficiently recognize and classify complex human behavior patterns in real-time under various environmental conditions (Arshad et al., 2022; Najeh et al., 2022; Yalçinkaya et al., 2023).

Traditional methods, such as template matching, have limitations in handling the inherent variability and complexity of human action (Saleem et al., 2023). Template matching relies on predefined templates for each action, which makes it difficult to handle variations in appearance, viewpoint, and timing (Pham et al., 2022). Moreover, template matching is sensitive to noise and occlusions, resulting in reduced accuracy and robustness (Solorzano & Tsai, 2023).

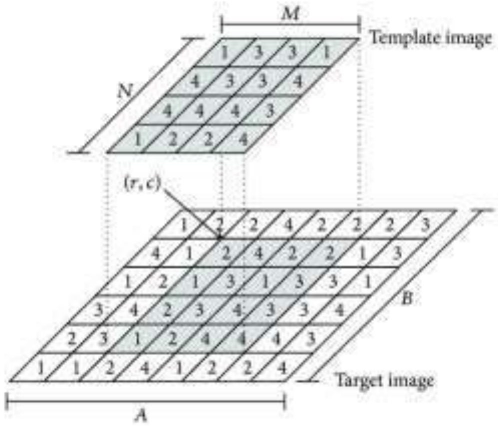

Figure 1: Object detection based on Template matching through use of best-so-far ABC (Banharnsakun & Tanathong, 2014)

Hidden Markov Models (HMMs) assume conditional independence between observations given the hidden states, which cannot hold for complex human actions (Lorek et al., 2022; Rojas-Salazar, 2022). They struggle to capture long-term dependencies and can be sensitive to noisy or incomplete data (Ahrends & Vidaurre, 2023). Additionally, acquiring labeled training data for HMMs can be a time-consuming and costly process (Şengönül et al., 2023).

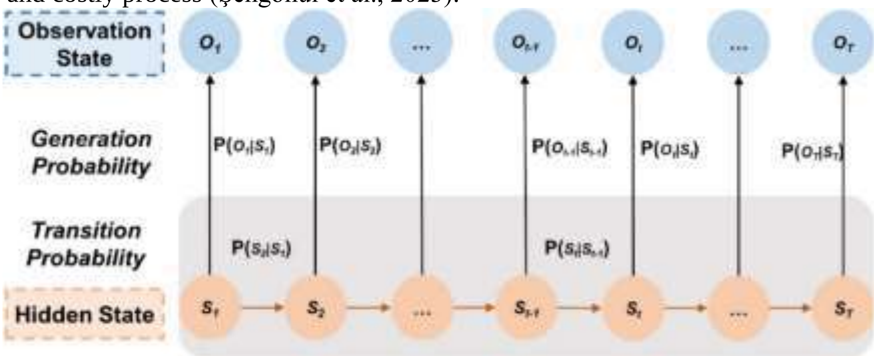

Figure 2: A fusion of a deep neural network and a hidden Markov model to recognize the multiclass abnormal behavior of elderly people (L. Wang et al., 2022)

Dynamic Time Warping (DTW) requires a predefined template sequence and is less effective for recognizing novel or complex actions (Singh et al., 2021). It is computationally expensive, particularly with large datasets or long sequences (Sedmidubsky et al., 2021). Furthermore, DTW does not capture the underlying structure or semantics of actions, limiting its ability to provide meaningful insights into the action patterns (Cortiñas-Lorenzo & Lacey, 2023).

Recurrent Neural Networks (RNNs) have difficulty accurately capturing long-term dependencies due to the vanishing gradient problem (Giovanni et al., 2023). To generalize effectively, they require a substantial quantity of labelled training data, which can be laborious and costly to acquire (Shahabi et al., 2022). RNNs are sensitive to noisy and incomplete data, as well as viewpoint and illumination variations (Deotale et al., 2022). Given these limitations, it is imperative to develop more effective and efficient systems for recognizing human actions. Deep learning techniques, such as CNNs

(Bilal et al., 2022; Javed et al., 2022; Lu et al., 2022; Uchiyama et al., 2023) and LSTM networks (Bansal & Chandra, 2022; Domingo et al., 2022; Sahoo et al., 2022; Stein et al., 2022), have shown promising results in various domains and can potentially address the challenges faced by traditional methods.

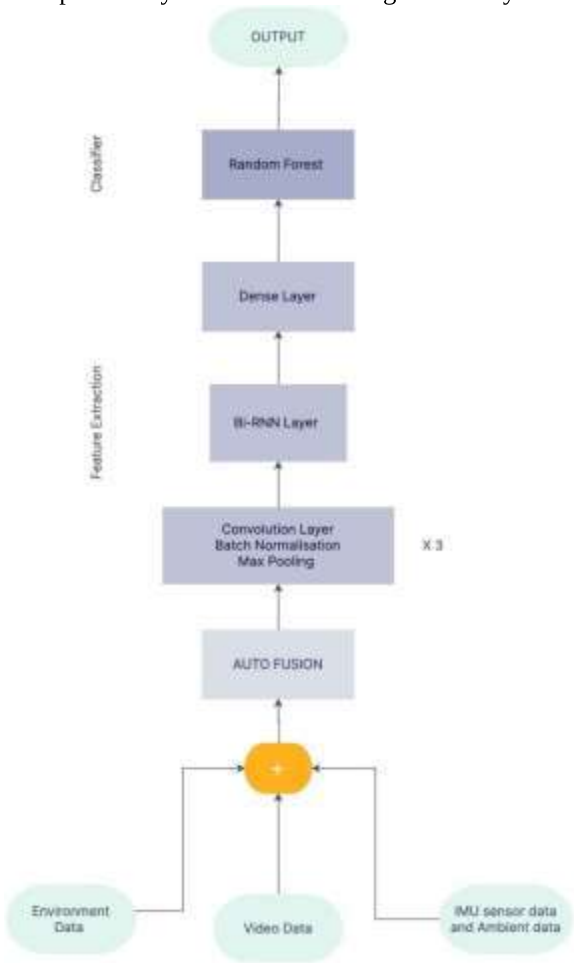

Figure 3: Ambient intelligence-based multimodal human action recognition for autonomous systems (Jain et al., 2023)

Researchers can capture meaningful representations of human action patterns by utilizing CNNs, which are designed to extract spatial features from images or video frames (Sun et al., 2022). This allows the system to effectively manage variations in appearance and perspective. In addition, CNNs can acquire features automatically from data, reducing the need for predefined templates (Tang et al., 2022). On the other hand, LSTM networks are explicitly designed to model temporal dependencies and long-term dependencies in sequential data (Al-Selwi et al., 2023). By integrating LSTM networks into human action recognition systems, the temporal dynamics of human actions can be precisely captured (Zhong et al., 2022). LSTMs can effectively manage noisy and incomplete data, making them suitable for real-world scenarios where these challenges are prevalent (Choudhury & Soni, 2023).

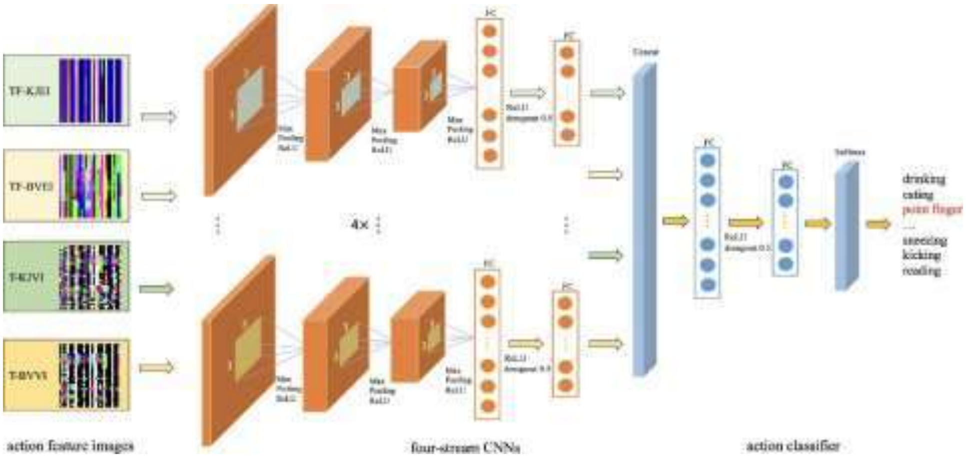

Figure 4: AFE-CNN: 3D Skeleton-based Action Recognition with Action Feature Enhancement (Guan et al., 2022)

Significant quantities of labelled training data are still required to train deep learning models. Recent advancements in data augmentation techniques, such as generative adversarial networks (GANs) (Qu et al., 2022), have made it possible to generate synthetic training data, thereby reducing reliance on large-scale manually labelled datasets. In addition, transfer learning and domain adaptation methods can be used to leverage pre-trained models on massive action datasets. This method permits the transfer of knowledge from related tasks or domains, allowing for more efficient training and improved generalization (Huang et al., 2022).

Table 1: SWOT Analysis of existing deep learning-based methodologies

	Algorithm	Strengths	Weaknesses	Opportunities	Threats
1	Hidden Markov Models	Assume conditional independence between observations given the hidden states	Can not hold for complex human actions	HMMs can model sequential data by capturing the underlying states and transitions between them. In the context of action recognition, HMMs can represent actions as sequences of states, allowing for the recognition of different action stages or sub-actions.	HMMs rely on the availability of labeled training data, which is labor-intensive and challenging to obtain for some action classes, and the computational complexity of training and inference is high
2	Dynamic Time Warping	It can lie in its flexibility to handle sequences of different lengths and speeds, its robustness to noise and variations within sequences, and its ability to perform nonlinear alignment, capturing complex temporal relationships.	It rely on a predefined template sequence, limited effectiveness in recognizing novel or complex actions, computational expense with large datasets or long sequences, and inability to capture the underlying structure or semantics of actions.	It allows for the comparison of actions with varying speeds or durations, making it suitable for recognizing actions with flexible timing.	DTW requires pairwise comparisons between sequences, resulting in high computational cost, especially for large datasets

3	Recurrent Neural Networks	Can model a collection of records (i.e. time collection) so that each pattern can be assumed to be dependent on previous ones	Suffer from the vanishing gradient problem, where gradients diminish exponentially over time, limiting their ability to capture long-term dependencies accurately.	RNNs can assign labels to individual frames or segments, allowing for detailed analysis of specific parts or sub-actions. By leveraging transfer learning and combining with other architectures, RNNs can learn from large datasets and fine-tune on specific human action recognition tasks, leading to improved performance.	RNNs in human action recognition face certain threats. They can struggle when the input data is noisy or incomplete, which can negatively impact their performance. Variations in factors like viewpoint or lighting conditions can introduce noise or distort the sequential patterns that RNNs rely on.
4	Convolutional Neural Networks	Can automatically learn features from data, reducing the reliance on predefined templates	CNNs can be sensitive to variations in viewpoint and lighting conditions, which affect their ability to generalize well across different action scenarios.	Can learn hierarchical representations of actions, enabling to recognize complex patterns and distinguish between different action categories	CNNs struggle with recognizing actions that involve fine-grained movements or subtle variations, as they primarily focus on capturing spatial features rather than temporal dynamics.
5	Long Short-Term Memory	Specifically designed to model temporal dependencies and long-term dependencies in sequential data	More complicated than traditional RNNs and require more training data in order to learn effectively. They are not well-suited for online learning tasks, such as prediction or classification tasks where the input data is not a sequence. LSTMs can be slow to train on large datasets	LSTMs can effectively handle noisy and incomplete data, making them suitable for real-world scenarios where such challenges are prevalent.	LSTMs struggle with Gradients decrease exponentially over time, hindering the ability to accurately capture long-term dependencies

As shown in Table 1, in the domain of human action recognition, while traditional algorithms such as HMM and DTW have demonstrated proficiency in certain areas, they face limitations when handling large-scale data, intricate actions, and varied environmental conditions. On the other hand, deep learning approaches, specifically RNNs, CNNs, and LSTMs, have showcased their superiority in the field of human action recognition.

Firstly, compared to HMMs and DTWs, RNNs and LSTMs are especially adept at processing extended time sequences and capturing long-term dependencies. LSTMs, in particular, were designed to address the vanishing gradient problem associated with RNNs. Additionally, CNNs excel at automatically learning features from data, reducing the reliance on predefined templates and offering a stark contrast to the constraints of DTW.

While deep learning algorithms have their set of challenges, such as sensitivity to data and training demands on extensive datasets, they bring unparalleled flexibility and accuracy to action recognition. Notably, as demonstrated by LSTMs, deep learning offers robust advantages when dealing with noisy or incomplete data.

THE PROPOSAL

In order to address variability, complexity, and chaotic or incomplete data in human action recognition, CNN or LSTM network structures are recommended. Utilizing CNNs or LSTMs requires a multistep process. The initial stage involves data preprocessing, which includes tasks such as resizing frames, normalizing pixel values, and dealing with missing or noisy data. Utilizing convolutional and pooling layers, which facilitate the acquisition of hierarchical representations, a CNN is then employed to extract spatial features from individual frames. Additionally, LSTM or other recurrent layers are incorporated to capture temporal dependencies and effectively model the dynamics inherent in action sequences.

The fusion of spatial and temporal information is achieved by establishing connections between the output of the CNN and the LSTM layers. During the training phase, the dataset is partitioned, and a suitable loss function is optimized through the use of an optimization algorithm, while applying regularization techniques like dropout or L2 regularization to mitigate overfitting. Augmenting the training data with diverse transformations is imperative to enhance the dataset's variability and adaptability, and when labeled data is limited, transfer learning techniques can be employed. Lastly, the model's performance is evaluated on an independent test set, employing appropriate metrics for assessment. The attainment of optimal performance necessitates a diligent process of experimentation and fine-tuning to cater to the specific characteristics of the dataset and the requirements of the task.

CONCLUSION

The development of advanced methodologies that leverage deep learning techniques, such as CNNs and LSTMs, can address the limitations of traditional human action recognition systems by improving accuracy, enabling automation, facilitating multimodal integration, and contributing to future technological advancements. These techniques offer the potential for more accurate, robust, and adaptable solutions for recognizing and classifying complex human action patterns and under various environmental conditions such as less light exposure, device noise, camera aperture. By overcoming the challenges of noisy and incomplete data, variations in viewpoint, and long-term dependencies, these methodologies pave the way for significant advancements in human action recognition research and applications. This paper therefore seeks to design an algorithm based on deep learning to improve the accuracy for human action recognition.

Successful research will aid in various fields and applications. In the field of surveillance and security, accurate human action recognition can improve the detection and classification of suspicious or abnormal behaviors, enhancing public safety and security. In healthcare and rehabilitation, it can assist in monitoring patients' physical activities, evaluating their functional abilities, and designing personalized therapy programs. Moreover, in sports and performance analysis, human action recognition enables the assessment of athletes' technique and form, providing insights for training optimization. Additionally, in the realm of human-computer interaction, advancements in action recognition can enhance gesture-based interfaces, enabling more intuitive and natural interactions between humans and computers. Furthermore, it can contribute to the development of robotics, enabling robots to understand and imitate human actions, enhancing their capabilities in various domains. Overall, successful research in human action recognition holds significant potential to benefit these areas and drive technological advancements.

AUTHOR BIOGRAPHY

Liang Zhantu is a Ph.D. in IT candidate at IUKL. His research direction is in human action recognition in the field of artificial intelligence. Liang Zhantu is currently working as an Information and Technology teacher at DGUT in China. *Email: 223923795@s.iukl.edu.my*

Jin Yuanrong is student of the postgraduate programme PhD (Information Technology) at Infrastructure University Kuala Lumpur (IUKL) Faculty of Engineering, Science and Technology. Her research interests include Cloud Computing and Load Prediction algorithms. *Email: 222923382@s.iukl.edu.my*

Li Zhenxiang is student of the postgraduate programme PhD (Information Technology) at Infrastructure University Kuala Lumpur (IUKL) Faculty of Engineering, Science and Technology. His research interests include Blockchain and Cloud Computing. *Email: 222923380@s.iukl.edu.my*

Sun xuekai is student of the postgraduate programme PhD (Information Technology) at Infrastructure University Kuala Lumpur (IUKL) Faculty of Engineering, Science and Technology. His research interests include the application of blockchain technology in smart education. *Email: 223923752@s.iukl.edu.my*

Tadiwa Elisha Nyamasvisva, PhD is a faculty member at the Faculty of Engineering and Science Technology in IUKL. His research interests are in Computer Algorithms, Data Analysis, Networking and Network Security, IT in Education. He is currently the Deputy Dean at the Center of Postgraduate Studies in IUKL and also Head of Postgraduate Programme (HOPP) for PhD in Information Technology and Master in Information Technology (MIT) programmes at the Centre of Postgraduate Studies in IUKL, *Email: tadiwa.elisha@iukl.edu.my*

REFERENCES

Afza, F., Khan, M. A., Sharif, M., Kadry, S., Manogaran, G., Saba, T., Ashraf, I., & Damaševičius, R. (2021). A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection. *Image and Vision Computing*, *106*, 104090. https://doi.org/10.1016/j.imavis.2020.104090

Ahrends, C., & Vidaurre, D. (2023). Dynamic Functional Connectivity. *ArXiv Preprint ArXiv:2301.03408*.

Al-Selwi, S. M., Hassan, M. F., Abdulkadir, S. J., & Muneer, A. (2023). LSTM Inefficiency in Long-Term Dependencies Regression Problems. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *30*(3), Article 3. https://doi.org/10.37934/araset.30.3.1631

Arabi, A. A. M., Nyamasvisva, T. E., & Valloo, S. (n.d.). ZERO TRUST SECURITY IMPLEMENTATION CONSIDERATIONS IN DECENTRALISED NETWORK RESOURCES FOR INSTITUTIONS OF HIGHER LEARNING. *International Journal of Infrastructure Research and Management Vol. 10 (1), June 2022*.

Arshad, M. H., Bilal, M., & Gani, A. (2022). Human Activity Recognition: Review, Taxonomy and Open Challenges. *Sensors*, *22*(17), 6463.

Banharnsakun, A., & Tanathong, S. (2014). Object Detection Based on Template Matching through Use of Best-So-Far ABC. *Computational Intelligence and Neuroscience*, *2014*, e919406. https://doi.org/10.1155/2014/919406

Bansal, N., & Chandra, S. (2022). Solving basic and advanced human activities using LSTM. *Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing*, 204–207.

Bilal, M., Maqsood, M., Yasmin, S., Hasan, N. U., & Rho, S. (2022). A transfer learning-based efficient spatiotemporal human action recognition framework for long and overlapping action classes. *The Journal of Supercomputing*, *78*(2), 2873–2908.

Choudhury, N. A., & Soni, B. (2023). An Adaptive Batch Size based-CNN-LSTM Framework for Human Activity Recognition in Uncontrolled Environment. *IEEE Transactions on Industrial Informatics*.

Cortiñas-Lorenzo, K., & Lacey, G. (2023). Toward Explainable Affective Computing: A Review. *IEEE Transactions on Neural Networks and Learning Systems*.

Deotale, D., Verma, M., Suresh, P., & Kotecha, K. (2022). Optimized hybrid RNN model for human activity recognition in untrimmed video. *Journal of Electronic Imaging*, *31*(5), 051409–051409.

Diao, H., Lu, Y., Deng, A., Zou, L., Li, X., & Pedrycz, W. (2022). Convolutional rule inference network based on belief rule-based system using an evidential reasoning approach. *Knowledge-Based Systems*, *237*, 107713. https://doi.org/10.1016/j.knosys.2021.107713

Domingo, J. D., Gómez-García-Bermejo, J., & Zalama, E. (2022). Improving human activity recognition integrating lstm with different data sources: Features, object detection and skeleton tracking. *IEEE Access*, *10*, 68213–68230.

Du, Z., & Mukaidani, H. (2022). Linear dynamical systems approach for human action recognition with dual-stream deep features. *Applied Intelligence*, *52*(1), 452–470. https://doi.org/10.1007/s10489-021-02367-6

Elangovan, V. (2021). *Indoor Group Activity Recognition using Multi-Layered HMMs* (arXiv:2101.10857). arXiv. https://doi.org/10.48550/arXiv.2101.10857

Elharrouss, O., Almaadeed, N., Al-Maadeed, S., Bouridane, A., & Beghdadi, A. (2021). A combined multiple action recognition and summarization for surveillance video sequences. *Applied Intelligence*, *51*(2), 690–712. https://doi.org/10.1007/s10489-020-01823-z

Fridriksdottir, E., & Bonomi, A. G. (2020). Accelerometer-Based Human Activity Recognition for Patient Monitoring Using a Deep Neural Network. *Sensors*, *20*(22), Article 22. https://doi.org/10.3390/s20226424

Gaafar, A. S., Dahr, J. M., & Hamoud, A. K. (2022). Comparative Analysis of Performance of Deep Learning Classification Approach based on LSTM-RNN for Textual and Image Datasets. *Informatica*, *46*(5), Article 5. https://doi.org/10.31449/inf.v46i5.3872

Ghadi, Y. Y., Javeed, M., Alarfaj, M., Shloul, T. A., Alsuhibany, S. A., Jalal, A., Kamal, S., & Kim, D.-S. (2022). MS-DLD: Multi-Sensors Based Daily Locomotion Detection via Kinematic-Static Energy and Body-Specific HMMs. *IEEE Access*, *10*, 23964–23979. https://doi.org/10.1109/ACCESS.2022.3154775

Giovanni, F. D., Giusti, L., Barbero, F., Luise, G., Lio, P., & Bronstein, M. M. (2023). On Over-Squashing in Message Passing Neural Networks: The Impact of Width, Depth, and Topology. *Proceedings of the 40th International Conference on Machine Learning*, 7865–7885. https://proceedings.mlr.press/v202/di-giovanni23a.html

Gong, L., Chen, B., Xu, W., Liu, C., Li, X., Zhao, Z., & Zhao, L. (2022). Motion Similarity Evaluation between Human and a Tri-Co Robot during Real-Time Imitation with a Trajectory Dynamic Time Warping Model. *Sensors*, *22*(5), Article 5. https://doi.org/10.3390/s22051968

Guan, S., Lu, H., Zhu, L., & Fang, G. (2022). AFE-CNN: 3D Skeleton-based Action Recognition with Action Feature Enhancement. *Neurocomputing*, *514*, 256–267. https://doi.org/10.1016/j.neucom.2022.10.016

Huang, H., Wu, R., Li, Y., & Peng, C. (2022). Self-supervised transfer learning based on domain adaptation for benign-malignant lung nodule classification on thoracic CT. *IEEE Journal of Biomedical and Health Informatics*, *26*(8), 3860–3871.

Islam, Md. M., Nooruddin, S., Karray, F., & Muhammad, G. (2022). Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects. *Computers in Biology and Medicine*, *149*, 106060. https://doi.org/10.1016/j.compbiomed.2022.106060

Jain, V., Gupta, G., Gupta, M., Sharma, D. K., & Ghosh, U. (2023). Ambient intelligence-based multimodal human action recognition for autonomous systems. *ISA Transactions*, *132*, 94–108. https://doi.org/10.1016/j.isatra.2022.10.034

Javed, M. H., Yu, Z., Li, T., Rajeh, T. M., Rafique, F., & Waqar, S. (2022). Hybrid two-stream dynamic CNN for view adaptive human action recognition using ensemble learning. *International Journal of Machine Learning and Cybernetics*, 1–10.

Khan, M. A., Javed, K., Khan, S. A., Saba, T., Habib, U., Khan, J. A., & Abbasi, A. A. (2020). Human action recognition using fusion of multiview and deep features: An application to video surveillance. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-020-08806-9

Lai, Z., Xu, J., Bowen, C. R., & Zhou, S. (2022). Self-powered and self-sensing devices based on human motion. *Joule*, *6*(7), 1501–1565. https://doi.org/10.1016/j.joule.2022.06.013

Lauriola, I., Lavelli, A., & Aiolli, F. (2022). An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. *Neurocomputing*, *470*, 443–456. https://doi.org/10.1016/j.neucom.2021.05.103

li, X., & Cao, X. (2023). Human motion recognition information processing system based on LSTM Recurrent Neural Network Algorithm. *Journal of Ambient Intelligence and Humanized Computing*, *14*(7), 8509–8521. https://doi.org/10.1007/s12652-021-03614-x

Lin, M., Zheng, Z., Yang, L., Luo, M., Fu, L., Lin, B., & Xu, C. (2022). A High-Performance, Sensitive, Wearable Multifunctional Sensor Based on Rubber/CNT for Human Motion and Skin Temperature Detection. *Advanced Materials*, *34*(1), 2107309. https://doi.org/10.1002/adma.202107309

Loh, S. B., Roy, D., & Fernando, B. (2022). *Long-Term Action Forecasting Using Multi-Headed Attention-Based Variational Recurrent Neural Networks*. 2419–2427. https://openaccess.thecvf.com/content/CVPR2022W/ABAW/html/Loh_Long-Term_Action_Forecasting_Using_Multi-Headed_Attention-Based_Variational_Recurrent_Neural_Networks_CVPRW_2022_paper.html

Lorek, P., Nowak, R., Trzcinski, T., & Zieba, M. (2022). FlowHMM: Flow-based continuous hidden Markov models. *Advances in Neural Information Processing Systems*, *35*, 8773–8784.

Lu, L., Zhang, C., Cao, K., Deng, T., & Yang, Q. (2022). A multichannel CNN-GRU model for human activity recognition. *IEEE Access*, *10*, 66797–66810.

Luo, J., Huang, D., Li, Y., & Yang, C. (2022). Trajectory Online Adaption Based on Human Motion Prediction for Teleoperation. *IEEE Transactions on Automation Science and Engineering*, *19*(4), 3184–3191. https://doi.org/10.1109/TASE.2021.3111678

Majid, F. F., Mazlan, S. S., & Mohamed, N. (2022). DEVELOPMENT OF SURVEILLANCE SYSTEM WITH AUTOMATED EMAIL AND TELEGRAM NOTIFICATION USING OPEN-SOURCE APPLICATION PROGRAMMING INTERPHASE (API). *International Journal of Infrastructure Research and Management Vol. 10 (2), December 2022*, 39.

Meng, Z., Zhang, M., Guo, C., Fan, Q., Zhang, H., Gao, N., & Zhang, Z. (2020). Recent Progress in Sensing and Computing Techniques for Human Activity Recognition and Motion Analysis. *Electronics*, *9*(9), Article 9. https://doi.org/10.3390/electronics9091357

Moon, H.-S., & Seo, J. (2022). Fast User Adaptation for Human Motion Prediction in Physical Human–Robot Interaction. *IEEE Robotics and Automation Letters*, *7*(1), 120–127. https://doi.org/10.1109/LRA.2021.3116319

Najeh, H., Lohr, C., & Leduc, B. (2022). Real-Time Human Activity Recognition in Smart Home on Embedded Equipment: New Challenges. *Participative Urban Health and Healthy Aging in the Age of AI: 19th International Conference, ICOST 2022, Paris, France, June 27–30, 2022, Proceedings*, 125–138.

Nasir, I. M., Raza, M., Shah, J. H., Wang, S.-H., Tariq, U., & Khan, M. A. (2022). HAREDNet: A deep learning based architecture for autonomous video surveillance by recognizing human actions. *Computers and Electrical Engineering*, *99*, 107805. https://doi.org/10.1016/j.compeleceng.2022.107805

Nyamasvisva, T. E., & Arabi, A. A. M. (n.d.). A COMPREHENSIVE SWOT ANALYSIS FOR ZERO TRUST NETWORK SECURITY MODEL. *International Journal of Infrastructure Research and Management Vol. 10 (1), June 2022*.

Paul, S. J., Elizabeth, I., & Gupta, B. K. (2021). Ultrasensitive Wearable Strain Sensors based on a VACNT/PDMS Thin Film for a Wide Range of Human Motion Monitoring. *ACS Applied Materials & Interfaces*, *13*(7), 8871–8879. https://doi.org/10.1021/acsami.1c00946

Pham, H. H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2022). Video-based human action recognition using deep learning: A review. *ArXiv Preprint ArXiv:2208.03775*.

Qiu, S., Zhao, H., Jiang, N., Wu, D., Song, G., Zhao, H., & Wang, Z. (2022). Sensor network oriented human motion capture via wearable intelligent system. *International Journal of Intelligent Systems*, *37*(2), 1646–1673. https://doi.org/10.1002/int.22689

Qu, L., Wang, Y., Yang, T., & Sun, Y. (2022). Human activity recognition based on WRGAN-GP-Synthesized micro-Doppler spectrograms. *IEEE Sensors Journal*, *22*(9), 8960–8973.

Ram, D. D., Muthukumaran, U., & Fatima, N. S. (2023). Enhanced Human Action Recognition with Ensembled DTW Loss Function in CNN LSTM Architecture. In S. Shakya, V. E. Balas, & W. Haoxiang (Eds.), *Proceedings of Third International Conference on Sustainable Expert Systems* (pp. 491–508). Springer Nature. https://doi.org/10.1007/978-981-19-7874-6_36

Rojas-Salazar, S. (2022). *Modeling State Duration and Emission Dependence in Hidden Markov and Hidden Semi-Markov Models* [PhD Thesis]. University of Missouri-Columbia.

Sahoo, S. P., Modalavalasa, S., & Ari, S. (2022). DISNet: A sequential learning framework to handle occlusion in human action recognition with video acquisition sensors. *Digital Signal Processing*, *131*, 103763.

Saleem, G., Bajwa, U. I., & Raza, R. H. (2023). Toward human activity recognition: A survey. *Neural Computing and Applications*, *35*(5), 4145–4182. https://doi.org/10.1007/s00521-022-07937-4

Sedmidubsky, J., Elias, P., Budikova, P., & Zezula, P. (2021). Content-based management of human motion data: Survey and challenges. *IEEE Access*, *9*, 64241–64255.

Şengönül, E., Samet, R., Abu Al-Haija, Q., Alqahtani, A., Alturki, B., & Alsulami, A. A. (2023). An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey. *Applied Sciences*, *13*(8), 4956.

Shahabi, F., Gao, Y., & Alshurafa, N. (2022). ActiveSense: A novel active learning framework for human activity recognition. *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops)*, 224–229.

Shahzad, A. R., & Jalal, A. (2021). A Smart Surveillance System for Pedestrian Tracking and Counting using Template Matching. *2021 International Conference on Robotics and Automation in Industry (ICRAI)*, 1–6. https://doi.org/10.1109/ICRAI54018.2021.9651452

Sharma, V., Gupta, M., Pandey, A. K., Mishra, D., & Kumar, A. (2022). A Review of Deep Learning-based Human Activity Recognition on Benchmark Video Datasets. *Applied Artificial Intelligence*, *36*(1), 2093705. https://doi.org/10.1080/08839514.2022.2093705

Shen, C., Yu, S., Wang, J., Huang, G. Q., & Wang, L. (2022). *A Comprehensive Survey on Deep Gait Recognition: Algorithms, Datasets and Challenges* (arXiv:2206.13732). arXiv. https://doi.org/10.48550/arXiv.2206.13732

Shen, X., & Ding, Y. (2022). Human skeleton representation for 3D action recognition based on complex network coding and LSTM. *Journal of Visual Communication and Image Representation*, *82*, 103386. https://doi.org/10.1016/j.jvcir.2021.103386

Singh, P. K., Kundu, S., Adhikary, T., Sarkar, R., & Bhattacharjee, D. (2021). Progress of human action recognition research in the last ten years: A comprehensive survey. *Archives of Computational Methods in Engineering*, 1–41.

Slade, S., Zhang, L., Yu, Y., & Lim, C. P. (2022). An evolving ensemble model of multi-stream convolutional neural networks for human action recognition in still images. *Neural Computing and Applications*, *34*(11), 9205–9231. https://doi.org/10.1007/s00521-022-06947-6

Solorzano, C., & Tsai, D.-M. (2023). Accurate Vision-Based PCB Positioning Using Cosine-Convolutional Neural Networks. *IEEE Transactions on Automation Science and Engineering*.

Stein, N., Bremer, G., & Lappe, M. (2022). Eye tracking-based lstm for locomotion prediction in vr. *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 493–503.

Sun, W., Min, X., Lu, W., & Zhai, G. (2022). A deep learning based no-reference quality assessment model for ugc videos. *Proceedings of the 30th ACM International Conference on Multimedia*, 856–865.

Tang, Y., Zhang, L., Min, F., & He, J. (2022). Multiscale deep feature learning for human activity recognition using wearable sensors. *IEEE Transactions on Industrial Electronics*, *70*(2), 2106–2116.

Trelinski, J., & Kwolek, B. (2021). CNN-based and DTW features for human activity recognition on depth maps. *Neural Computing and Applications*, *33*(21), 14551–14563. https://doi.org/10.1007/s00521-021-06097-1

Uchiyama, T., Sogi, N., Niinuma, K., & Fukui, K. (2023). Visually explaining 3D-CNN predictions for video classification with an adaptive occlusion sensitivity analysis. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1513–1522.

Wang, L., Zhou, Y., Li, R., & Ding, L. (2022). A fusion of a deep neural network and a hidden Markov model to recognize the multiclass abnormal behavior of elderly people. *Knowledge-Based Systems*, *252*, 109351. https://doi.org/10.1016/j.knosys.2022.109351

Wang, X., He, M., Yang, L., Wang, H., & Zhong, Y. (2023). Human Activity Recognition Based on an Efficient Neural Architecture Search Framework Using Evolutionary Multi-Objective Surrogate-Assisted Algorithms. *Electronics*, *12*(1), Article 1. https://doi.org/10.3390/electronics12010050

Yalçinkaya, B., Couceiro, M. S., Soares, S. P., & Valente, A. (2023). Human-Aware Collaborative Robots in the Wild: Coping with Uncertainty in Activity Recognition. *Sensors*, *23*(7), Article 7. https://doi.org/10.3390/s23073388

Yang, L.-H., Liu, J., Wang, Y.-M., Nugent, C., & Martínez, L. (2021). Online updating extended belief rule-based system for sensor-based activity recognition. *Expert Systems with Applications*, *186*, 115737. https://doi.org/10.1016/j.eswa.2021.115737

Zhang, C., Liang, J., Li, X., Xia, Y., Di, L., Hou, Z., & Huan, Z. (2022). Human action recognition based on enhanced data guidance and key node spatial temporal graph convolution. *Multimedia Tools and Applications*, *81*(6), 8349–8366. https://doi.org/10.1007/s11042-022-11947-8

Zhong, C., Hu, L., & Xia, S. (2022). Spatial–temporal modeling for prediction of stylized human motion. *Neurocomputing*, *511*, 34–42.