IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3026~3035
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3026-3035

Long short-term memory based activity detection using skeleton joints data: a systematic review

Bakkala Santha Kumar, R. Shankar
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India

Article Info

Article history:

Received Nov 13, 2024
Revised Jun 9, 2026
Accepted Jul 9, 2026

Keywords:

Abnormal activity recognition
Human activity recognition
Long-short term memory
Real-time activity recognition
Skeleton joints data analysis

ABSTRACT

In today's security and surveillance applications, recognizing abnormal activity is critical component of identifying possible hazardous or unusual human behaviors. There is need for new technologies that can detect abnormal human behaviors precisely. The present review investigates various aspects of detection process, focusing on skeleton joints input data. It then explores adoption of deep learning (DL) architectures, such as long short-term memory (LSTMs) and transformers, to improve accuracy and robustness of recognition models. The review explores synergistic integration of LSTMs and transformers to improve recognition of unusual activity. By integrating LSTMs' processing capabilities and attention mechanisms of transformers, enhanced models can accurately identify intricate patterns of activity. Despite the advancements that have been made in the field, the challenges that remain are still related to recognition of unusual activities. These include lack of scalability for large datasets, need for models that can recognize complex behaviors across diverse applications, and need to ensure that detection is performed in a low-latency manner. The paper explores future directions of developing LSTM-based models that can recognize unusual activity using skeleton joint data in a cloud-based environment. The review emphasizes the potential of such solutions that can take advantage of the processing power of graphics processing units (GPUs) and tensor processing units (TPUs) and enable real-time recognition of activity in large datasets.

Corresponding Author:

Bakkala Santha Kumar
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation
Vaddeswaram, Andhra Pradesh, India
Email: santhakumar.bakkala@gmail.com

1. INTRODUCTION

In recent times, the domain of human activity recognition (HAR) has experienced remarkable progress, especially through the incorporation of deep learning (DL) techniques. Among these techniques, long short-term memory (LSTM) networks [1]–[3] have surfaced as an effective instrument for capturing temporal relationships in sequential information. This systematic review focuses on the application of LSTM-based activity detection techniques that leverage skeleton joint data, which provides a compact and informative representation of human posture and movement. The use of skeleton joints, derived from motion capture systems or depth cameras, allows for robust activity recognition, even in challenging environments. Nevertheless, even with the encouraging findings highlighted in the literature, numerous obstacles continue to exist, such as discrepancies in joint configurations, obstructions, and the intricacy of human activities. Additionally, computational factors like model efficiency, real-time processing abilities, and scalability are crucial for practical implementations in ever-changing environments. This review intends to consolidate current research on LSTM-based activity detection utilizing skeleton joints, underscore the fundamental challenges encountered by researchers, and examine computational approaches that could improve performance and applicability in real-world contexts. By tackling these elements, the review aids in fostering a more comprehensive understanding of cutting-edge developments and lays the groundwork for future research trajectories in this advancing field.

Recent progress in skeleton-based HAR has been based on either recurrent networks, e.g. LSTM networks, or attention-based transformer models [4]–[7] to capture temporal dependence in joint-level motion data. Although LSTM networks are best applied in learning short and medium-term time dynamics in skeleton sequences, they tend to fail to capture long-range temporal dynamics and multi-joint interactions in activities over long time scales. Transformer architectures, on the side, are effective in being able to compute global temporal modelling via self-attention. But transformers consume large-scale data and high computational resources, which restrict their application in real-time or resource-constrained settings. The novelty of this review is in the systematic study of hybrid LSTM-transformer models. The synergistic combination provides a balanced modelling paradigm, which overcomes the singular architecture limitations when used individually. Moreover, the intersection of LSTM-transformer models with cloud-based computing areas opens a novel dimension of research that has not been adequately examined in earlier reviews. Cloud infrastructures [8] can be scaled to support training, distributed inference [9], and real-time deployment of computationally intensive hybrid models, which are especially useful to large-scale skeleton-based HAR systems, including smart surveillance, healthcare monitoring, and human-robot interaction (HRI). The review also indicates that cloud-enabled LSTM-transformer pipelines are novel, where industrial sophistications at edge parts consist of lightweight skeleton extraction, and deep temporal models and attention-infused inference are run on cloud servers. This study stands out as one of the first review papers to fill the sequence modelling innovation (LSTM-transformer fusion) and practical cloud-based HAR deployment, creating gaps in the research and identifying opportunities towards scalable and smart activity recognition systems.

To continue enhancing the structural purity of HAR research of skeletal type form, this review suggests a graphical taxonomy, which previsualizes a methodically categorized set of model infrastructure and deployment strategies utilized by LSTM- and transformer-based frameworks. In contrast to other previous surveys, which boil relevant studies down to comparisons on a tabular case or narrative analysis, the proposed taxonomy will provide a multi-dimensional graphical structure which ranks existing studies on 3 orthogonal scales. These scales include: i) model architecture, which is housed in standalone LSTM models, pure transformer-based models and mixed LSTM-transformer pipelines; ii) temporal and spatial modeling strategies, including joint-level encoding, spatio-temporal graph representations and attention-based inter-joint dependency learning; and iii) deployment schemes, which include edge-based in such visual abstraction allows the readers to easily recognize architectural trend, prevailing modeling preferences, and uncharted options in the literature.

The taxonomy also factors in deployment-conscious design factors, pointing to the effects of task-oriented constraints on computational costs, latency tolerance, and scalability concerns in the operational tip of the iceberg on the architecture. The mapping of hybrid LSTM-transformer models to cloud-centered and edge-cloud collaborative models allows exploration of an emerging direction in research with the involvement of satisfying timeliness modeling, intimacy, and deployment efficiency. It is especially the contribution that is very new to a systematic review, as it brings the algorithmic innovation and the system-level deployment strategies into contact, providing researchers and practitioners with a structured viewpoint on the current work and the future opportunities of research.

The suggested visual taxonomy as shown in Figure 1 can be used as a summarization tool, as well as a design template for next-generation, scalable skeleton-based HAR systems. The study will explore innovative approaches that leverage DL techniques to improve accuracy and robustness in activity recognition systems. Furthermore, the integration of multimodal data sources [10], [11] will be examined, which can provide richer contextual information and potentially lead to more nuanced interpretations of human activities.

Furthermore, transfer learning will be examined for its contribution to the adjustment of models for various environments, allowing systems to achieve better generalization across a range of contexts and populations [12]. The ethical considerations involved in implementing these technologies in practical applications will also be considered, ensuring that privacy and security issues are thoroughly addressed while optimizing the advantages of improved activity recognition. Moreover, the significance of continuous learning and adaptation within these models will be highlighted, stressing the necessity for regular assessment and enhancement to keep up with changing user demands and technological progress.

2. REVIEW METHOD

The entire review process is carried out by framing the research questions that can contribute to the understanding of abnormal activity recognition using skeleton joints, how to enhance the detection accuracy, the role of LSTM-based models in analyzing the skeleton joints, challenges to overcome, and future directions that can guide in analyzing the analysis of skeleton joints data and perform the recognition with high efficiency. The major contribution of the review is to address the following framed research questions.

- RQ1: how do skeleton joints contribute to the detection of normal and abnormal human behavior?
- RQ2: how do the transformation models contribute to activity detection?
- RQ3: how are LSTM-based models efficient in analyzing the skeleton joints data and conducting the detection well?
- RQ4: what are the different challenges in detecting different activities using the current recognition frameworks and future scope?

Google Search engine is used to search for relevant articles published in various reputed journals. The search engine performed the search in many reputed databases like Springer, IEEE Xplore, and Science Direct. The search process is conducted using keywords combined with the AND Boolean operator. All the databases were explored to identify the articles to be considered for review systematically and provide an opportunity to improve the performance of the models and predict the activities most efficiently. Four keywords were selected based on the research questions. These search words were combined during the advanced search process to identify more relevant articles. Table 1 shows the list of keywords used during the review process.

Table 1. Keywords used during the search

Database	Initial search words	Advanced search words
IEEE Xplore	LSTM models	LSTM models [AND] transformers
Springer	Transformer models	LSTM models [AND] human activity recognition
Science Direct	Human activity recognition	HAR [AND] skeleton joints data
Others	Skeleton joints data	LSTM models [AND] transformers [AND] HAR [AND] skeleton joints data analysis

A systematic screening procedure was employed in accordance with established systematic literature review and PRISMA reporting guidelines [13], [14]. Studies published between 2019 and 2024 were considered, focusing on HAR, using skeleton joint data, and mentioning LSTM, transformer, or related DL architectures. A few papers were excluded from the review that did not apply the proposed techniques to skeleton-based activity detection. The remaining articles were screened through identification, duplicate removal, relevance assessment, eligibility checking, and full-text evaluation before final inclusion in the review. Figure 1 summarizes all the stages using a systematic article-selection flow.

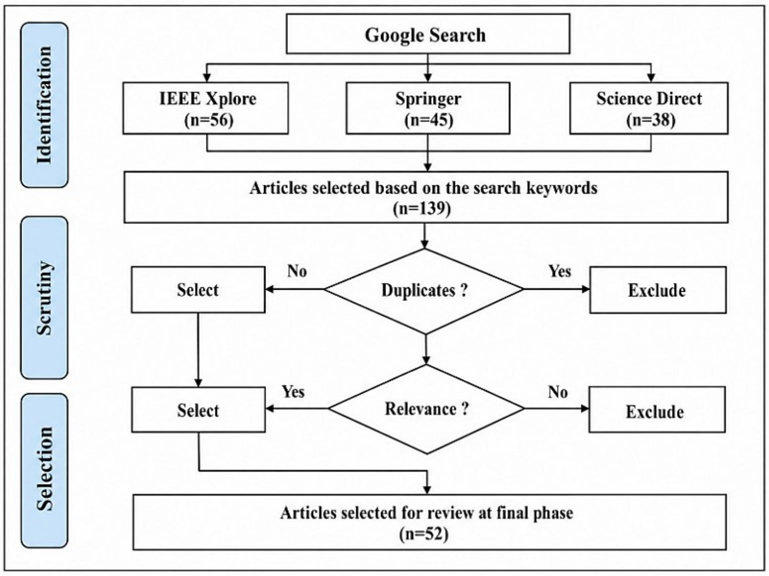

Figure 1. Systematic article selection process

The initial search generated 126 articles, but only 85 were considered based on the relevance and the year of publication. The further search process is conducted by fine-tuning the search keywords using the common keyword 'human activities recognition'. Finally, 52 articles were considered for the final review process. Figure 1 shows the process flow diagram summarizing all the steps.

3. DISCUSSIONS

The section presents the findings of the reviewed articles in relation to the four research questions formulated for this study. The discussion is organized according to each research question to ensure a systematic interpretation of the literature. The subsections highlight the major approaches, findings, challenges, and research opportunities identified across the selected studies.

3.1. RQ1: understanding abnormal activity recognition: the role of skeleton joints

For this research question, 35 papers were considered by filtering them with the role of skeleton joints in understanding and recognizing abnormal behavior. At last, 9 papers were considered for analysis. This review aims to explore various methodologies employed in detecting and interpreting abnormal activities through the analysis of skeletal joint movements, highlighting advancements in machine learning techniques and their applications in real-time monitoring systems. Furthermore, the integration of depth sensors and computer vision has significantly enhanced the accuracy of these systems, enabling more nuanced interpretations of human behavior in diverse environments.

Additionally, the use of convolutional neural networks (CNNs) has proven effective in extracting features from skeletal data, allowing for improved classification of activities and the identification of subtle deviations that may indicate abnormal behavior [15], [16]. In addition to the foundational role of skeleton joints in recognizing abnormal activities, recent advancements have highlighted the importance of integrating multimodal data for enhanced performance [17]. By employing CNNs alongside attention-based LSTM models, researchers can effectively merge deep spatial information with temporal dynamics, thereby improving the extraction of key behaviors from complex datasets [18].

In addition to the ethical considerations surrounding abnormal activity recognition systems, it is imperative to address the technical challenges that persist in achieving optimal performance. Recent studies have shown that employing recurrent neural networks (RNNs) can significantly enhance the analysis of temporal dependencies within skeletal data, thereby improving accuracy in recognizing both normal and anomalous activities [19], [20]. As the landscape of abnormal activity recognition evolves, the potential for integrating advanced machine learning techniques with real-time data streams presents a compelling frontier. For instance, the application of attention mechanisms within LSTM frameworks can enhance the model's sensitivity to critical moments in human behavior, allowing for more accurate detection of anomalies as they occur.

Moreover, exploring the use of generative adversarial networks (GANs) could provide innovative ways to augment training datasets, addressing issues of bias and improving generalizability across diverse populations [21]. As the integration of advanced machine learning techniques continues to reshape abnormal activity recognition, it becomes increasingly important to examine the role of data quality and feature selection in model performance. The effectiveness of LSTM models can be significantly influenced by the choice of input features; thus, employing methods such as dimensionality reduction or geometrical feature extraction from skeletal data is essential for optimizing accuracy [21].

Moreover, the integration of advanced techniques such as hybrid models that combine LSTM networks with CNNs has shown promise in enhancing activity recognition systems by capturing both spatial and temporal features effectively. For instance, recent studies have demonstrated that employing a visual geometry group (VGG)-LSTM model can significantly improve accuracy across various datasets, achieving impressive results like 95.85% on the wireless sensor data mining (WISDM) dataset [2]. Furthermore, exploring multimodal data sources—such as integrating audio or environmental sensors—can enrich the context surrounding human activities, leading to a more comprehensive understanding and interpretation of complex behaviors.

3.2. RQ2: transformers in motion: enhancing abnormal activity detection

For this research question, 32 papers were considered by filtering them based on the role of Transformers in enhancing the performance of activity detection. Finally, 7 papers were considered for analysis. Recent studies indicate that integrating Transformer models with skeleton joint data can yield even higher accuracy rates than traditional methods, further pushing the boundaries of what is achievable in this domain [22]. In addition to the integration of edge computing and human expertise, the role of advanced anomaly detection techniques in addressing cybersecurity threats cannot be overstated. As digital infrastructures become increasingly complex, traditional methods often fall short in identifying sophisticated cyber-attacks that exploit system vulnerabilities; hence, DL approaches have emerged as a vital solution [23]. For instance, utilizing models such as CNNs combined with LSTM networks has shown remarkable success in detecting network anomalies by analyzing vast amounts of sequential data, thus enhancing both accuracy and efficiency in real-time threat identification [24], [25].

Furthermore, these methodologies not only bolster defenses against immediate attacks but also contribute to a more profound understanding of attack patterns over time, allowing organizations to pre-emptively strengthen their security postures and mitigate risks effectively. As organizations increasingly rely on sophisticated anomaly detection techniques, the integration of unsupervised learning methods has emerged as a vital complement to traditional approaches. These methods can autonomously identify patterns without extensive labeled datasets, making them particularly useful in dynamic environments where new types of anomalies frequently arise.

For instance, leveraging convolutional long short-term memory (Conv-LSTM) models enables systems to learn spatiotemporal features from video data effectively, thereby enhancing surveillance capabilities and improving response times in critical situations [26]. By embracing such innovative methodologies, organizations not only fortify their defenses against evolving threats but also pave the way for more adaptive and resilient operational frameworks that can respond intelligently to unforeseen challenges. In addition to the advancements in hybrid models, the integration of unsupervised DL techniques has emerged as a transformative force in anomaly detection [27], [28]. These methods are particularly adept at uncovering subtle deviations from normal behavior patterns without relying on extensive labeled datasets, thereby addressing one of the significant challenges faced by traditional approaches.

3.3. RQ3: LSTM-based models: a supportive framework for skeleton joint analysis

For this research question, 27 papers were considered by filtering them with the role of LSTM models in analyzing skeleton joint data during the process of recognizing abnormal behavior. At last, 7 papers were considered for analysis. As the field progresses, integrating these advanced methodologies will be crucial for developing comprehensive solutions that address both technical and ethical considerations inherent in deploying HAR technologies effectively. By capturing the dynamics of motion over time, LSTM-based approaches can effectively distinguish between normal and abnormal activities, leading to improved detection rates in various applications such as surveillance, healthcare, and human-computer interaction. Furthermore, the incorporation of attention mechanisms within these LSTM frameworks can further enhance performance by allowing the model to focus on critical joints and movements that are most indicative of abnormal behavior. Additionally, the integration of multi-modal data sources, such as audio and environmental context, can provide a richer understanding of the scenarios being analyzed, enabling even more robust predictions and insights into human behavior. By leveraging these diverse data streams, researchers can develop more comprehensive models that account for the complexities of human interactions and environmental influences, ultimately leading to improved outcomes in various fields, including security and healthcare management [29].

The primary challenges in activity recognition using skeleton joints data include the difficulty in capturing discriminative spatiotemporal features, computational inefficiency, and the handling of occlusions and noise. Traditional methods often struggle with recognizing actions that have subtle differences in motion patterns due to their reliance on naive attention mechanisms, which are not adept at distinguishing such nuances [30]. Additionally, the complexity of transformer models can lead to high computational costs and resource demands, which are significant barriers to their practical application [31]. Another challenge is the limited receptive field of graph convolutional networks (GCNs), which restricts their ability to capture long-range dependencies between joints across frames [4].

Moreover, the aggregation of joints into body parts can obscure fine-grained actions, making it difficult to fully exploit part-level data. The IIP-transformer leverages intra- and inter-part interactions to improve recognition of both coarse- and fine-grained actions [5]. Additionally, the language-guided graph transformer (LGGT) incorporates semantic guidance to enhance the model's understanding of action relationships [6]. These innovations demonstrate the potential of transformers to overcome the inherent challenges in skeleton-based action recognition, offering more robust, efficient, and accurate models for real-world applications.

In skeleton-based HAR systems, interpretability is also crucial. Attention weights, key joint movements, motion trajectories, and patterns at the frame level focus on how LSTM and transformer models classify an activity. These explainable features can be helpful in sensitive applications as they provide information on which body joints or time steps most influenced the final classification.

3.4. RQ4: focusing on skeleton joints and transformers for superior activity recognition

For this research question, 23 papers were considered by filtering them with the concept of integrating skeleton joints data and Transformers for better activity recognition. Finally, 8 papers were considered for analysis. As the integration of skeleton joint data with advanced architectures like transformers and LSTMs gains traction, attention must also be directed toward enhancing user interaction with these systems. For example, developing intuitive interfaces that allow users to provide real-time feedback can significantly improve model training and adaptability, fostering a more personalized experience. The multi-scale temporal transformer (MTT) also addresses the need for capturing long-term temporal dependencies, which are crucial for sequential skeleton data, by integrating multi-scale temporal embeddings and transformer encoders [24]. The cloud environment offers a scalable and efficient platform for deploying these advanced models, allowing for real-time processing and analysis of large datasets. The integration of LSTM and transformers in the cloud can leverage the strengths of both models, with LSTM handling temporal sequences and transformers capturing complex dependencies, thus enhancing the overall accuracy and efficiency of HAR systems.

Moreover, the use of DL techniques, such as those combining CNN and LSTM, has shown promise in extracting spatiotemporal features from skeleton data, further improving recognition accuracy [32]. The development of intelligent HAR systems using skeleton information, as discussed [33], underscores the potential of these technologies in various applications, from monitoring the elderly to surveillance. The cloud environment not only facilitates the deployment of these models but also supports the integration of multiple data sources, such as image feature extraction and object detection, to improve action recognition [10]. The spatio-temporal image formation technique proposed by Tasnim *et al.* [11] also highlights the potential of using pre-trained models and fusion techniques to enhance the performance of HAR systems in the cloud. Finally, addressing challenges such as missing joint recovery, as explored by Zhalgasbayev and Tu [34], is crucial for improving the robustness of skeleton-based activity recognition systems in dynamic and uncertain environments [35], [36]. Overall, the combination of LSTM, transformers, and cloud computing presents a powerful approach to advancing the field of HAR using skeleton joints.

Latency, privacy, bandwidth, and computational cost are critical challenges for cloud-enabled HAR systems for real-world deployment. A feasible solution is to do the skeleton extraction and basic processing on edge devices, and the large-scale storage and complex LSTM-transformer inference on the cloud server. The integration of edge and cloud can help to cut latency and increase scalability. But, in certain applications like healthcare, surveillance, and elderly monitoring, privacy-preserving techniques are required for using HAR system. The results of this review are applicable to a few practical applications of HAR such as fall detection for the elderly, smart surveillance, rehabilitation, sports analysis, HRI, and healthcare monitoring. Transfer learning, domain adaptation, lightweight edge processing, and scalable inference to the cloud are all possible techniques to adapt the models to various environments. The models can be transferred to different environments with transfer learning, adapted to various environments with domain adaptation, processed at the edge with lightweight models, and scaled to the cloud for inference.

4. COMPARATIVE ANALYSIS OF EXISTING WORKS

The field of HAR has significantly advanced with the emergence of DL and machine learning models [37], [38], particularly in environments that utilize cloud computing. This technology enables the prediction and classification of human activities based on the collected sensor data [39]. Due to the diverse computing environments and methods, the development of various processing techniques and models has been carried out. Some of these include the use of transformer networks and long-term memory. This subsection aims to review the various implementations of HAR using transformer networks, LSTM models, and others in cloud environments. There was no formal meta-analysis conducted due to the differences in the data sets, data format, evaluation metrics, and experimental conditions used across the studies. Table 2 summarizes the existing HAR studies based on method, dataset, performance, and key influencing factors. The comparison allows the evaluation of the relative performance of the LSTM, CNN-LSTM, gated recurrent unit (GRU), transformer, GCN, and hybrid models on various HAR datasets.

In addition to LSTM models and transformer models, other methods, such as GCNs and multimodal fusion in skeleton-based HAR, are also significant. The advantage of GCN-based models is that they can learn spatial relationships between body joints, and LSTM models and transformer models are primarily used to model temporal and long-range movements. By 4 combining skeleton data and RGB video, wearable sensors, audio, or environmental information, the performance can further be improved by multimodal fusion.

Table 2. Comparative analysis with the state-of-the-art works

Ref.	Problem statement	Methods proposed	Dataset used	Results obtained	Influencing factors
[8]	HAR using edge-computing	OI-EleAttG-GRU	UCI machine learning repository, UCI HAR, and self-created	85.4% average accuracy	Feature extraction
[9]	HAR for pedestrian navigation	DL algorithms with sensor data, LSTM, and CNN	UCI	Up to 91.92% accuracy and 95.55% on smartphone posture recognition	Seven traditional ML algorithms
[12]	HAR using mobile health data	CNN with LSTM and SVM transfer learning	MHEALTH	99.8% impressive accuracy	Transfer learning
[40]	HAR	A GRU algorithm	WISDM	97.08% testing accuracy and 0.221 testing loss	Hyper-parameter tuning and k-fold cross-validation
[41]	Performance evaluation w.r.t HAR	Feature extraction using 3D-DCNN using LSTM.	UCF50 action recognition	86.57% accuracy	3DCNN-LSTM approach
[42]	HAR using DL	A residual BiLSTM network	UCF-Crime Kaggle	100% classification accuracy	Combining convolutional and recurrent layers
[43]	Sensor-based HAR	A modified BRNN with LSTM	Mobifall	4% improvement over previous methods	Feature extraction
[44]	HAR using hybrid DL	Hybrid DL combined with CNN and LSTM	20 participants dataset with 12 different physical activities	90.89% accuracy	Combination of CNN and LSTM
[45]	Deep neural network (DNN)-based HAR	LSTM combined with RNN	WISDM	94% accuracy with 30% loss	L2 regularization and dropout.
[46]	HAR using sensor data	BiLSTM, CNN, and ConvLSTM	WISDM	84.8% with BiLSTM and 84.3% with ConvLSTM.	Watch accelerometer data
[47]	HAR using DNN	A hybrid model combining CNN and LSTM algorithms	UCI-HAR	96.8% accuracy.	Principal component analysis (PCA) and feature extraction
[48]	HAR using LSTM and RNN	LSTM-RNN DL	OPPORTUNITY	94% overall accuracy. 91% accuracy on non-repetitive motions and 80% accuracy on repetitive ones	Comparative analysis and performance metrics
[49]	HAR using wearable sensor data	CNN with GRU	WISDM smartphone and smartwatch activity and biometrics	The CNN-GRU model achieved the highest classification accuracy	Leveraging spatial and temporal features
[50]	HAR using self-attention model	CNN with LSTM	H-activity, MHEALTH, and UCI-HAR	99.93% accuracy.	A self-attention layer to extract temporal features
[51]	HAR based on CNN	2D-CNN	WISDM	achieving an accuracy of 97.20% on the WISDM dataset	2D-CNN with hyperparameter tuning
[52]	HAR using transitional activities	LSTM with BiLSTM and CNN	Dataset A (HAPT) and dataset B	96.1% accuracy for transitional activities and 98.4% for basic activities	Hybrid model combined with decision fusion
[53]	HAR using noisy data	CNN with BiLSTM parallel model using attention mechanism	UCI HAR dataset and the WISDM HAR	96.71% accuracy	Dimensionality reduction and hyperparameter tuning
[54]	HAR using DL	CNN with LSTM	UCI-HAR, OPPORTUNITY, UNIMIB-SHAR, PAMAP2, and WISDM	98.82% impressive accuracy on specific datasets	Hybrid architecture combined with LSTM and ML algorithms

5. CONCLUSION

The comprehensive review of LSTM-based activity detection utilizing skeleton joints underscores the transformative capabilities of DL techniques within the sphere of HAR. The combination of LSTM networks with emerging transformer architectures signifies a notable advance in effectively capturing the temporal and spatial intricacies of human movements, thus improving the precision and reliability of activity recognition systems. Nonetheless, despite the encouraging progress, the domain still faces considerable obstacles, including implementing technologies in real-world scenarios. Tackling these issues through innovative strategies like ensemble learning, multimodal data integration, and federated learning will be essential for creating resilient systems that emphasize user privacy while enhancing their operational efficiency. As research advances, the knowledge acquired from this review will lay the groundwork for future investigations aimed at refining activity recognition systems, paving the way for a deeper understanding and interpretation of human behaviors in intricate contexts. The importance of detecting unusual activities in modern surveillance has been highlighted in this review. It emphasizes the importance of having skeleton joint data as a starting point for identifying such activities. The integration of the Transformer and LSTM models can provide a more accurate and robust solution to identifying complex activity patterns. The development of a cloud-based LSTM framework that is capable of handling joint skeleton analysis is expected to help improve the detection of unusual activities.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Bakkala Santha Kumar	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
R. Shankar		✓				✓		✓	✓	✓	✓	✓		

C	: **C**onceptualization	I	: **I**nvestigation	Vi	: **Vi**sualization
M	: **M**ethodology	R	: **R**esources	Su	: **Su**pervision
So	: **So**ftware	D	: **D**ata Curation	P	: **P**roject administration
Va	: **Va**lidation	O	: Writing - **O**riginal Draft	Fu	: **Fu**nding acquisition
Fo	: **Fo**rmal analysis	E	: Writing - Review & **E**diting		

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interests to declare that are relevant to the content of this article.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] M. Ghislieri, G. L. Cerone, M. Knaflitz, and V. Agostini, "Long short-term memory (LSTM) recurrent neural network for muscle activity detection," *Journal of NeuroEngineering and Rehabilitation*, vol. 18, no. 1, 2021, doi: 10.1186/s12984-021-00945-w.

[2] K. Xia, J. Huang, and H. Wang, "LSTM-CNN architecture for human activity recognition," *IEEE Access*, vol. 8, pp. 56855–56866, 2020, doi: 10.1109/ACCESS.2020.2982225.

[3] C. Si, W. Chen, W. Wang, L. Wang, and T. Tan, "An attention enhanced graph convolutional LSTM network for skeleton-based action recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1227–1236, doi: 10.1109/CVPR.2019.00132.

[4] J. Do and M. Kim, "SkateFormer: skeletal-temporal transformer for human action recognition," in *Computer Vision – (ECCV 2024)*. Cham, Switzerland: Springer, 2024, doi: 10.1007/978-3-031-72940-9.

[5] Q. Wang, S. Shi, J. He, J. Peng, T. Liu, and R. Weng, "IIP-transformer: intra-inter-part transformer for skeleton-based action recognition," in *2023 IEEE International Conference on Big Data, BigData 2023*, 2023, pp. 936–945, doi: 10.1109/BigData59044.2023.10386970.

[6] L. Weng, W. Lou, and F. Gao, "Language guided graph transformer for skeleton action recognition," in *Communications in Computer and Information Science*, Singapore: Springer, 2024, pp. 283–299, doi: 10.1007/978-981-99-8141-0_22.

[7] J. Kong, Y. Bian, and M. Jiang, "MTT: multi-scale temporal transformer for skeleton-based action recognition," *IEEE Signal Processing Letters*, vol. 29, pp. 528–532, 2022, doi: 10.1109/LSP.2022.3142675.

[8] X. Huang, Y. Yuan, C. Chang, Y. Gao, C. Zheng, and L. Yan, "Human activity recognition method based on edge computing-assisted and GRU deep learning network," *Applied Sciences,* vol. 13, no. 16, 2023, doi: 10.3390/app13169059.

[9] J. Ye, X. Li, X. Zhang, Q. Zhang, and W. Chen, "Deep learning-based human activity real-time recognition for pedestrian navigation," *Sensors*, vol. 20, no. 9, 2020, doi: 10.3390/s20092574.

[10] J. D. Domingo, J. G.-G.-Bermejo, and E. Zalama, "Improving human activity recognition integrating LSTM with different data sources: features, object detection and skeleton tracking," *IEEE Access*, vol. 10, pp. 68213–68230, 2022, doi: 10.1109/ACCESS.2022.3186465.

[11] N. Tasnim, M. K. Islam, and J. H. Baek, "Deep learning-based human activity recognition using spatio-temporal image formation of skeleton joints," *Applied Sciences*, vol. 11, no. 6, 2021, doi: 10.3390/app11062675.

[12] R. A. Praba and L. Suganthi, "HARNet: automatic recognition of human activity from mobile health data using CNN and transfer learning of LSTM with SVM," *Automatika*, vol. 65, no. 1, pp. 167–178, 2024, doi: 10.1080/00051144.2023.2290736.

[13] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," *EBSE Technical Report*, EBSE-2007-01, Keele University and University of Durham, United Kingdom, 2007.

[14] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, 2021, doi: 10.1136/bmj.n71.

[15] A. Elkholy, M. E. Hussein, W. Gomaa, D. Damen, and E. Saba, "Efficient and robust skeleton-based quality assessment and abnormality detection in human action performance," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 280–291, Aug. 2020, doi: 10.1109/JBHI.2019.2904321.

[16] C. N. Phyo, T. T. Zin, and P. Tin, "Deep learning for recognizing human activities using motions of skeletal joints," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 2, pp. 243–252, May 2019, doi: 10.1109/TCE.2019.2908986.

[17] B. X. B. Yu, Y. Liu, and K. C. C. Chan, "Skeleton-based detection of abnormalities in human actions using graph convolutional networks," in *Proceedings - 2020 2nd International Conference on Transdisciplinary AI, TransAI 2020*, 2020, pp. 131–137, doi: 10.1109/TransAI49837.2020.00030.

[18] T. Ahmad, S. T. H. Rizvi, and N. Kanwal, "Transforming spatio-temporal self-attention using action embedding for skeleton-based action recognition," *Journal of Visual Communication and Image Representation*, vol. 95, 2023, doi: 10.1016/j.jvcir.2023.103892.

[19] D. O. Esan, P. A. Owolawi, and C. Tu, "Detection of anomalous behavioural patterns in university environment using CNN-LSTM," in *2020 23rd International Conference on Information Fusion, FUSION 2020*, 2020, pp. 1–6, doi: 10.23919/FUSION45008.2020.9190406.

[20] K. Zhou, T. Wu, C. Wang, J. Wang, and C. Li, "Skeleton based abnormal behavior recognition using spatio-temporal convolution and attention-based LSTM," *Procedia Computer Science*, vol. 174, pp. 424–432, 2020, doi: 10.1016/j.procs.2020.06.110.

[21] R. Nale, M. Sawarbandhe, N. Chegogoju, and V. Satpute, "Suspicious human activity detection using pose estimation and LSTM," in *2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation, IRIA 2021*, 2021, pp. 197–202, doi: 10.1109/IRIA53009.2021.9588719.

[22] G. Pandey, A. K. Karn, and M. Jha, "Human Activity recognition using CNN-LSTM-GRU model," *International Research Journal on Advanced Engineering Hub*, vol. 2, no. 04, pp. 889–894, 2024, doi: 10.47392/irjaeh.2024.0125.

[23] P. Kallepally and M. Rajesh, "LSTM-based deep learning architecture for recognition of human activities," in *Advances in Intelligent Systems*, Singapore: Springer, 2023, pp. 163–172, doi: 10.1007/978-981-19-5443-6_14.

[24] Y. Li, T. Guo, X. Liu, and R. Xia, "Skeleton-based action recognition with Lie group and deep neural networks," in *2019 IEEE 4th International Conference on Signal and Image Processing, ICSIP 2019*, 2019, pp. 26–30, doi: 10.1109/SIPROCESS.2019.8868548.

[25] H. Yun *et al.*, "Development of action-recognition technology using LSTM based on skeleton data," in *The 9th IIAE International Conference on Industrial Application Engineering 2021 (ICIAE2021)*, 2021, pp. 35–39, doi: 10.12792/iciae2021.009.

[26] G. Wu, Z. Guo, L. Li, and C. Wang, "Video abnormal event detection based on CNN and LSTM," in *2020 IEEE 5th International Conference on Signal and Image Processing, ICSIP 2020*, 2020, pp. 334–338, doi: 10.1109/ICSIP49896.2020.9339428.

[27] G. Paoletti, C. Beyan, and A. D. Bue, "SKELTER: unsupervised skeleton action denoising and recognition using transformers," *Frontiers in Computer Science*, vol. 5, 2023, doi: 10.3389/fcomp.2023.1203901.

[28] T. Acharya, A. Annamalai, and M. F. Chouikha, "Efficacy of CNN-bidirectional LSTM hybrid model for network-based anomaly detection," in *13th IEEE Symposium on Computer Applications and Industrial Electronics (ISCAIE 2023)*, 2023, pp. 348–353, doi: 10.1109/ISCAIE57739.2023.10165088.

[29] J. Liu, J. Bai, H. Li, and B. Sun, "Improved LSTM-based abnormal stream data detection and correction system for internet of things," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1282–1290, Jan. 2022, doi: 10.1109/TII.2021.3079504.

[30] L. Wen *et al.*, "Frequency guidance matters: skeletal action recognition by frequency-aware mixed transformer," in *Proceedings of the 32nd ACM International Conference on Multimed*, Melbourne, Australia, 2024.

[31] S. Oraki, H. Zhuang, and J. Liang, "LORTSAR: low-rank transformer for skeleton-based action recognition," in *Advances in Visual Computing (ISVC 2024)*, Cham, Switzerland: Springer, 2024, doi: 10.1007/978-3-031-77392-1.

[32] R. K. Murthy, S. Dhanraj, T. N. Manjunath, A. N. Prasad, P. K. Pareek, and H. N. Kumar, "Human activity recognition using CNN and long-term short-term memory," *International Journal of Health Sciences*, vol. 6, no. S6, pp. 10797–10809, 2022, doi: 10.53730/ijhs.v6ns6.12919.

[33] S. A. Roseline, S. Karthik, and I. N. V. D. Sruti, "Intelligent human anomaly detection using LSTM autoencoders," in *3rd International Conference on Advances in Computing, Communication and Applied Informatics, ACCAI 2024*, 2024, pp. 1–6, doi: 10.1109/ACCAI61061.2024.10602454.

[34] Y. Zhalgasbayev and N. A. Tu, "Two-Branch stacked transformer for 2D skeleton-based action recognition," in *2023 17th International Conference on Ubiquitous Information Management and Communication, IMCOM 2023*, 2023, pp. 1–6, doi: 10.1109/IMCOM56909.2023.10035660.

[35] J. Zhang, Y. Jia, W. Xie, and Z. Tu, "Zoom transformer for skeleton-based group activity recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 12, pp. 8646–8659, Feb. 2022, doi: 10.1109/TCSVT.2022.3193574.A.

[36] A. Zhu *et al.*, "Exploring a rich spatial–temporal dependent relational model for skeleton-based action recognition by bidirectional LSTM-CNN," *Neurocomputing*, vol. 414, pp. 90–100, 2020, doi: 10.1016/j.neucom.2020.07.068.

[37] B. Lindemann, B. Maschler, N. Sahlab, and M. Weyrich, "A survey on anomaly detection for technical systems using LSTM networks," *Computers in Industry*, vol. 131, 2021, doi: 10.1016/j.compind.2021.103498.

[38] S. Natha, "A systematic review of anomaly detection using machine and deep learning techniques," *Quaid-e-Awam University Research Journal of Engineering, Science & Technology*, vol. 20, no. 1, pp. 83–94, 2022, doi: 10.52584/qrj.2001.11.

[39] S. Mekruksavanich and A. Jitpattanakul, "LSTM networks using smartphone data for sensor-based human activity recognition in smart homes," *Sensors*, vol. 21, no. 5, pp. 1–25, 2021, doi: 10.3390/s21051636.

[40] S. Mohsen, "Recognition of human activity using GRU deep learning algorithm," *Multimedia Tools and Applications*, vol. 82, no. 30, pp. 47733–47749, 2023, doi: 10.1007/s11042-023-15571-y.

[41] K. S. Low and S. K. Eng, "Performance evaluation of deep learning techniques for human activity recognition system," *Journal of Physics: Conference Series*, vol. 2641, no. 1, 2023, doi: 10.1088/1742-6596/2641/1/012012.

[42] S. Sabbu and V. Ganesan, "LSTM-based neural network to recognize human activities using deep learning techniques," *Applied Computational Intelligence and Soft Computing*, 2022, doi: 10.1155/2022/1681096.

[43] P. Maragathavalli, R. Vivitha, M. Saranya, and U. Dinesh, "Incorporating LSTM method on modified deep learning technique for sensor based human activity recognition system," *IOP Conference Series: Materials Science and Engineering*, 2021, doi: 10.1088/1757-899X/1065/1/012045.

[44] I. U. Khan, S. Afzal, and J. W. Lee, "Human activity recognition via hybrid deep learning based model," *Sensors*, vol. 22, no. 1, 2022, doi: 10.3390/s22010323.

[45] S. W. Pienaar and R. Malekian, "Human activity recognition using LSTM-RNN deep neural network architecture," in *2019 IEEE 2nd Wireless Africa Conference, WAC 2019*, 2019, pp. 1–5, doi: 10.1109/AFRICA.2019.8843403.

[46] B. Oluwalade, S. Neela, J. Wawira, T. Adejumo, and S. Purkayastha, "Human activity recognition using deep learning models on smartphones and smartwatches sensor data," in *HEALTHINF 2021 - 14th International Conference on Health Informatics; Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021*, 2021, pp. 645–650, doi: 10.5220/0010325906450650.

[47] A. Kh Aboo and L. M. Ibrahim, "Human activity recognition using a hybrid CNN-LSTM deep neural network," *Webology*, vol. 19, no. 1, 2022.

[48] D. Gaur and S. K. Dubey, "Development of activity recognition model using LSTM-RNN deep learning algorithm," *Journal of Information and Organizational Sciences*, vol. 46, no. 2, pp. 277–291, 2022, doi: 10.31341/jios.46.2.1.

[49] S. Gupta, "Deep learning based human activity recognition (HAR) using wearable sensor data," *International Journal of Information Management Data Insights*, vol. 1, no. 2, 2021, doi: 10.1016/j.jjimei.2021.100046.

[50] M. A. Khatun *et al.*, "Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 1–16, 2022, doi: 10.1109/JTEHM.2022.3177710.

[51] R. Raj and A. Kos, "An improved human activity recognition technique based on convolutional neural network," *Scientific Reports*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-49739-1.

[52] S. Irfan, N. Anjum, N. Masood, A. S. Khattak, and N. Ramzan, "A novel hybrid deep learning model for human activity recognition based on transitional activities," *Sensors*, vol. 21, no. 24, 2021, doi: 10.3390/s21248227.

[53] X. Yin, Z. Liu, D. Liu, and X. Ren, "A novel CNN-based Bi-LSTM parallel model with attention mechanism for human activity recognition with noisy data," *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-11880-8.

[54] W. Z. Tee, R. Dave, J. Seliya, and M. Vanamala, "A close look into human activity recognition models using deep learning," in *2022 3rd International Conference on Computing, Networks and Internet of Things, CNIOT 2022*, 2022, pp. 201–206, doi: 10.1109/CNIOT55862.2022.00043.

BIOGRAPHIES OF AUTHORS

Bakkala Santha Kumar received M.Tech. in Computer Science and Engineering from JNTU University, Hyderabad. He is now working as an assistant professor in the Department of Computer Science and Engineering at K L Deemed to be University, Vaddeswaram, Vijayawada, Andhra Pradesh. He is a highly accomplished researcher with a strong track record of publications and patents. He is also a globally certified trainer for Azure and AWS. He has 14 years of academic experience. His research interests lie in the areas of image processing, cloud computing, and machine learning. He is currently pursuing a Ph.D. in Department of Computer Science and Engineering at Koneru Lakshmaiah Education Foundation, Vaddeswaram, India. He can be contacted at email: santhakumar.bakkala@gmail.com.

Dr. R. Shankar received B.E. in Computer Science and Engineering from the University of Madras, Chennai, M.Tech. in Computer Science and Engineering from Anna University, Chennai, and Ph.D. in Computer Science and Engineering from Manonmaniam Sundaranar University, Tirunelveli. He is now working as an associate professor in the Department of Computer Science and Engineering at Koneru Lakshmaiah Education Foundation, Vaddeswaram, India. He has 24 years of academic experience. He has published his research work in several reputed journals and conferences in the fields of mobile computing, cloud computing, and artificial intelligence. He can be contacted at email: shankarrajendran75@kluniversity.in.