SN Computer Science (2023) 4:600
https://doi.org/10.1007/s42979-023-02031-5

ORIGINAL RESEARCH

Human Activity Identification System for Video Database Using Deep Learning Technique

Ram Kumar Yadav[1] · Subhrendu Guha Neogi[2] · Vijay Bhaskar Semwal[3]

Received: 31 January 2023 / Accepted: 8 June 2023

Abstract

This paper proposes a deep learning-based computation model for different activity recognition. The Motion Video dataset (MoVi) investigates human activity recognition. The MoVi dataset contains 90 volunteers (30 males and 60 females) and 44 activities. The experiment utilized only 20 activities. It contains a series of poses, body movements, high-quality videos, and many activities such as phone talking, stretching, pointing, walking, jogging, and many more. ConvNet or convolutional neural network (CNN) and ConvLSTM2D deep learning models recognize and categorize human activities. The overall accuracy attained by both deep learning models is 84% (ConvNet) and 97% (ConvLSTM2D), respectively. The effectiveness of this research is to improve the classification result of the ConvLSTM2D network with quick processing and decreased execution time. The proposed model has tried to solve the problem of interclass variability, different terrain, and inclination adaptability. The present research reduced the execution time complexity of data and simultaneously improved the model accuracy.

Keywords Human activity · CNN · LSTM · MoVi dataset · Deep learning

Introduction

One of the main reasons human activity recognition [1] is essential due to its effectiveness in safety and security implementations. These technologies can be used to monitor potential safety hazards in the workplace, such as workers operating machinery without proper safety precautions. It can also be used in public spaces to detect suspicious behaviour and potential securing threats. Additionally, human activity recognition can be used in healthcare to monitor patients with physical or cognitive impairments, allowing for early intervention if necessary. These technologies can also be used in sports and fitness to track and analyze movement [2] patterns, allowing for improved training and injury prevention. These technologies can also be used for surveillance and security [3]. Human activity recognition research aims to develop algorithms [4] and systems that can automatically recognize, categorize and understand human behaviour and actions through various datasets. Biometric identification of human activity recognition [5] can be made using physiological or behavioural traits such as walking, jogging, jumping, kicking, sit-down, jumping, and so on [6, 7]. With a growing population, many researchers are provided various automated human activity recognition systems to recognize human activity with various conditions, such as health care monitoring systems, intelligent home monitoring and many more Internet of Things-based monitoring systems [8, 9] which are containing sensors and actuators, etc. the various system for HAR (human activity recognition) are categorized into two parts sensors based and vision based, as depicted in Fig. 1.

This article is part of the topical collection "Machine Intelligence and Smart Systems" guest edited by Manish Gupta and Shikha Agrawal.

✉ Ram Kumar Yadav
yadav20072@gmail.com

Subhrendu Guha Neogi
sgneogi@gwa.amity.edu

Vijay Bhaskar Semwal
vsemwal@gmail.com

[1] Amity University Gwalior, Gwalior, India

[2] Department of Computer Science and Engineering, Amity University Gwalior, Gwalior, India

[3] Department of Computer Science and Engineering, MANIT, Bhopal, India

Published online: 09 August 2023

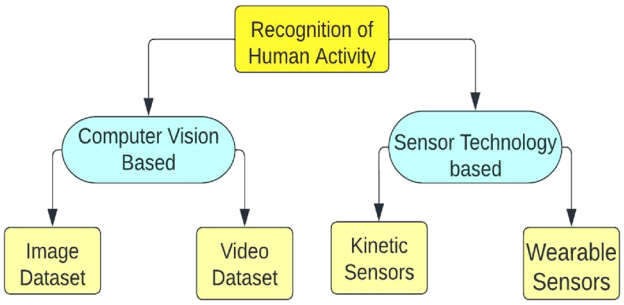

Fig. 1 Human activity recognition approaches

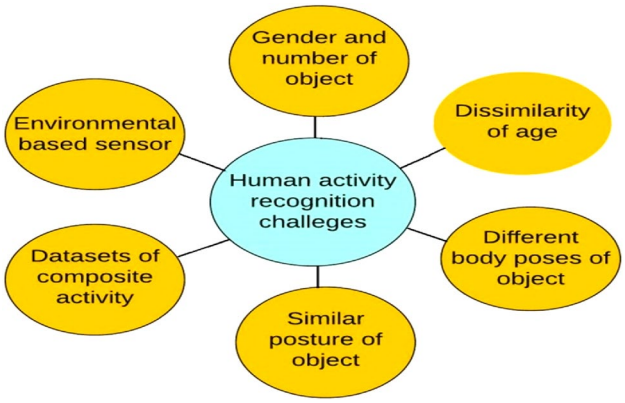

Fig. 2 Challenges in video-based and sensor-based HAR

In computer vision, routine human actions are captured by cameras and video-based systems, with automated recognition based on image sequencing. Due to advancements in microelectronics and computer systems, there has been a significant development in low-power, high-capacity, inexpensive sensors and wired and wireless communication networks [10, 11]. Although the video-based method frequently yields good results indoors, it can achieve a different precision outside or in realistic conditions [12, 13]. Wearable sensors can control physiological characteristics, making them easier to measure. Wearable sensors such as environmental or video sensors are attached [14] to the monitored object and are not dependent on external infrastructure.

Difficulties in Human Activity Recognition

Based on some researcher view, several issues were faced, such as extracting features from the recorded databases, constructing models, and identifying and classifying distinct activities. This research aims to identify the human activities in BML (Bio Motion Lab) video database using various human characteristics. Various sensors [15] are utilized in HAR. HAR is an active and well-liked research area to obtain raw data. These sensors are significant in this domain, and picking the best sensors in the right way might be challenging after reading about several sensing technologies and examining thirty five publications in the last decade. The three main categories of sensors are studied wearable sensors devices such as accelerometers, gyroscopes, magnetometer, GPS, and video sensors device such as cameras fixed in one place to detect actions, and detecting user interaction with the environment using environmental sensors and radio-based sensors have sensed the radio signals like Bluetooth and Wi-Fi [16, 17], etc. and one more infrared sensor, also known as an infrared camera. At the time of preparing the dataset of different volunteers in ideal and unstructured surface using video and sensor-based approaches, facing the various challenges and difficulties are observed, as depicted in Fig. 2 in which it has seven general difficulties and challenges such as dissimilarity of age, similar posture of object, composite human activities etc.

The authors employed a distinct method, such as segmentation, extraction of features, and visualization, to accomplish this activity recognition [18] and evaluate algorithms for recognizing human activity in terms of the challenges and complexity of activity recognition using sensors. The difficulty of the actions can vary and depends on various factors, such as the number of states of activities and kinds of activities, the sensors used, and the protocols used for data gathering. Locomotor activities are categorized into three categories: stationary activity, kinetic activity, and activity involving postural changes, as shown in Fig. 3.

Especially in comparison to kinetic activities (running, walking, etc.), static activities (sleeping, sitting, etc.) and activities involving postural changes are easier to identify. However, due to significant feature space overlap, separating from broadly similar postures, such as standing and sitting, presents significant challenges.

Moreover, dynamic activities such as walking upstairs and downstairs with high feature space similarity are also challenging to distinguish due to comparable movement patterns. Most of the time, completed actions do not coincide with each other during the whole activity period, making the recognition even more difficult. For instance, while sitting and standing are closely connected (therefore, challenging to distinguish), walking is considerably different from both, which means easily separable. Figure 3 represents the three main types of human activity [19] with each sub-category.

The authors [20] proposed a method with deep recurrent neural networks, which were suggesting with maximum throughput from crude accelerometer data, which conformed faster recognition times than another simple technique.

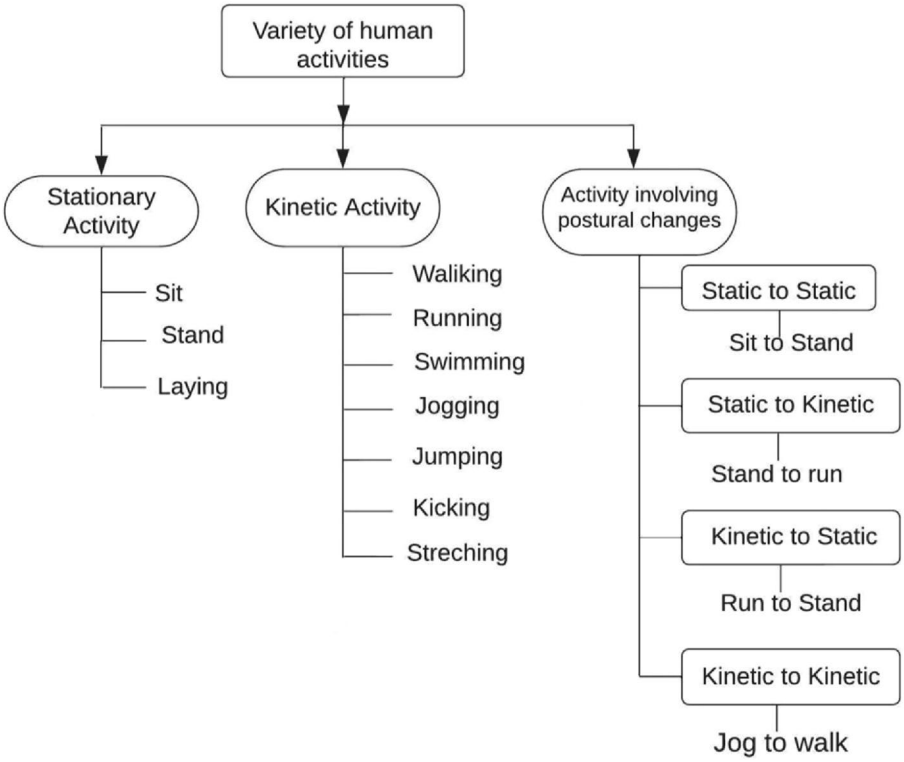

Fig. 3 Categorization of human activity, including the sub-category

Reference [21] suggested combining the Random Forest classifiers with a post-processing method called the Mode of approach for categorizing locomotion and transportation activities.

Objective of the Research

The primary goal of this present work is to offer a method for Human Activity Recognition (HAR). The goal is to offer a deep learning algorithm analysis to deal with HAR. Moreover, extract the most important features out of the Bio-Motion Lab's raw MoVi dataset.

Organization of the Paper

The remainder of this article is structured as follows: The related work is reviewed in the next section. The proposed method and calculation of measuring parameters are covered in "Proposed Method". Result evaluation and discussion with experimental setup and data preparation are included in "Result evaluation and Discussion". "Conclusion and Future Work" addresses the conclusion and future work of the research article.

Related Work

Many deep learning-based techniques have been presented for human activity identification in the last decay. The authors [22] have presented deep learning techniques and performed examinations with published three datasets; after concluding the result, the GRU-based approach was performed with the best result for human activity classification. Authors [16] have proposed Hidden Markov Model (HMM) to identify human activity. HMM is one of the more powerful statistics techniques. The author used the available dataset to find the performance parameters such as accuracy, precision etc.

The authors [23] have provided and demonstrated the 3DCNN + LSTM framework to recognize different human activities, evaluate experimental results with available public datasets and compare the results with existing results.

The authors [24] have suggested a novel method for early finding health-related issues based on human activity or motion recognition patterns. The initial goal is activity detection using motion patterns and deep learning methods. The suggested technique's architecture is made using the pre-processing, engineering layer, and classification layers. The classification of activity uses the CNN layer technique.

A publicly accessible dataset called opportunity was used to train and test the proposed approach and achieved an improved higher accuracy rate of 88.57%. Applications of this novel method include smart homes, intelligent monitoring devices, virtual healthcare, and health advisors.

The authors [25] have investigated whether the HAR problem can be solved using machine learning and deep learning technique, including conventional dimensionality reduction and TDA feature extraction methods. The experiments are carried out with the help of WISDM and UCI-HAR datasets. Various data balancing approaches are used to remove the issue of the imbalanced dataset. In addition to topological data Analysis (TDA), some conventional dimensionality reduction approaches are used. HAR is carried out by seven machine learning (ML) algorithms. Deep learning techniques are also used, including 1DCNN, BiLSTM, and GRU. Three experiments are used: DL, ML with TDA, and ML with standard characteristics. The best reported accuracy and WSM scores for the first category experiments for the WISDM dataset are 99.10% and 86.61%, respectively. The best-reported accuracy and WSM scores for the UCI-HAR dataset were 100% and 100%, respectively. The best-reported results for the WISDM dataset for the second category experiments are accuracy and WSM, which are 95.34% and 89.62%, respectively. The best-reported scores for the UCI-HAR dataset are accuracy and WSM, which are 96.70% and 92.57%, respectively. The accuracy and WSM scores for the third category experiments with the WISDM dataset are best reported at 99.90% and 99.76%, respectively. The best-reported scores for the UCI-HAR dataset were 100% accuracy and 100% WSM, respectively. At the end of the conclusive outcomes, the proposed method is examined with existing research using the same datasets.

The authors [26] have experimented with classifying and predicting human activities using a supervised (XG Boost) machine learning technique. It is the goal of this research paper. The report shows a precision rate of 97% and a recall rate of 97% during the categorization phase, and classification accuracy is 97%. The new one should sprint and produce extremely accurate results compared to previous models.

The authors [27] have provided an overview of human activity recognition while contrasting and analyzing current studies and measures. The methods have recognized and classified abnormal activity using CNN, and LSTM approaches. A brand-new hybrid deep learning structure that combined CNN and LSTM was suggested to combine the extracted feature. First, CNN pre-processed the video and retrieved the visual elements. Then learned, the temporal features of visual elements via LSTM, and add attention mechanism was added to help choose the most crucial elements. An experiment is used on the standard dataset UMN to assess the model's capacity to detect the abnormality.

The authors [7] have provided an overview of extracting and predicting human body motions, frequently occurring indoors, using any embedded hardware device like a camera or sensor device, known as "human activity recognition." Before the advent of smartphones and other personal wearable devices with accelerometer-based sensors that monitor our movements, data collection from sensors was quite expensive. People are very interested in the classification method known as HAR because it allows us to identify different movements of the human body, such as sitting, running, jumping, and jogging, by using wearable sensors like an accelerometer and gyroscope and requesting techniques like convolution neural networks and deep learning techniques. This review examines various human actions, substances, and techniques to identify human activity and body posture.

The authors [28] have proposed modern technology can detect, recognize, and monitor human actions thanks to the widespread usage of HAR and sensor-based data. The status of the human activity recognition publication needed to be revised even though numerous studies and reviews on human activity recognition have already been printed. As a result, this review intends to shed light on the state of the HAR literature as of publications made after 2018. To emphasize application domains, data origins, methodologies, and available investigation issues in human activity recognition, the 95 articles assessed for this study were divided into different categories. Daily living activities have received most of the attention in the literature, followed by user activities centred on particular and group-based activities. Yet, more research must be done on real-time tasks, including surveillance, healthcare, and suspicious activity. Previous studies have extensively used data from mobile sensors and closed-circuit television (CCTV) videos. The most popular methods in the literature examined that are being used for HAR are CNN, LSTM deep learning, and a support vector machine learning technique.

A comparison of the existing research and the current investigation is shown in Table 1.

Proposed Method

Deep learning (DL) is a sub-part of machine learning (ML) that involves artificial neural networks (ANN) to learn from data. Neural networks are made from various layers. The input layer receives the data, and the output layer creates the

Table 1 Comparison of the existing study with the proposed investigation

Initial of authors	Year	Dataset	Methods/models	No. of dataset	Performance metrics
Abbaspour et al. [29]	2020	PAMAP2 (12 activities considered out of 18 activities)	CNN-BiGRUs, CNN-BiLSTM, BiGRU, BiLSTM	1	Accuracy 99.80%, 99.65%, 99.57%, 99.53%, respectively
Bianchi et al. [30]	2019	Self recorded dataset with nine activities	CNN approach	1	Accuracy 97%
Mekruksavanich et al. [31]	2020	WISDM (six activities)	Hybrid LSTM Network	1	Accuracy 96.20%
Barut et al. [32]	2020	Self-recorded dataset with seven activities)	Multi-task LSTM	1	Accuracy 97.76% and F1-score 83.43%
Xia et al. [33]	2021	WISDM, UCI HAR, (six activities)	CNN and LSTM approach	3	Accuracy 95.8%
Present research	2023	MoVi, BML (Twenty activities)	ConvNet, ConvLSTM2D	1	Training Accuracy 84%, 97%. Testing accuracy 74%, 82%

prediction or classification. Among these, one or more hidden layers can transform the inputs. In this paper, two hybrid deep learning approaches are proposed to classify various human activities, defined one by one, and to finalize which is best among them based on its evaluated results.

ConvNet, or convolutional neural network (CNN)**,** is a class of deep neural networks most commonly applied to analyzing visual imagery datasets. ConvNet designs consist of five elements: input phase, convolutional phase, pooling phase, fully connected phase, and output phase. A basic rule for ConvNet architectures is to apply convolutional layers to the input in succession, periodically down-sample the spatial dimensions, and then use Pooling Layers to reduce the numeral of feature mappings [34]. FCL (fully connected layers) layers in a neural network are those where each activation unit on one layer is coupled to every input on the layer above it.

The proposed method modifies the existing ConvNet model, adding extra layers after the existing ConvNet architecture layers. These layers are not part of ConvNet architecture, shown in Fig. 4.

The long short-term memory (LSTM) network is often used for modelling long-term dependencies. ConvLSTM2D [35] is an artificial neural network component that gives the time series character to the convolutional layers, resulting in a model that captures both extended- and short-term dependencies. In short, ConvLSTM2D is the prominent feature of our design, such as All of the inputs X_1, $X_{\_iput}$, cell outputs C_out-1, C_out_t, hidden states Hidden_1, Hidden_t, and gates I_gate, Forget, OP_gate of the ConvLSTM. It is a distinctive characteristic of our architecture (rows and columns). We can visualize the inputs and states as vectors standing on a spatial grid to understand them better. By using the inputs and previous states of its local neighbours, the ConvLSTM predicts the future state of each cell in the grid. Using a convolution operator in the state-to-state and input-to-state transitions makes this simple. In the below, where '$*$' stands for the convolution operator and '$\circ$' (Hadamard product), the main equations of ConvLSTM are displayed in Eqs. (1–5) [36].

$$\text{I\_gate} = \sigma\left(W_{\text{xi}} * X_{\text{iput}} + W_{\text{hi}} * \text{ Hidden}_t - 1 + W_{\text{ci}} \circ C_{\text{out}} - 1 + b_i\right) \tag{1}$$

$$\begin{aligned}\text{Forget } &= \sigma\left(W_{\text{xi}} f * X_{\text{iput}} + W_{hi} * \text{ Hidden\_t}\right. \\ &\left. -1 \ + \ W_{\text{ci}} \circ C_{\text{out}} - 1 + b_f\right)\end{aligned} \tag{2}$$

$$\begin{aligned}\text{C\_out } &= \text{Forget} \circ \text{C\_out\_t} - 1 + \text{I\_gate} \circ \tanh\left(W_{\text{xc}} * X_t\right. \\ &\left. + W_{\text{hc}} * \text{ Hidden}_t - 1 + b_c\right)\end{aligned} \tag{3}$$

$$\sigma\left(W_{\text{xo}} * X_t + W_{\text{ho}} * \text{ Hidden\_t} - 1 + W_{\text{co}} \circ \text{C\_out\_t} + b_o\right) \tag{4}$$

$$\text{Hidden\_t } = \text{OP\_gate} \circ \tanh\{\text{C\_out\_t}\} \tag{5}$$

A ConvLSTM [37, 38] with a larger transitional kernel must capture faster motions if we consider the states as the concealed representations of object tracking. One with a smaller kernel, in contrast, may record slower motions. On a Moving BML dataset, we compared our ConvLSTM network to the ConvNet network to better understand the behaviour of our model. Different layer levels and kernel values are used to run our model.

ConvLSTM2D architecture [39] combines the gating of the LSTM layer with 2D convolutions layers architecture. ConvLSTM layers do a similar task to LSTM [31, 40, 41], but instead of matrix multiplication, it does convolution operations and retains the input dimensions [35].

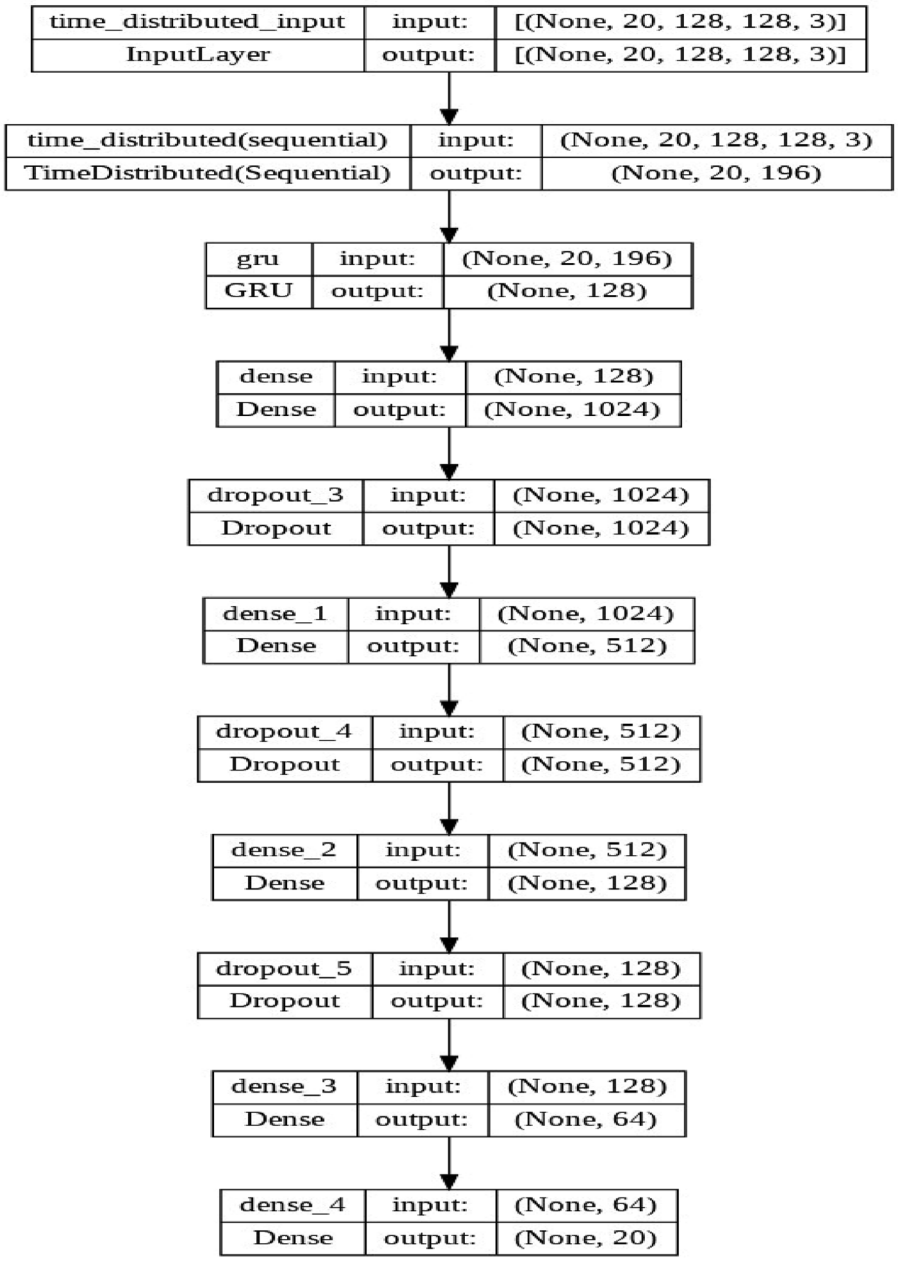

Fig. 4 Proposed architecture of ConvNet model

The input of Keras ConvLSTM layer is a 5_Dimensional with structure (samples, time, channels, rows, cols) if it is first channels, (samples, time, rows, cols, channels) if it is last channels. The ConvLSTM deep learning approach involves using the channel's last ConvLSTM layer with the "data_format" parameter set to "channels_last."

The output of ConvLSTM layer If return_sequence = True, then it is a 5-D tensor with structure (samples, time, filters, rows, cols). If return_sequence = False, it is a 4D tensor in TensorFlow with structure (samples, filters, rows, cols). This ConvLSTM deep learning approach involves using return_sequence = True in the ConvLSTM layer implementation. The model's input is of the shape (None, 20, 128, 128, 3) with 20 frames extracted from each video with each frame of the size 128 × 128. Each frame is of the RGB format having three layers for each layer. It is depicted in Fig. 5 and the data sapling criteria are represented in Table 2.

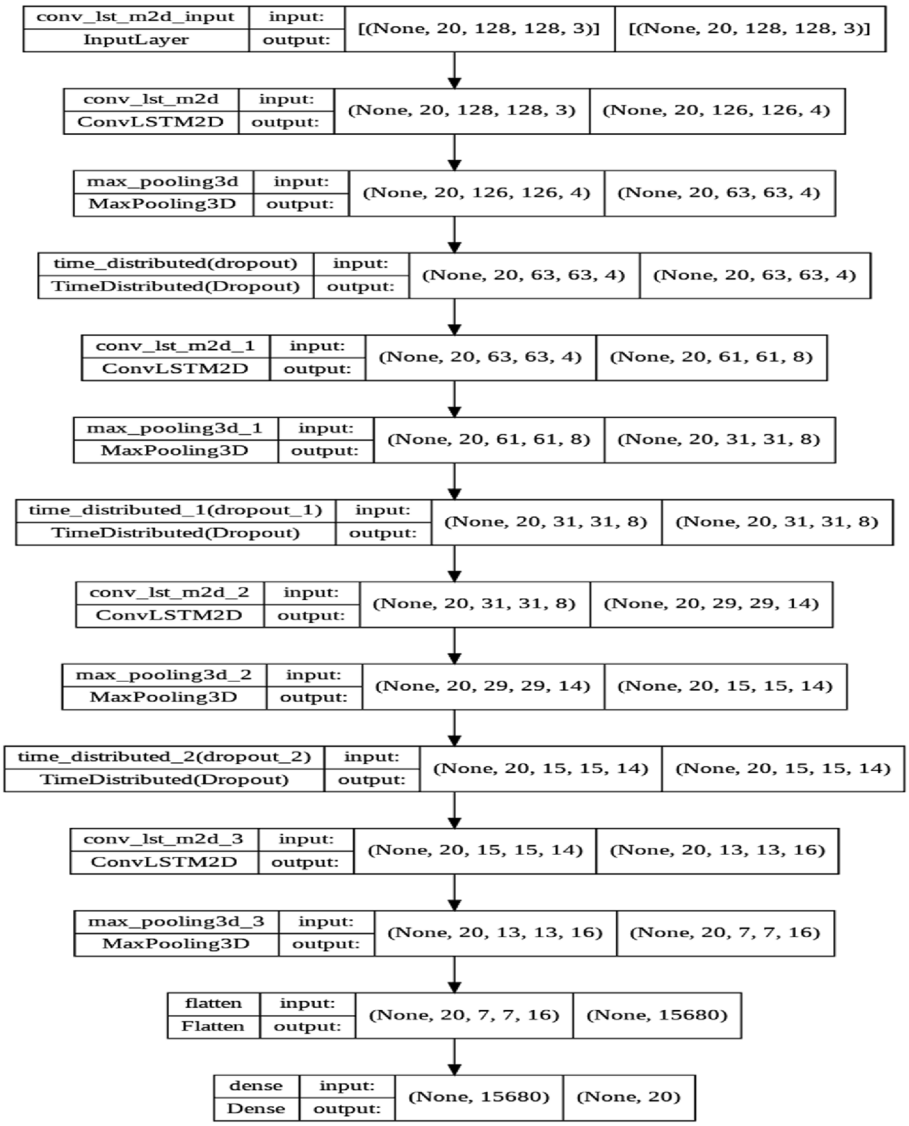

Fig. 5 Proposed architecture of ConvLSTM2D

Table 2 Sampling criteria of data

Database	Model	Size of sample (input shape)	Length of single array (flatten)	Output shape	Number of iteration	Batch size
MoVi dataset published by BML	ConvNet	128*128	64	20	20	32
	ConvLSTM2D	128*128	15,680	20	20	32

Algorithm

Algorithm: An activity Identification system for video Database using a deep learning technique

Input: Activity video_dataset.csv file

Output: Prediction of the accuracy of human activities recognition

Step I: **Begin**
Step II: Generate video data from sensory data
a. Select 20 frames from each video
b. Selecting 10 alternative frames from the selected 20 frames
Step III: Generate and store activity labels and features
Step IV: Array data input to deep learning models such as ConvNet, ConvLSTM2D
Step V: Get evaluation measures and compare the outcomes of each model
Step VI: Build the results tables and plots
StepVII: **Finish**

These deep learning architectures and concepts are relevant to tracking and recognizing human activities. Using these architectures and techniques makes it possible to achieve high accuracy in recognizing and classifying complex human activities from raw sensor data. Finally, ConvLSTM2D handles the BML dataset better than ConvNet, delivering more significant results.

Calculation of Measuring Parameters

Equations 6–9 describe how the evaluation of results depends on different performance measures. True Positive: the deep learning model accurately [42, 43] identified the accurate activity label for the test sample. True Negative: the model successfully excludes the test sample from a specific label. False Positive: an inaccurate label from the actual sample is used to predict the test sample. False Negative: an inaccurate match between a predicted sample and its original label [44].

$$\text{Accuracy} = \frac{\text{True}_{\text{Positive}} + \text{True}_{\text{Negative}}}{\text{True}_{\text{Positive}} + \text{True}_{\text{Negative}} + \text{False}_{\text{Positive}} + \text{False}_{\text{Negative}}} \tag{6}$$

$$\text{Measurement\_Precision} = \frac{\text{True}_{\text{Positive}}}{\text{True}_{\text{Positive}} + \text{False}_{\text{Positive}}} \tag{7}$$

$$\text{Measurement\_Recall} = \frac{\text{True}_{\text{Positives}}}{\text{True}_{\text{Positive}} + \text{False}_{\text{Negative}}} \tag{8}$$

$$\text{Measurement\_F1\_Score} = 2 * \frac{\text{Measurment\_Precision} * \text{Measurment\_Recall}}{\text{Measurment\_Precision} + \text{Measurment\_Recall}} \tag{9}$$

In convolutional neural networks (CNNs), the cross-entropy loss (logistic loss) is a regularly used loss function to determine the gap between the expected probability distribution and the original probability distribution of the output. The goal of training the CNN is to minimize the cross-entropy loss. It means that the aim is to reduce the model loss. On the other hand smaller the loss betters the model. Moreover, a perfect model has zero cross-entropy loss. In the case of the ConvLSTM2D technique, a variant of the LSTM (long short-term memory) model that includes convolutional layers, the cross-entropy loss can be used to train the network for classification tasks such as images or video classification. Equations 10 and 11 describe how the evaluation of the results of model loss.

$$\text{Measurement of loss} = -\sum_{k=1}^{n} \text{X}_{\text{k}} * \log(softmax\_Prob.\_k) \tag{10}$$

$$\text{Measurement of loss} = -[\text{X}_{\text{k}} * \log(softmx\_Prob.\_k) + \text{X}_{1-\text{k}} * \log(softmax\_Prob.\_1-k)] \tag{11}$$

where n is number of classes, X_k is a class, softmax_Prob$_k$ is a softmax probability of Kth class, X_{1-k} is a previous class, softmax_Prob_1$-k$ is the probability of (1$-k$)th class.

Result Evaluation and Discussion

This section explains the database utilized the results of the experiments, and a comparison with other HAR systems.

Experiment Setup

The experimental setup involves using Google Colaboratory, which uses python version 3.7.15 to run the python scripts. TensorFlow 2.9.2 and Keras 2.9.0 are the primary libraries for building the deep learning neural network model and also used GPU support in google colab to increase the training and testing procedure speed. Google colab provides Intel®

Table 3 Description of forty four activities including video of each activity

Human activity	Recorded video	Human activity	Recorded video	Human activity	Recorded video	Human activity	Recorded video
bicep_curls_rm	1	free_throw_rm	1	pointing	81	stretching	81
checking_watch	81	front_swimming_rm	4	punching_rm	2	stretching_rm	2
coughing_rm	1	hand_clapping	81	pushups_rm	2	swinging_arms_rm	4
crawling	81	hand_waving	81	random_motion_rm	9	swinging_racket_rm	3
crossarms	14	jogging	81	rowing_rm	2	taking_photo	81
cross_arms	67	juggling_rm	1	running_in_spot	81	throwing_frisbee_rm	1
cross_legged_sitting	81	jumping_jack	14	scratching_head	80	throw_catch	81
dancing_rm	29	jumping_jacks	67	serving_rm	1	vertical_jumping	81
dj-ing_rm	1	kicking	81	sideways	81	walking	81
dribbling_rm	2	lunges_rm	1	sitting_down	81	wearing_belt_rm	1
fencing_rm	1	phone_talking	81	squatting_rm	10	yoga_rm	2

Table 4 The databases that were used in this paper

Database	Class(activity) Detail	Name of used classes (activities)	Recorded video	Name of used classes (activities)	Recorded video	Degree of record	Feature
MoVi dataset published by BML	20 activity Out of 44 Activity (Video of each activity ≥ 50)	Phone_talking	81	Jogging	81	540,000	20
		Strething	81	Trow_catcch	81		
		Running_in_spot	81	Hand_waving	81		
		Pointing	81	Cross_arms	67		
		Cross_legged_sitting	81	Vertical_jumping	81		
		Hand_clapping	81	Checking_watch	81		
		Sitting_down	81	Talking_photo	81		
		Jumping_jacks	67	Scratching_head	80		
		Walking	81	Sideways	81		
		Kicking	81	Crawling	81		

Xenon® CPU @2.00 GHz with a total RAM of 12 GB. Google colab uses Linux based operating system to execute its processes. Tesla T4 GPU supported by colab provides CUDA Version 11.2 and 15 GB of graphic memory, which reduces the time complexity for training and testing our deep learning model.

Dataset and Result Analysis

The dataset used in this experiment is the BML MoVi dataset produced by BIO MOTION LAB. It contains synchronized poses, body meshes and video recordings. This dataset has videos for 44 different activities captured from four distinct points of view for 60 women and 30 men candidates performing day-to-day actions and sporting acts like kicking, sitting down, walking, etc. which is shown in Table 3. Twenty activities out of 44 were selected, for each activity has more than 50 videos available for better training of the model [2] and the Table 4 represents the whole detail of MoVi dataset. Made stick-figure video recordings of each participant performing each activity using the available sensor movement data and then fed it into our deep-learning model for activity classification. After finishing the pre-processing, the human activity dataset is partitioned into two phases, of which 80% is used for model training, and 20% is used for model testing. To evaluate the general measures such as accuracy, precision, recall, and F1_score of the suggested approach by combing different optimizers such as Adam, Adamax, RMSprop, and Nadam, shown in Figs. 6, 7, and Table 1, which clearly shows the incremented accuracy of ConvLSTM2D with RMSprop, which equals 97% in the period of training and 82% in the phase of testing, which is greater than the proposed ConvNet approach with various optimizers.

Using ConvNet, we have achieved maximum accuracy of 84%, but in ConvLSTM2D, we have achieved an average of 97% accuracy with the RMSprop optimizer.

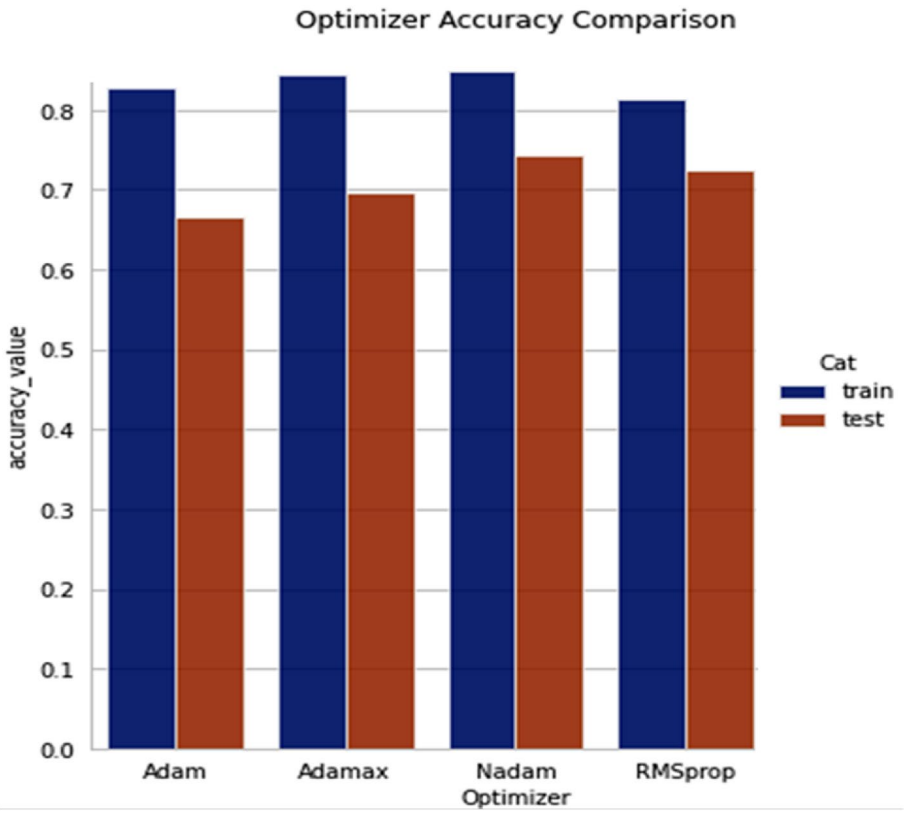

Fig. 6 Histogram of optimizer's accuracy in the form of training and testing

In implementing the ConvLSTM2D model, we have to use various optimizers to achieve better results with various measurements such as accuracy, precision, recall, and F1_Score. After performing the analysis process, it is represented in Figs. 8 and 9.

Figure 10a represents the confusion matrix for all 20 activities in the ConvLSTM2D model (training), and Fig. 10b represents the confusion matrix for all 20 activities in the ConvLSTM2D model (testing).

The result is concluded by using a video dataset of each activity. The achieved model accuracy with various optimizers such as Adam, Adamax, RMSprop, and Nadam are displayed in Table 5, including precision, recall, and F1_score, as shown in Table 6.

Conclusion and Future Work

The article shows deep learning techniques for tracking and recognizing human activities. The extraction and recognition of features from 20 human activities are investigated, and proposed methods are tested on the MoVi dataset, which contains 60 female and 30 male volunteers. Participants were randomly chosen to create the dataset to test the strategy's effectiveness. The proposed technique can deal with different terrain, inclination and viewpoint challenges. The final accuracy result is achieved at 97% from the ConvLSTM2D network, and the measuring parameters (Precision, Recall, and F1_score) still need to be increased. The proposed techniques solved the problem of interclass variability and below given points clearly explain the novel contribution of the research work.

1. It is necessary to reduce the computational cost of models regarding memory, CPU, sensors, and battery utilization.
2. The proposed approach has achieved the best recognition accuracy, precision, and resource utilization results.
3. A classifier-based approach can also be used to accurately identify the most similar activity, such as standing and sitting or walking, walking upstairs, and walking downstairs. Most previous investigations have needed help to identify similar activities.
4. Investigation can be performed on integrating sensory, video, and similar activity data with a shorter execution time than the existing method.

	precision	recall	f1-score	support
phone_talking	0.86	0.93	0.89	27
stretching	0.96	1.00	0.98	22
running_in_spot	0.96	1.00	0.98	25
pointing	1.00	0.96	0.98	25
cross_legged_sitting	0.93	1.00	0.97	28
hand_clapping	1.00	0.85	0.92	27
sitting_down	0.88	0.96	0.92	23
jumping_jacks	1.00	1.00	1.00	26
walking	1.00	0.92	0.96	24
kicking	0.95	0.83	0.88	23
jogging	0.96	1.00	0.98	27
throw_catch	0.97	0.97	0.97	30
hand_waving	0.93	1.00	0.96	27
cross_arms	1.00	0.96	0.98	28
vertical_jumping	1.00	1.00	1.00	28
checking_watch	1.00	1.00	1.00	27
taking_photo	1.00	1.00	1.00	28
scratching_head	1.00	0.96	0.98	28
sideways	1.00	1.00	1.00	28
crawling	0.96	1.00	0.98	24
accuracy			0.97	525
macro avg	0.97	0.97	0.97	525
weighted avg	0.97	0.97	0.97	525

(A)

	precision	recall	f1-score	support
phone_talking	0.88	0.88	0.88	8
stretching	1.00	1.00	1.00	13
running_in_spot	0.90	0.90	0.90	10
pointing	0.83	1.00	0.91	10
cross_legged_sitting	0.86	0.86	0.86	7
hand_clapping	0.38	0.38	0.38	8
sitting_down	0.92	0.92	0.92	12
jumping_jacks	1.00	1.00	1.00	9
walking	1.00	0.73	0.84	11
kicking	0.62	0.42	0.50	12
jogging	0.88	0.88	0.88	8
throw_catch	0.71	1.00	0.83	5
hand_waving	0.58	0.88	0.70	8
cross_arms	0.67	0.57	0.62	7
vertical_jumping	0.83	0.71	0.77	7
checking_watch	0.78	0.88	0.82	8
taking_photo	0.71	0.71	0.71	7
scratching_head	1.00	0.57	0.73	7
sideways	0.70	1.00	0.82	7
crawling	1.00	1.00	1.00	11
accuracy			0.82	175
macro avg	0.81	0.81	0.80	175
weighted avg	0.83	0.82	0.81	175

(B)

Fig. 7 Percentage of accuracy (training and testing) of ConvLSTM2D Model using RMSprop optimizer

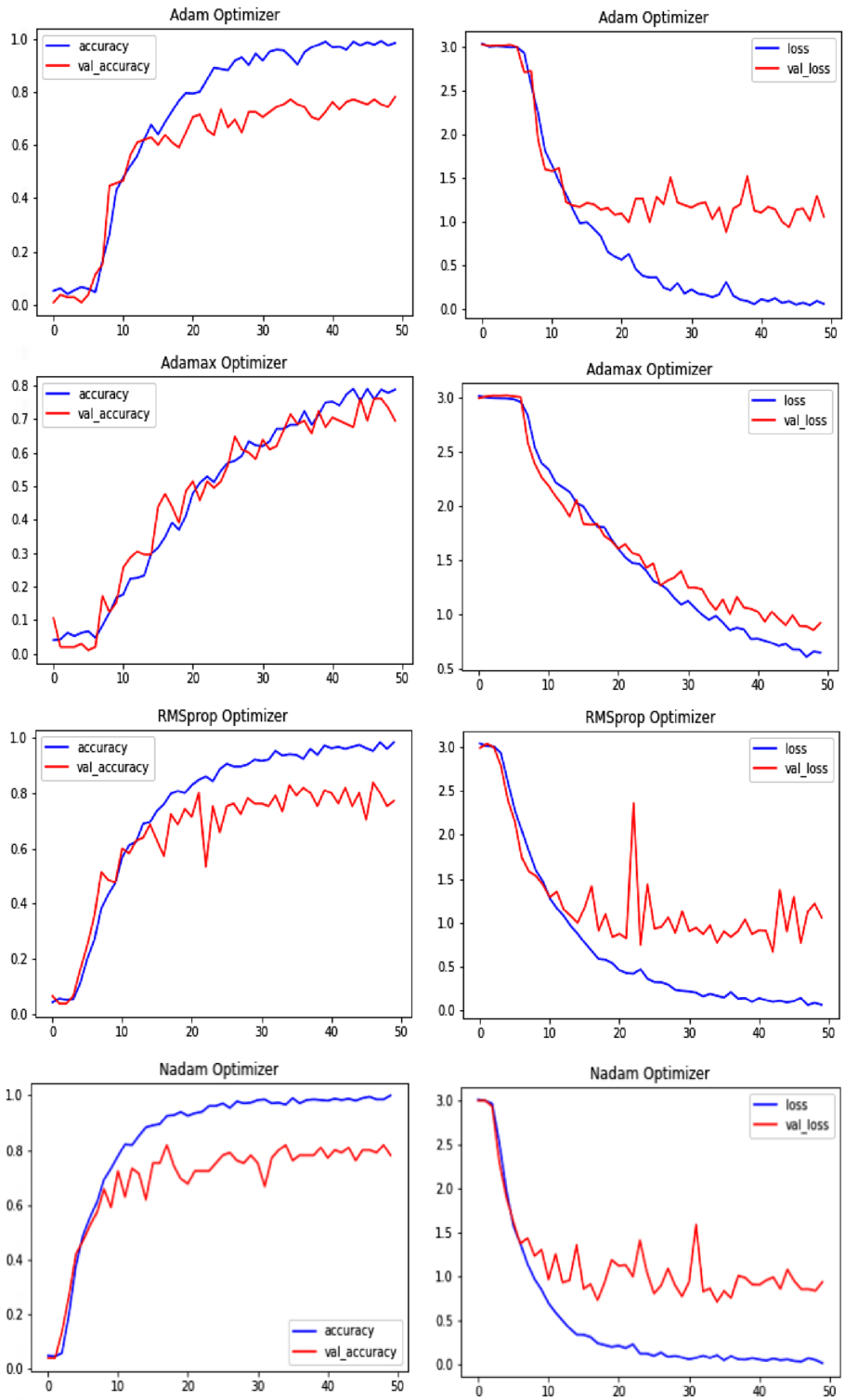

Fig. 8 Accuracy and loss represent in the form of a line graph of each optimizer

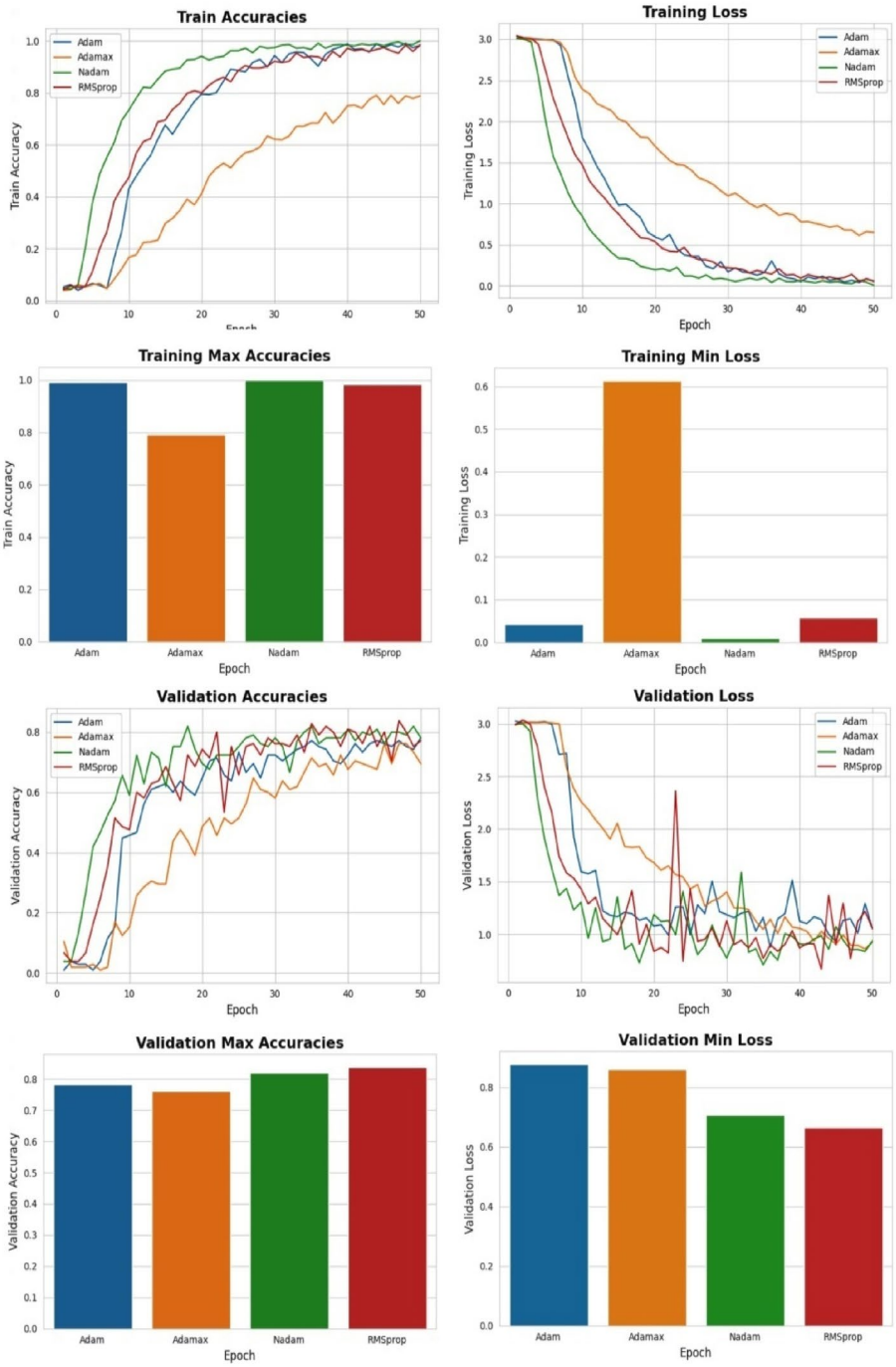

Fig. 9 Line and histogram plot for training accuracy and loss, training max accuracy and min loss, Validation accuracy and loss, and Validation max accuracy and min loss

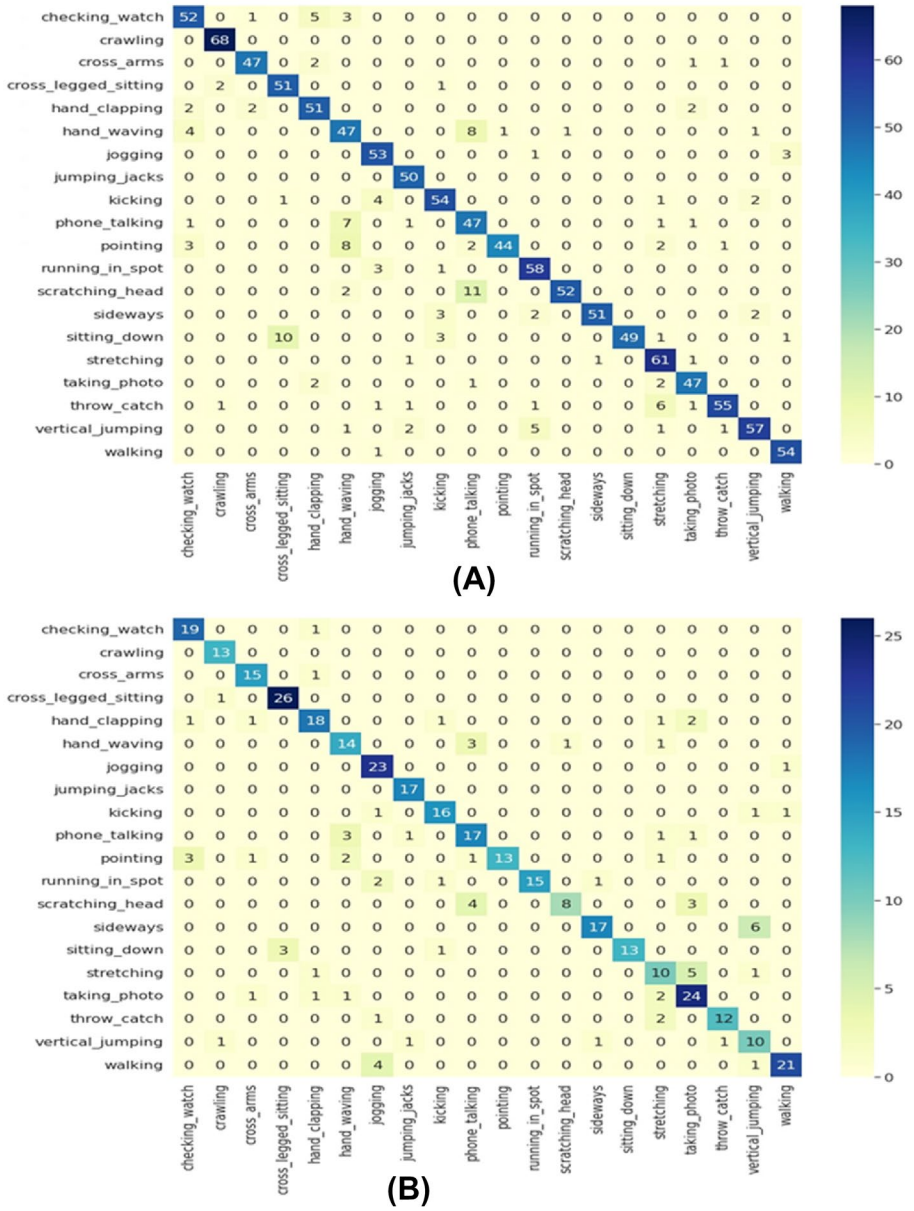

Fig. 10 **a** Training data (confusion matrix). **b** Testing data (confusion matrix)

Table 5 Comparative outcome for the proposed model using different optimizers

Method/technique	Optimizer	Percentage of accuracy (%)	
		Training	Testing
ConvNet model	Using Adam	82	66
	Using Adamax	84	69
	Using RMSprop	81	72
	Using Nadam	84	74
ConvLSTM2D model	Using Adam	96	81
	Using Adamax	94	86
	Using RMSprop	97	82
	Using Nadam	88	81

Table 6 Comparative outcomes of measuring parameters for the proposed model

Method/technique	Percentage of measuring parameters (%)				
	Accuracy		Precision	Recall	F1_score
ConvNet model	Train	84	81	81	80
	Test	74			
ConvLSTM2D model	Train	97	97	97	97
	Test	82			

Acknowledgements We would like to thank the Bio Motion Lab for providing the human activity MoVi dataset, which provides a video database of twenty human activities, and allowing for the use of the dataset in this paper.

Funding SERB, DST of government of India, ECR/2018/000203 ECR dated 04-June-2019.

Data Availability BML published the human activity MoVi video dataset.

Declarations

Conflict of Interest There is no conflict of interest.

Human Participants or Animals Not applicable.

Informed Consent Not relevant.

References

1. Ranasinghe S, Al-MacHot F, Mayr HC. A review on applications of activity recognition systems with regard to performance and evaluation. Int J Distrib Sens Netw. 2016. https://doi.org/10.1177/1550147716665520.
2. Friday NH, Al-Garadi MA, Mujtaba G, Alo UR, Waqas A. Deep learning fusion conceptual frameworks for complex human activity recognition using mobile and wearable sensors. In: 2018 Int. Conf. Comput. Math. Eng. Technol. Inven. Innov. Integr. Socioecon. Dev. iCoMET 2018—Proc., vol. 2018-Jan, pp. 1–7, 2018, https://doi.org/10.1109/ICOMET.2018.8346364.
3. Jordao A, Torres LAB, Schwartz WR. Novel approaches to human activity recognition based on accelerometer data. Signal, Image Video Process. 2018;12(7):1387–94. https://doi.org/10.1007/s11760-018-1293-x.
4. De-La-Hoz-Franco E, Ariza-Colpas P, Quero JM, Espinilla M. Sensor-based datasets for human activity recognition—a systematic review of literature. IEEE Access. 2018;6(c):59192–210. https://doi.org/10.1109/ACCESS.2018.2873502.
5. Yadav RK, Neogi SG, Semwal VB. Special session on recent advances in computational intelligence & technologies (SS_10_RACIT). In: Proceedings of Third International Conference on Computing, Communications, and Cyber-Security. Springer, Singapore. 2023; 595–608. https://doi.org/10.1007/978-981-19-1142-2_47.
6. Ghorbani S, Mahdaviani K, Thaler A, Kording K, Cook DJ, Blohm G, Troje NF. Movi: a large multipurpose motion and video dataset. arXiv preprint 2020. https://arXiv.org/2003.01888. https://doi.org/10.1371/journal.pone.0253157.
7. Saini R, Maan V. Human activity and gesture recognition: a review. In: 2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3) 2020; 1–2. IEEE. https://doi.org/10.1109/ICONC345789.2020.9117535.
8. Bouchabou D, Nguyen SM, Lohr C, LeDuc B, Kanellos I. A survey of human activity recognition in smart homes based on IoT sensors algorithms: taxonomies, challenges, and opportunities with deep learning. Sensors. 2021;21(18):6037. https://doi.org/10.3390/s21186037.
9. Zhang X, Zhang H, Hu J, Zheng J, Wang X, Deng J, Wang Y. Gait pattern identification and phase estimation in continuous multilocomotion mode based on inertial measurement units. IEEE Sens J. 2022. https://doi.org/10.1109/JSEN.2022.3175823.
10. Antar AD, Ahmed M, Ahad MAR. Challenges in sensor-based human activity recognition and a comparative analysis of benchmark datasets: a review. In: *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 2019; 134–139. IEEE. https://doi.org/10.1109/ICIEV.2019.8858508.
11. Ding D, Cooper RA, Pasquina PF, Fici-Pasquina L. Sensor technology for smart homes. Maturitas. 2011;69(2):131–6. https://doi.org/10.1016/j.maturitas.2011.03.016.
12. Tan TH, Gochoo M, Huang SC, Liu YH, Liu SH, Huang YF. Multi-resident activity recognition in a smart home using RGB activity image and DCNN. IEEE Sens J. 2018;18(23):9718–27. https://doi.org/10.3390/s21186037.
13. Liu J, Shahroudy A, Xu D, Kot AC, Wang G. Skeleton-based action recognition using spatiotemporal LSTM network with trust gates. IEEE Trans Pattern Anal Mach Intell. 2017;40(12):3007–21. https://doi.org/10.1109/TPAMI.2017.2771306.
14. Uddin MZ, Hassan MM. Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network. IEEE Sens J. 2018;19(19):8413–9.
15. Khandnor P, Kumar N. A survey of activity recognition process using inertial sensors and smartphone sensors. In: 2017 International Conference on Computing, Communication and Automation (ICCCA). 2017; 607–612. IEEE. https://doi.org/10.1109/JSEN.2018.2871203.
16. Manouchehri N, Bouguila N. Human activity recognition with an HMM-based generative model. Sensors. 2023;23(3):1390.
17. Kwapisz JR, Weiss GM, Moore SA. Activity recognition using cell phone accelerometers. ACM SIGKDD Explor Newsl. 2011;12(2):74–82.
18. Bhardwaj R, Singh PK. Analytical review on human activity recognition in video. In: 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence) 2016; 531–536. IEEE.
19. Khan YA, Imaduddin S, Singh YP, Wajid M, Usman M, Abbas M. Artificial intelligence based approach for classification of human activities using MEMS sensors data. Sensors. 2023;23(3):1275.
20. Inoue M, Inoue S, Nishida T. Deep recurrent neural network for mobile human activity recognition with high throughput. Artif Life Robot. 2018;23(2):173–85.
21. Antar AD, Ahmed M, Ishrak MS, Ahad MAR. A comparative approach to classification of locomotion and transportation modes using smartphone sensor data. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers 2018; 1497–1502.
22. Wang X, Shang J. Human activity recognition based on two-channel residual–GRU–ECA module with two types of sensors. Electronics. 2023;12(7):1622.
23. Vrskova R, Kamencay P, Hudec R, Sykora P. A new deep-learning method for human activity recognition. Sensors. 2023;23(5):2816.
24. Javeed M, Jalal A. Deep activity recognition based on patterns discovery for healthcare monitoring. In: 2023 4th International Conference on Advancements in Computational Sciences (ICACS), 2023;1–6. IEEE.
25. Balaha HM, Hassan AES. Comprehensive machine and deep learning analysis of sensor-based human activity recognition. Neural Comput Appl. 2023;35:1–39.
26. Khattak NY, Hussnain EG, Ahmad W, Islam SU, Haq IU. Computer vision-based human activity classification and prediction. The Sciencetech, 4(1).
27. Fu Y, Liu T, Ye O. Abnormal activity recognition based on deep learning in crowd. In: 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), vol 1. 2019; 301–304). IEEE.
28. Arshad MH, Bilal M, Gani A. Human activity recognition: review, taxonomy and open challenges. Sensors. 2022;22(17):6463.

29. Abbaspour S, Fotouhi F, Sedaghatbaf A, Fotouhi H, Vahabi M, Linden M. A comparative analysis of hybrid deep learning models for human activity recognition. Sensors. 2020;20(19):5707.
30. Bianchi V, Bassoli M, Lombardo G, Fornacciari P, Mordonini M, De Munari I. IoT wearable sensor and deep learning: an integrated approach for personalized human activity recognition in a smart home environment. IEEE Internet Things J. 2019;6(5):8553–62.
31. Mekruksavanich S, Jitpattanakul A. Smartwatch-based human activity recognition using hybrid lstm network. In: 2020 IEEE SENSORS, 2020; 1–4. IEEE.
32. Barut O, Zhou L, Luo Y. Multitask LSTM model for human activity recognition and intensity estimation using wearable sensor data. IEEE Internet Things J. 2020;7(9):8760–8.
33. Xia K, Huang J, Wang H. LSTM-CNN architecture for human activity recognition. IEEE Access. 2020;8:56855–66.
34. Huang CD, Wang CY, Wang JC. Human action recognition system for elderly and children care using three stream convnet. In: 2015 International Conference on Orange Technologies (ICOT), 2015; 5–9. IEEE.
35. Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Adv Neural Inform Process Syst, 2015; 28.
36. Hu WS, Li HC, Ma TY, Du Q, Plaza A, Emery WJ. Hyperspectral image classification based on tensor-train convolutional long short-term memory. In: IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, 2020; 858–861. IEEE.
37. Shah D, Malhotra A, Patidar M. Sensor-based human activity recognition by multi-headed ConvLSTM. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 1, 2021; 1127–1129. IEEE.
38. Tsutsumi H, Kondo K, Takenaka K, Hasegawa T. Sensor-based activity recognition using frequency band enhancement filters and model ensembles. Sensors. 2023;23(3):1465.
39. Hu WS, Li HC, Deng YJ, Sun X, Du Q, Plaza A. Lightweight tensor attention-driven ConvLSTM neural network for hyperspectral image classification. IEEE J Select Top Signal Process. 2021;15(3):734–45.
40. Mutegeki R, Han DS. A CNN-LSTM approach to human activity recognition. In: 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). 2020; 362–366. IEEE.
41. Yu S, Qin L. Human activity recognition with smartphone inertial sensors using bidir-lstm networks. In: 2018 3rd international conference on mechanical, control and computer engineering (ICMCCE) 2018; 219–224. IEEE.
42. Yen CT, Liao JX, Huang YK. Human daily activity recognition is performed using wearable inertial sensors combined with deep learning algorithms. IEEE Access. 2020;8:174105–14.
43. Kumar S, Gornale SS, Siddalingappa R, Mane A. Gender classification based on online signature features using machine learning techniques. Int J Intell Syst Appl Eng. 2022;10(2):260–8.
44. Yadav RK, Neogi SG, Semwal VB. A computational approach to identify normal and abnormal person gait using various machine learning and deep learning classifiers. In: Machine Learning, Image Processing, Network Security and Data Sciences: 4th International Conference, MIND 2022, Virtual Event, January 19–20, 2023, Proceedings, Part I. Cham: Springer Nature Switzerland, 2023; 14–26.