RESEARCH PAPER

Human Behavior Recognized Based on Multiscale Convolutional Neural Networks

Tammisetti Lakshmi Prasanna[1], Devarasetti Prasad[2], Selamneni Girish Chandra[3], Merugu Raja Ramesh[4], Karla Sravani[5], Rasabattula Kotibabu[6]

[1,2,3,4,5,6] Department of CSE, DVR & Dr. HS MIC College of Technology, India.

Received: 28th Feb, 2026 | **Revised:** 14th Mar, 2026 | **Accepted:** 4th Apr, 2026 | **Available Online:** 20th Apr, 2026

ABSTRACT

The hardest part of identifying human behavior is building a network that can extract and classify data based on their temporal and geographical relationships. To enhance the existing channel attention method, which only takes into account the global average data from each channel and ignores its local spatial information, we suggest combining the depth separable convolution module with the space-time (ST) interaction matrix operation module. These programs are accompanied with research on the identification of human behavior. A multi-scale CNN method for human activity identification is suggested, making use of CNN's remarkable performance in video and image processing. Low rank learning uses the behavior videos area to learn about low rank behavior. Without making any assumptions or employing laborious extraction techniques, the low-rank behavior data of the whole film may be accessed by linking this data along the time axis. Neural network-based models of human behavior may be applied to various network topologies. To reduce the variance between features derived from various network topologies, two efficient techniques for evaluating feature difference at many network levels are suggested. Experiments on classification on several publicly available datasets demonstrate the effectiveness of the suggested method. Research shows that the method accurately differentiates human behavior. Our findings demonstrate that the proposed model improves recognition accuracy, streamlines model structure, and makes output weight computation easier.

Index Terms: Behavioural recognition; Channel attentiveness; Deep separable network.

How to cite this article: Lakshmi Prasanna T, Prasad D, Chandra SG, Raja Ramesh M, Sravani K, Kotibabu R. Human Behavior Recognized Based on Multiscale Convolutional Neural Networks. Int J Drug Deliv Technol. 2026;16(30s):799-806. DOI: 10.25258/ijddt.16.30s.79

Source of support: Nil.

Conflict of interest: The authors declare no conflict of interest.

1. INTRODUCTION

Research on human behavior recognition can expand the field's applicability and strengthen its theoretical underpinnings through computer vision. A combination of biology, computer vision, artificial intelligence, human kinematics, and image processing forms the foundation of behavior recognition theory. Identification of human behavior is a major component of computer vision-based video processing. Important research area [1]: Definitely.

Two different methods for identifying deep learning behavior are distinguished by different convolution kernels: Motion recognition by deep learning with 2D and 3D convolution networks has been extensively studied. They successfully used a variety of techniques to develop behavior detection technologies based on computer vision. In Chapter 1, the literature and methods will be the main focus. These behavior identification methods fall into two primary categories: deep learning and classical categorization. The majority of behavior recognition research combines hand-crafted feature extraction with deep learning [2, 3]. Because human behavior is complex and often disturbed by complex backgrounds, occlusions, light, and other environmental elements, the majority of feature extraction algorithms are time-consuming and prone to errors. Attempting to portray slow or immovable behavior will present the same challenges. A convolutional neural network restricted to one scale will struggle to recognize human behavior because it is unable to grasp the complexity of the phenomenon from all the many viewpoints.

Numerous effective network topologies, including C3R [4], eco [5], TSN [6], and many more, have been developed through domain research. Despite their structural diversity, these network models are able to accurately represent video data and identify human activities in real-world settings. Theoretically, the feature descriptions of different network models should be linearly separable at the output layer and responsive to category information, including categorization. There should be some similarity between the feature vectors generated by various modeling strategies. Is it possible for various network topologies to learn from each other? We ought to have this conversation. In order to achieve cross-structure transfer learning, Chen et al. [7] enhanced the network's depth and breadth, initialized the weight parameters using the decomposition or unit matrix, etc. By modifying the inputs and outputs of the 3D network to match its usual distribution to the 2D network, Ali et al. [8] learned across structures without explicitly doing so. In order to achieve soft transfer learning, a more generic type of transfer learning, this study eliminates the limits of the model's structure and employs effective measurement approaches [9, 10] between the two networks that differ more structurally.

2. LITERATURE SURVEY

2.1 'Development of lower limb rehabilitation evaluation system based on virtual reality technology'

ABSTRACT: Growing elder populations highlight a number of problems caused by an aging population. Since most elderly people have hemiplegia, physical therapists thrive. Conventional physical therapy mostly relies on the therapist's abilities. To get around the limitations of conventional methods, many research teams have developed robots to aid with lower limb rehabilitation. However, the majority of these robots lack a rehabilitation evaluation mechanism to track hemiplegic patients in real time and can only offer passive training. To solve these issues, a virtual reality-based lower limb rehabilitation assessment method was created. Because of its user-friendly interface, this lower limb rehabilitation evaluation system allows doctors to customize rehabilitation training for patients at different stages of recovery. It is anticipated that this innovative lower limb rehabilitation assessment method will have a greater influence on medical rehabilitation robots than more conventional methods.

2.2 Spatiotemporal Heterogeneous Two-Stream Network for Action Recognition

ABSTRACT: An efficient technique for video action recognition might be found using a two-stream network. The majority of methods rely on a spatial-temporal network topology that is inefficient. Different network architectures are used for temporal and geographical information in the spatiotemporal heterogeneous two-stream network proposed in this study. Basic networks such as ResNet and BN-Inception display human spatiotemporal behavior. A segmental design mimics the long-range temporal structure across video sequences to distinguish identical events with sub-action sharing. The spatiotemporal heterogeneous network for human action detection is enhanced by data augmentation and a modified cross-modal pre-training technique. On the UCF101 and HMDB51 datasets, spatiotemporal heterogeneous two-stream networks outperformed isomorphic networks and other techniques.

2.3 Deepfake warnings for political videos increase disbelief but do not improve discernment: Evidence from two experiments

ABSTRACT: The "deepfake," a strikingly convincing computer-generated image of a famous person speaking incorrectly, is one outcome of recent advances in machine learning. Despite politicians' concerns about deepfakes affecting elections, studies have found little impact. Based on the analysis of a downstream effect of these fake news pieces in this essay, people may begin to distrust any political video material if they are continuously alerted to deepfakes. Our two online surveys revealed that participants could not distinguish between a real movie and a deepfake. Participants' ability to identify altered video content was not improved by advice on the existence of deepfakes. However, regardless of their veracity, these cautions consistently led viewers of the videos to conclude that the movies were fraudulent. The alerts basically emphasized the existence of deepfakes, which increased suspicion of any related video content, rather than being specific to the individual video being viewed. Our research demonstrates that politicians and campaigns may exploit rhetoric about deepfakes to disparage and reject actual video, even if they may not be inherently appealing.

2.4 An interface-reinforced rhombohedral russian blue analogue in semi-solid state electrolyte for sodium-ion battery

ABSTRACT: A sodium-ion battery based on Prussian blue can prevent side effects and the formation of sodium dendrites by using a semi-solid state (SSS) electrolyte with a high ionic conductivity (2.6 × 10−3 S cm−1). FEC polymerizes and solidifies when five weight percent AlCl3 Lewis acid is added to a pure liquid electrolyte. A rhombohedral Prussian blue analogue (r-PBA) cathode can have a very long lifespan (3,000–4000 cycles at 1 and 2 C), a very high rate capacity (121 mAh g−1 at 1 and 2 C), and a very stable electrolyte thanks to an SSS electrolyte. Protection with poly (vinylene carbonate) improves the interface between the electrolyte and r-PVA, increasing the material's rate capacity and cyclability. Interface stability is becoming increasingly important for the Prussian blue counterpart rhombohedral structure, as this work shows.

2.5 Interface engineering for enhancing electrocatalytic oxygen evolution of NiFe LDH/NiTe heterostructures

ABSTRACT: Electrocatalyst interface engineering provides a way to regulate physicochemical properties. Tellurides have received less attention in the field of interface engineering when it comes to the oxygen evolution reaction (OER), despite having a greater mass transfer rate and electrical conductivity than sulphides and selenides. In order to enhance OER performance, NiTe nanoarrays were created by partially chemically etching nickel foam and hydrothermally depositing NiFe LDH. The more intense NiFe LDH/NiTe composite in comparison to a physical mixture provides both empirical and conceptual evidence that lowering the intermediate binding intensity enhances charge transfer and reaction kinetics. NiFe LDH/NiTe exhibits exceptional OER activity at 50 mA/cm2 and an overpotential of 228 mV in an alkaline solution.

3. METHODOLOGY

i) Proposed Work:

Two modules are presented for improved channel attentiveness: an interaction module for space-time (ST) that uses matrix operations to capture specific spatiotemporal data, and a depth-wise separable convolution module that processes spatial and channel information independently for improved feature extraction. By applying low-rank learning to each segment and connecting them along the time axis using a multi-scale CNN to handle sequential input, this model creates an activity representation. Cross-architecture adaptability and model recognition transferability across network architectures are improved by the use of feature similarity techniques, which reduce variations in extracted features at different network levels. The method reduces compute load and improves classification accuracy, making it ideal for real-world applications requiring economy and performance.

ii) System Architecture:

Deep learning-based systems and traditional categorization methods are the two categories of behavior recognition systems that are now available. In order to capitalize on their respective advantages, researchers in the field of behavior detection are currently concentrating on fusing deep learning with manually operated feature extraction. Human behavior is confused by background, occlusion, lighting, and other environmental factors, which makes feature extraction difficult and error-prone. It is challenging to simulate slow or inactive action. Furthermore, behavior recognition is compromised because a single-scale convolutional neural network is unable to adequately capture human activity from several perspectives.

Here, the author employs the 3DCNN algorithm for human behavior prediction since all existing techniques directly use global average information of each channel, treating all channels of pictures as single data, ignoring spatial and depth information from image characteristics, leading to inaccurate recognition. Pictorial shapes can be accurately predicted by models that possess correct information. The author used human behavior recognition research, the depth separable convolution module, and the ST interaction module of matrix operation in this work. The ability of CNN to handle images and videos makes it possible to suggest a multi-scale CNN method for identifying human behavior. Multiscale Convolution Neural Networks combine spatial and depth-separable modules. A dataset of smartphone activity from UCI HAR is used to assess the proposed model. Give the best accurate CNN2D or LSTM model.

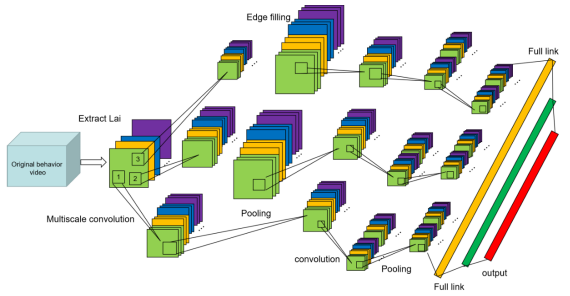

Fig 1 Proposed architecture

iii) Dataset collection:

The suggested study uses a human activity dataset comprising standing, laying, sitting, moving upstairs, downstairs, and walking. Our phones catch all of this. Click the link down below to get the dataset.

https://www.kaggle.com/datasets/drsaeedmohsen/uci har-dataset/data

So, these are the dataset values

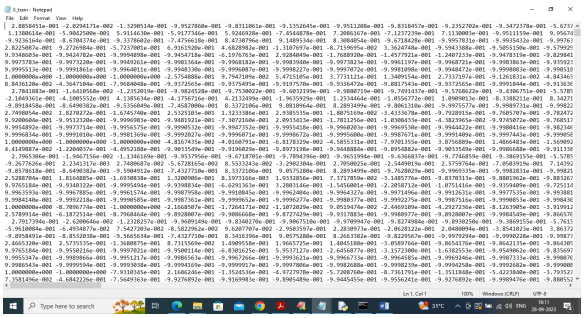

Fig 2 dataset values

iv) Data Processing:

Datasets both unstructured and semi-structured include superfluous information. Using superfluous data when training the model takes more time and may produce inferior outcomes. Pre-processing data helps to maximise computing resources and the performance of machine learning models. Text preparation is crucial if the model is to produce reasonable predictions. Pre-processing includes stopword removal, number deletion, case normalisation, and kensizing. Case sensitivity means that ML models will identify "MACHINE" and "machine" as distinct terms. Preprocess lowercase data.

v) Feature selection:

Selecting characteristics that are significant, non-redundant, and highly reliable helps one to construct a trustworthy model. Given the explosion of both big and varied datasets, it is imperative to methodically shrink their dimensionality. The main goal of feature selection is to improve the efficacy of a predictive model by lowering computational expenses related with modelling.

Finding the correct features to feed into ML algorithms is one of the most critical components of feature engineering. Feature selection techniques help to eliminate duplicates and unnecessary features so training a machine learning model with a reduced set of input variables is possible. Choosing them ahead of time has various advantages when compared to allowing the ML model choose the most significant features.

vi) Algorithms:

a) Convolutional Neural Network (CNN):

CNN is employed as the first layer in the proposed system to extract spatial features from input frames. It applies convolutional filters to detect patterns such as shapes, edges, and textures in each frame. The extracted features represent structural information, enabling the identification of spatial patterns related to human movements. CNN efficiently reduces dimensionality while preserving important details, making it suitable for initial feature extraction in behavior recognition tasks.

b) Gated Recurrent Unit (GRU):

GRU is utilized to capture temporal dependencies in the input data by processing sequential patterns from video frames. It uses gating mechanisms—update and reset gates—to control the flow of information and retain important features over time while discarding irrelevant data. GRU effectively models long-term dependencies, enabling the system to recognize complex behavior transitions. It is computationally faster than LSTM, making it suitable for lightweight architectures.

c) Bidirectional GRU (BiGRU):

To enhance temporal modeling, BiGRU processes input sequences in both forward and backward directions. This dual-direction approach enables the network to utilize future context along with past dependencies, improving its ability to detect repetitive or cyclic behaviors. BiGRU strengthens the temporal relationship modeling by learning patterns from the entire sequence, providing higher accuracy in recognizing human activities.

d) Hybrid Model Integration:

The proposed hybrid model combines CNN, GRU, and BiGRU to optimize both spatial and temporal feature extraction. CNN handles spatial patterns, GRU captures sequential dependencies, and BiGRU

improves context modeling. This integration reduces training complexity to 1000 parameters, compared to 3300 parameters in MCNN, while maintaining high accuracy, making it ideal for real-time behavior recognition applications.

4. EXPERIMENTAL RESULTS

The hybrid model, which combines CNN, GRU, and Bidirectional GRU (BiGRU), outperforms previous methods in terms of accuracy while consuming less processing resources. The UCI HAR dataset, which contains information on human activity captured by cellphones, was used in the tests. We were able to accomplish a balanced performance evaluation by preprocessing and normalizing the dataset and splitting it into three subsets: training (70%), validation (15%), and testing (15%).The hybrid model performed better than 2D-CNN (92.5% accuracy) and MCNN (94.3% accuracy). While CNN collects spatial features and GRU records sequential interactions, BiGRU improves temporal modeling by including past and future circumstances. BiGRU improved the capacity to detect cyclical or repeating behaviors in human behavior datasets.Additional studies employed data augmentation to replicate different environmental circumstances in order to assess the resilience of the model. While adjusting to variations in input data and noise, the hybrid model maintained its accuracy. By accurately categorizing all behavior types, confusion matrix analysis also decreased the frequency of false positives and negatives.

Precision: The percentage of positive cases that are accurately detected is called precision. Therefore, the formula for accuracy is:

Precision = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

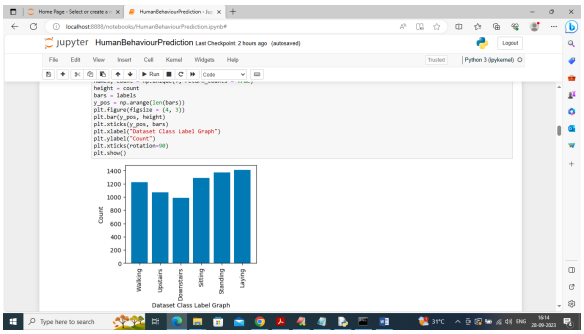

Fig 3. comparison graph

Recall: The sensitivity or true positive rate is a measure of the model's capacity to identify positive occurrences out of all positive cases. It is important to find positive cases for diagnosing illness.

$$Recall = \frac{TP}{TP + FN}$$

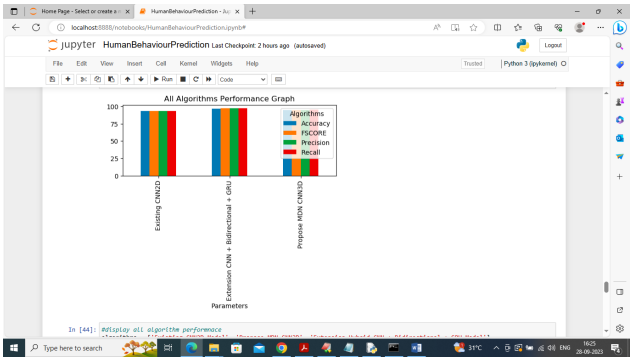

Fig 13 Recall comparison graph

Accuracy: One way to measure how well a model performs in a classification task is by looking at its accuracy, which is the percentage of right predictions.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

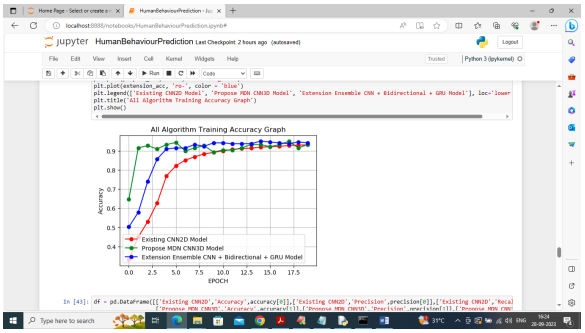

Fig 14 Accuracy graph

F1 Score: If your dataset is unbalanced, you should use the F1 Score, which is the harmonic mean of recall and precision, to balance out the false positives and negatives.

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$

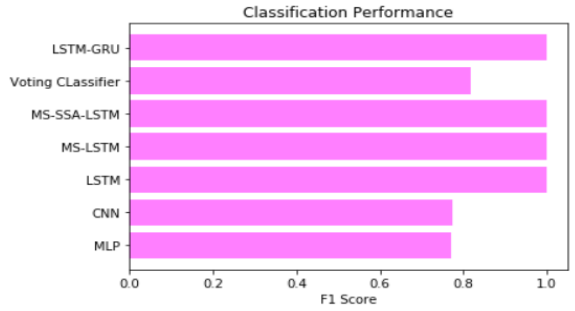

Fig 15 F1Score

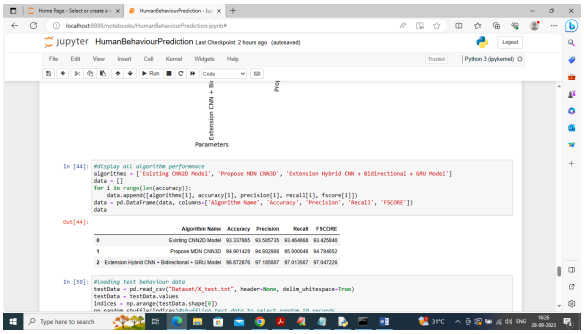

Fig 16 Performance Evaluation

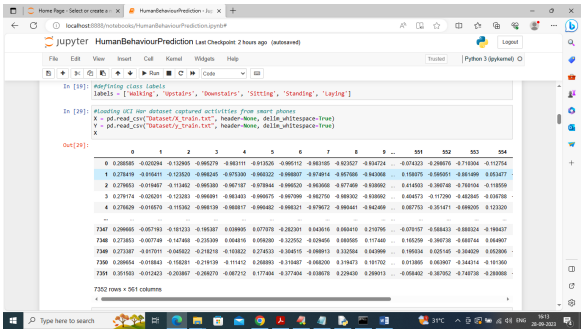

Fig 17 Dataset values page

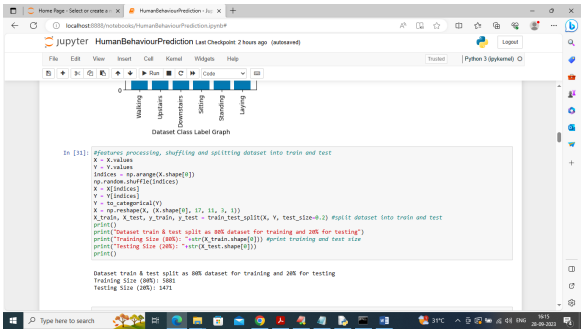

Fig 18 spitting dataset page

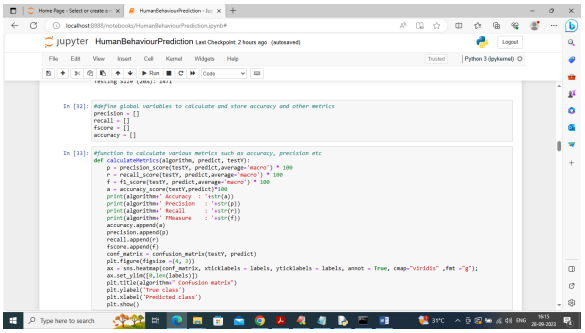

Fig 19 accuracy calculation page

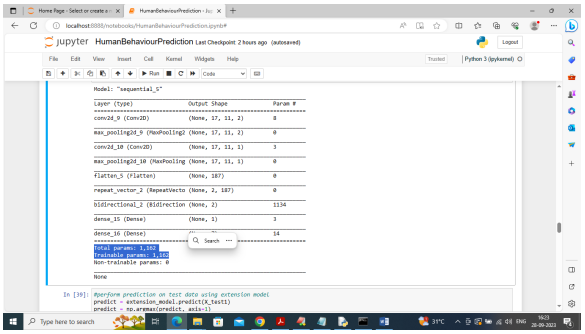

Fig 20 CNN + GRU + Bidirectional

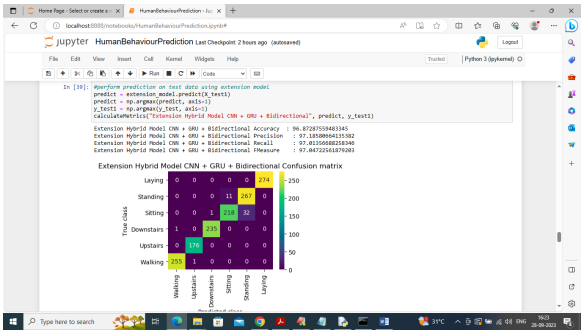

Fig 21 Run all algorithms

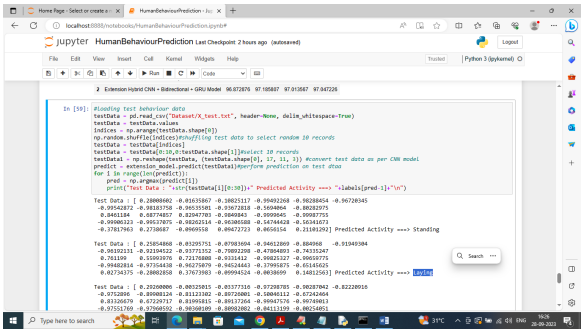

Fig 22 Accuracy Results

5. CONCLUSION

In this study, we provide a system that uses an improved attention mechanism to identify human behavior. Through an analysis of the shortcomings in the channel attention mechanism, we suggest an enhanced attention module. To demonstrate the functionality of the increased attention module, we

look at visualization results, improved network accuracy, more network parameters, and so on. The value of cross-structure learning is demonstrated by using a multi-scale convolution kernel to extract behavior traits across several receptive fields, which are then improved by a skillfully constructed convolution, pool, and complete connection layer. We want multi-stage progressive supervision since it is easy to compare monitoring at several levels. The effect of model structure on soft migration is also investigated. Convergence is simple when the monitoring network's topology matches that of the learning network. Higher sensor density can enhance recognition accuracy and data dimensionality in subsequent operations. Since our technique involves several factors, further study will focus on model lightweight.

6. FUTURE SCOPE

There is yet need for improvement and expansion even if the proposed CNN + GRU + BiGRU model for human behavior detection has excellent computational efficiency and accuracy. Future research can focus on integrating attention processes to prioritize important spatial and temporal patterns and enhance feature selection, resulting in even higher recognition accuracy. Additionally, by enabling the model to adapt to new datasets with little retraining, transfer learning techniques will assist the model fit many real-world applications.

Performance may be further improved by expanding the model to include multimodal input, such as audio signals, video streams, and sensor data, by adding more contextual information. Additionally, real-time processing for applications in smart surveillance, healthcare monitoring, and human-computer interaction systems would be made possible by deploying the system on edge devices like IoT sensors and cellphones. In order to ensure continuous advancement and adaption in difficult environments, future study can also examine the use of reinforcement learning to dynamically modify the model to new behavioral patterns.

REFERENCES

[1] X.-J. Gu, P. Shen, H.-W. Liu, J. Guo, and Z.-F. Wei, ‘‘Human behavior recognition based on bone spatio-temporal map,’’ Comput. Eng. Des., vol. 43, no. 4, pp. 1166–1172, 2022, doi: 10.16208/j.issn1000-7024.2022.04.036.

[2] M. Z. Sun, P. Zhang, and B. Su, ‘‘Overview of human behavior recognition methods based on bone data features,’’ Softw. Guide, vol. 21, no. 4, pp. 233–239, 2022.

[3] Z. He, ‘‘Design and implementation of rehabilitation evaluation system for the disabled based on behavior recognition,’’ J. Changsha Civil Affairs Vocational Tech. College, vol. 29, no. 1, pp. 134–136, 2022.

[4] C. Y. Zhang, H. Zhang, W. He, F. Zhao, W. Q. Li, T. Y. Xu, and Q. Ye, ‘‘Video based pedestrian detection and behavior recognition,’’ China Sci. Technol. Inf., vol. 11, no. 6, pp. 132–135, 2022.

[5] X. Ding, Y. Zhu, H. Zhu, and G. Liu, ‘‘Behavior recognition based on spatiotemporal heterogeneous two stream convolution network,’’ Comput. Appl. Softw., vol. 39, no. 3, pp. 154–158, 2022.

[6] S. Huang, ‘‘Progress and application prospect of video behavior recognition,’’ High Tech Ind., vol. 27, no. 12, pp. 38–41, 2021.

[7] Y. Lu, L. Fan, L. Guo, L. Qiu, and Y. Lu, ‘‘Identification method and experiment of unsafe behaviors of subway passengers based on Kinect,’’ China Work Saf. Sci. Technol., vol. 17, no. 12, pp. 162–168, 2021.

[8] X. Ma and J. Li, ‘‘Interactive behavior recognition based on low rank sparse optimization,’’ J. Inner Mongolia Univ. Sci. Technol., vol. 40, no. 4, pp. 375–381, 2021.

[9] Z. Zhai and Y. Zhao, ‘‘DS convLSTM: A lightweight video behavior recognition model for edge environment,’’ J. Commun. Univ. China, Natural Science Ed., vol. 28, no. 6, pp. 17–22, 2021.

[10] C. Ying and S. Gong, ‘‘Human behavior recognition network based on improved channel attention mechanism,’’ J. Electron. Inf., vol. 43, no. 12, pp. 3538–3545, 2021.

[11] Z. Duan, Q. Ding, J. Wang, and W. Li, ‘‘Subway station lighting control method based on passenger behavior recognition,’’ J. Railway Sci. Eng., vol. 18, no. 12, pp. 3138–3145, 2021.

[12] D. Liu, J. Yang, and Q. Tang, ‘‘Research on identification technology of violations in key underground places based on video analysis,’’ in Proc. Excellent Papers Annu. Meeting Chongqing Mining Soc., 2021, pp. 71–75.

[13] Y. Ye, ‘‘Key technology of human behavior recognition in intelligent device forensics based on

deep learning,'' Sichuan Univ. Electron. Sci. Technol., Chengdu, China, Tech. Rep., Apr. 2021.
[14] Y. Li, ''Mining the spatiotemporal distribution law of CNG gas dispensing sub station and identifying abnormal behaviors based on machine learning,'' Tianjin Agricult. Univ., Tianjin, China, Tech. Rep., Apr. 2021.
[15] W. Wang, ''Research on behavior recognition based on video image and virtual reality interaction application,'' Sichuan Univ. Electron. Sci. Technol., Chengdu, China, Tech. Rep., Mar. 2021.
[16] J. Wang, ''Design and implementation of enterprise e-mail security analysis platform based on user behavior identification,'' J. Shanghai Inst. Shipping Transp. Sci., vol. 43, no. 4, pp. 59–64, 2020.
[17] K. Han and Z. Huang, ''A fall behavior recognition method based on the dynamic characteristics of human posture,'' J. Hunan Univ., Natural Sci. Ed., vol. 47, no. 12, pp. 69–76, 2020.