Karbala International Journal of Modern Science

Volume 9 | Issue 1 | Article 11

Data Integration Based Human Activity Recognition using Deep Learning Models

Basamma Umesh Patil
Research Scholar, Department of Computer Science and Engineering, JSS Academy of Technical Education (affiliated to VTU), Bengaluru, India., bupatil25@gmail.com

D V Ashoka
Professor and Dean, Department of Information Science and Engineering, JSS Academy of Technical Education (affiliated to VTU), Bengaluru, India., dr.dvashoka@gmail.com

Ajay Prakash B. V
Department of Research and Development, Pentagon Space, Bengaluru, India., ajayprakas@gmail.com

Recommended Citation
Patil, Basamma Umesh; Ashoka, D V; and V, Ajay Prakash B. (2023) "Data Integration Based Human Activity Recognition using Deep Learning Models," *Karbala International Journal of Modern Science*: Vol. 9 : Iss. 1 , Article 11.
Available at: https://doi.org/10.33640/2405-609X.3286

Data Integration Based Human Activity Recognition using Deep Learning Models

Abstract

Regular monitoring of physical activities such as walking, jogging, sitting, and standing will help reduce the risk of many diseases like cardiovascular complications, obesity, and diabetes. Recently, much research showed that the effective development of Human Activity Recognition (HAR) will help in monitoring the physical activities of people and aid in human healthcare. In this concern, deep learning models with a novel automated hyperparameter generator are proposed and implemented to predict human activities such as walking, jogging, walking upstairs, walking downstairs, sitting, and standing more precisely and robustly. Conventional HAR systems are unable to manage real-time changes in the surrounding infrastructure. Improved HAR approaches overcome this constraint by integrating multiple sensing modalities. These multiple sensors can produce accurate information, leading to a better perception of activity recognition. The proposed approach uses sensor-level fusion to integrate gyroscope and accelerometer sensors. The analysis is carried out using the widely accepted benchmark UCI-HAR dataset. Based on several performance evaluation experiments, the classification accuracy of long short-term memory (LSTM), convolutional neural network (CNN), and deep neural network (DNN) classifiers is reported to be 96%, 92%, and 93%, respectively. Compared to state-of-the-art deep learning models, the proposed method gives better results.

Keywords

Accelerometer, Deep Learning, Gyroscope, Human Activity Recognition, Sensor Level Fusion.

RESEARCH PAPER

Data Integration Based Human Activity Recognition using Deep Learning Models

Basamma Umesh Patil [a,b,*], D.V. Ashoka [c], B.V. Ajay Prakash [d]

[a] Research Scholar, Department of Computer Science and Engineering, JSS Academy of Technical Education (affiliated to VTU), Bengaluru, India
[b] Assistant Professor, Department of Computer Science and Engineering, SJB Institute of Technology (affiliated to VTU), Bengaluru, India
[c] Professor and Dean, Department of Information Science and Engineering, JSS Academy of Technical Education (affiliated to VTU), Bengaluru, India
[d] Department of Research and Development, Pentagon Space, Bengaluru, India

Abstract

Regular monitoring of physical activities such as walking, jogging, sitting, and standing will help reduce the risk of many diseases like cardiovascular complications, obesity, and diabetes. Recently, much research shows that the effective development of Human Activity Recognition (HAR) will help in monitoring the physical activities of people and aid in human healthcare. In this concern, deep learning models with a novel automated hyperparameter generator are proposed and implemented to predict human activities such as walking, jogging, walking upstairs, walking downstairs, sitting, and standing more precisely and robustly. Conventional HAR systems are unable to manage real-time changes in the surrounding infrastructure. Improved HAR approaches overcome this constraint by integrating multiple sensing modalities. These multiple sensors can produce accurate information, leading to a better perception of activity recognition. The proposed approach uses sensor-level fusion to integrate gyroscope and accelerometer sensors. The analysis is carried out using the widely accepted benchmark UCI-HAR dataset. Based on several performance evaluation experiments, the classification accuracy of long short-term memory (LSTM), convolutional neural network (CNN), and deep neural network (DNN) classifiers is reported to be 96%, 92%, and 93%, respectively. Compared to state-of-the-art deep learning models, the proposed method gives better results.

Keywords: Accelerometer, Deep learning, Gyroscope, Human activity recognition, Sensor level fusion

1. Introduction

Human activity recognition has become a very active research topic because of its widespread applications across various regions such as healthcare, smart homes, safety systems, transportation mode identification, sports, gaming, disability assistance, and human–computer interaction [1–3]. HAR can analyze military actions, identify vehicle driving activities, and help in fall detection and prevention [4]. It is employed in indoor and outdoor surveillance cameras to detect suspicious behavior and activities in airports, public transportation, and correctional facilities.

Prolonged sitting and lack of physical activity result in severe health issues like depression, obesity, diabetes, poor metabolism, and cardiovascular disorders. HAR can continuously monitor the everyday physical activities of subjects and analyze them. It can recognize human behavior by tracking various physical actions, namely, walking, running, sitting, sleeping, standing, driving, jogging, and abnormal activities based on the series of measurements captured by sensors [2,5,6].

Sensor-based and vision-based are two primary methods to gather human activity data [7]. Computer vision-based data can be collected using cameras, while, the sensor-based data acquisition

Received 10 August 2022; revised 25 December 2022; accepted 26 December 2022.
Available online 20 January 2023

* Corresponding author.
E-mail addresses: bupatil25@gmail.com (B.U. Patil), dr.dvashoka@gmail.com (D.V. Ashoka), ajayprakas@gmail.com (B.V. Ajay Prakash).

https://doi.org/10.33640/2405-609X.3286

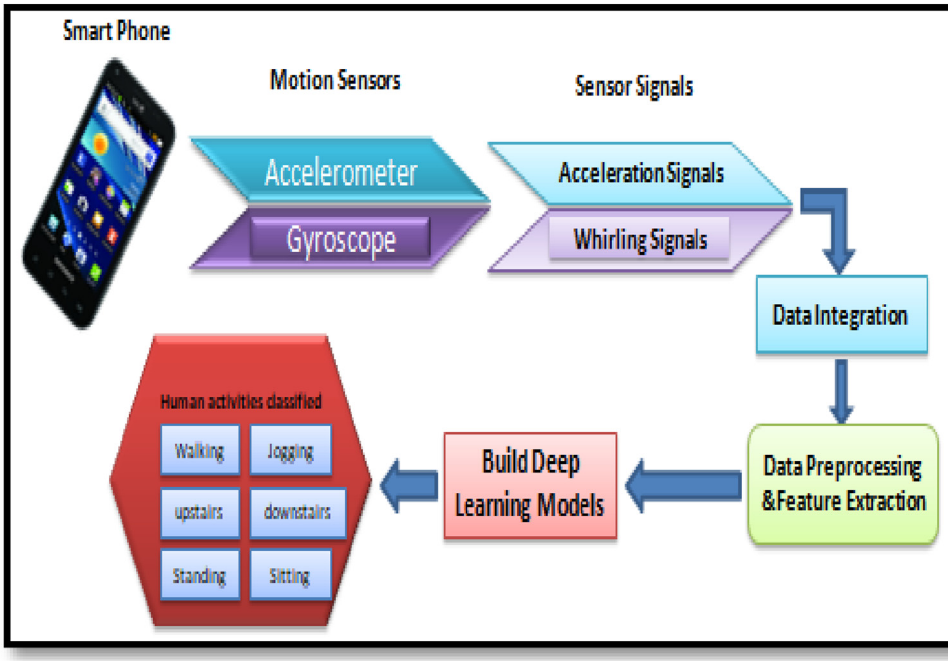

Fig. 1. Human activity recognition architecture.

approach is further classified into three groups based on how the sensor is deployed. These groups include wearable, object-tagged (device-bound), and dense sensing (device-free) methods [5,8]. Vision-based methods are more expensive, whereas, wearable sensor-based methods such as wristbands, wristwatches, smartphones, and smart glasses are more efficient and cheaper for capturing human activity data. Wearable sensor-based systems consist of built-in sensors such as an accelerometer, gyroscope, and GPS to continuously monitor real-time human activity. The accelerometer measures the linear acceleration of physical activity, whereas the gyroscope measures its angular velocity [1,9].

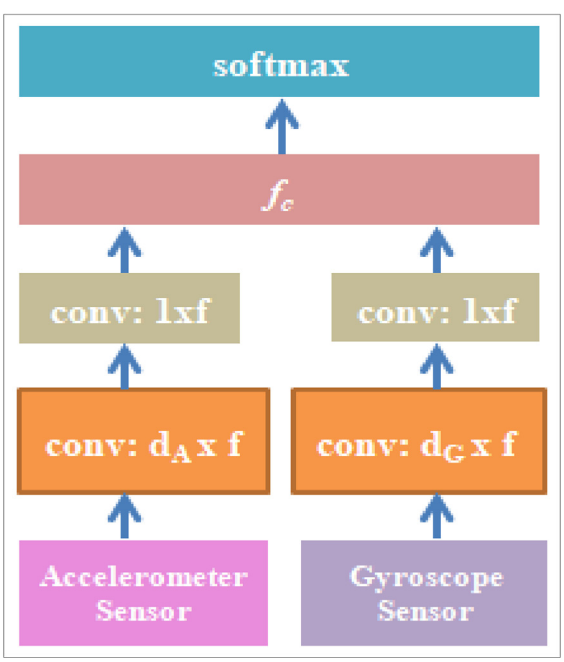

Fig. 2. Sensor level fusion using deep learning.

The combination of both accelerometer and gyroscope sensors often improves activity monitoring. According to recent studies, the multimodal data formed by integrating accelerometer and gyroscope signals increases HAR accuracy and is more reliable than using each signal alone [1,10]. The data integration technique helps in the assimilation of a variety of datasets from various domains to infer particular and precise knowledge [11]. Data integration of both sensors is paramount to measure multiple perspectives and get accurate object orientation, position, and velocity. This research approach collects standard HAR data (accelerometer and gyroscope sensors) from a publicly available benchmark UCI-HAR machine-learning repository [12].

1.1. Problem description and motivation

Presently, activity recognition plays a significant role in various healthcare services and applications.

Table 1. Comparison of DL models with training epoch of 50.

Optimizers	Model Name	Precision	Recall	F1-Score	Accuracy
Adam	LSTM	95	95	95	96
	DNN	89	91	90	91
	Single-headed CNN	88	89	88	88
	Multi-headed CNN	92	93	93	93
SGD	LSTM	78	72	73	82
	DNN	77	70	71	81
	Single-headed CNN	79	78	79	79
	Multi-headed CNN	91	91	91	91
Rmsprop	LSTM	95	95	95	97
	DNN	89	86	87	89
	Single-headed CNN	89	89	88	89
	Multi-headed CNN	91	91	91	91

For example, from daily life assistance promoting healthy behavior modifications to professional medical analysis such as fall detection and heart failure detection. As populations age, interpreting and tracking human activity with many sensor fusions will lead to better health outcomes and lower costs. The presence of individual criminals poses a significant danger to autonomous systems worldwide, making sensor fusion more crucial than ever. Eldercare and supported living, postural recognition, security and surveillance, athlete and first-responder status, and localization and navigation help for the disabled, are some examples of applications of wearable and ambient-sensor fusion for human activities.

Table 2. Comparison of DL models with training epoch of 100.

Optimizers	Model Name	Precision	Recall	F1-Score	Accuracy
Adam	LSTM	96	95	95	96
	DNN	91	94	92	93
	Single-headed CNN	92	92	91	91
	Multi-headed CNN	92	92	92	92
SGD	LSTM	82	76	78	85
	DNN	77	72	72	81
	Single-headed CNN	85	85	85	85
	Multi-headed CNN	91	92	91	91
Rmsprop	LSTM	95	96	95	96
	DNN	92	90	90	92
	Single-headed CNN	90	90	90	90
	Multi-headed CNN	93	93	93	93

1.2. Author's contribution

This activity recognition work integrates accelerometer and gyroscope signals using sensor-level fusion to generate accurate multimodal data. Sensor-level fusion combines raw signals from multiple sensors. Deep learning models with a novel automated hyperparameter generator are proposed and implemented to predict six human activities: walking, jogging, walking upstairs, walking downstairs, sitting, and standing. The work compares the effectiveness of deep learning models such as DNN, CNN, and LSTM. Machine learning models such as XGBoost, Logistic Regression, and SGD classifiers are also compared, and results are tabulated.

1.3. Organization of the paper

The remainder of this work is organized into different sections. Section II describes the literature review on HAR using data fusion techniques and different sensors; Section III details the proposed approach; Section IV explains performance evaluation; Section V explores experiments and results; and Section VI depicts the conclusion and future work.

2. Literature review

This section surveys existing methodologies for activity recognition using data integration techniques.

Challa et al. [13] identified HAR as an important area of research in human behavior analysis and healthcare services. Sensor-based and video-based systems are used for activity classification. A hybrid CNN − BiLSTM model is used for classification with an accuracy of 96%. Jain et al. [14] described HAR as a pattern recognition task. Its framework consists of four phases: data collection, data preprocessing, feature extraction, and activity classification. Y. Wang et al. [15] demonstrated that the combination of two or more sensors for activity recognition achieves the best accuracy compared to a single sensor alone. Ambient and wearable sensors, namely an accelerometer, gyroscope, barometer, magnetometer, temperature, and altitude, are integrated to predict the activities of older people. The data has been collected from 21 participants performing 17 different activities. However, the model was designed only for older people who live alone. Mitchell Webber and Raul Fernandez Rojas [1] compared the different data integration levels for multi-sensor activity details to find an optimal level

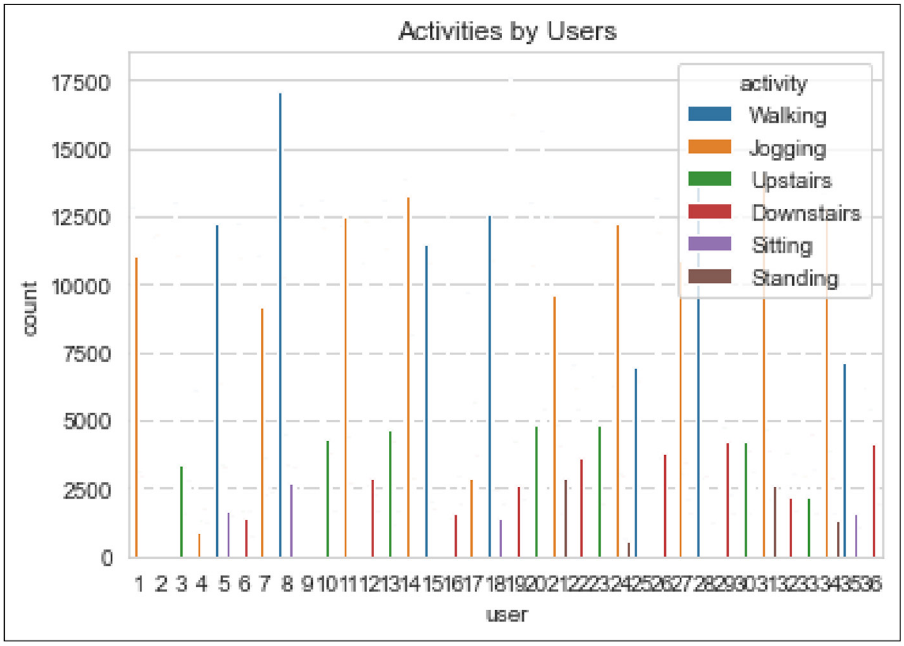

Fig. 3. Activities performed by 36 participants.

of data fusion. They fused gyroscope and accelerometer sensors by using a data processing pipeline at sensor fusion, feature fusion, and decision fusion steps. K-nearest neighbor, support vector machine, and decision tree classifier, are used to classify the public human activity dataset. The limitation of this work is the reduced efficiency while using HAR on mobile devices. M. Ehatisham-Ul-Haq et al. [8] used multimodal feature level integration to recognize the human activity. By combining publicly available multiple sensors of 27 different activity types, they achieved 97.6% accuracy. The data were classified using a support vector machine and K nearest neighbor. However, the method uses pre-segmented actions, which are not practical. Huynh-The et al. [16] employed a deep convolutional neural

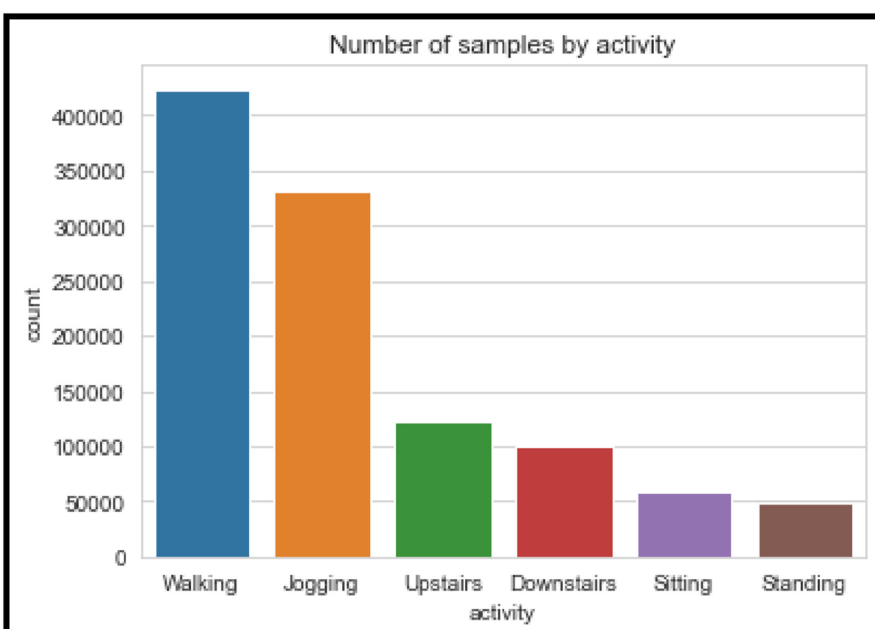

Fig. 4. Number of samples by activity.

network to derive multi-scale spatiotemporal signals and sensor-level correlations of an activity image with an accuracy of 96%.

X. Zhou et al. [2] developed a multisensory-based data fusion framework using context sensors, on-body sensors, and personal profile data. LSTM and deep Q-network are designed to improve the learning ability. N. S. Ghosh et al. [3] gathered six types of activity data from 30 volunteers ranging in age from 19 to 48 years. Random forest classifier, logistic regression, and logistic regression CV are used to classify the data. M. Muaaz et al. [4] presented a multimodal HAR system to classify physical actions with the use of Wi-Fi as well as wearable sensors. Feature-level fusion is employed. G. Ascioglu and Y. Senol [7] collected accelerometers and gyroscope sensors from 60 participants to recognize 13 different physical activities. The datasets were fed into convolutional neural networks, LSTM, and convolutional LSTM neural networks. Tianqi L. V. et al. [17] presented an activity recognition system using multimodal sensor data and the LSTM technique. A. Jain and V. Kanhangad [18] proposed an approach for activity classification by fusing accelerometers and gyroscope sensors. Data is fused using feature and score level fusion, with feature level fusion outperforming score level fusion. Support vector machine and k-nearest neighbor were employed on the public UCI-HAR data set. Sebastian Münzner et al. [19] showed that the prediction accuracy of CNN models is enhanced using sensor-specific normalization techniques. Henry Friday Nweke et al. [20] summarized various data fusion techniques and many classifier systems for physical activity recognition.

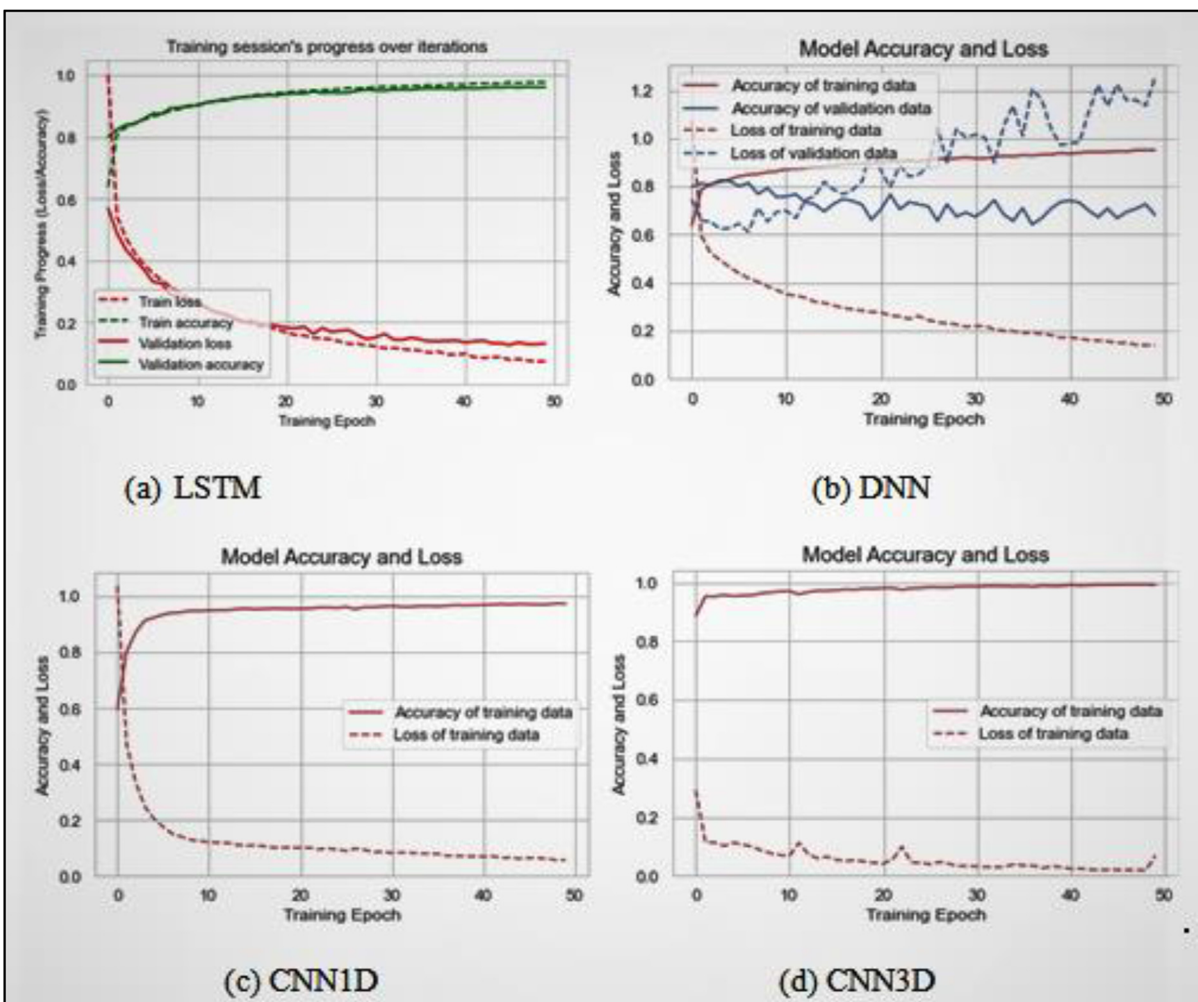

Fig. 5. Comparison of training and validation accuracy of different deep learning models (a) LSTM (b) DNN (c) CNN1D (d) CNN3D concerning adam optimizer for epoch 50.

Z. Chen et al. [21] collected 3D acceleration and gyroscope data to identify six human activities. Artificial neural networks, support vector machines, extreme learning machines, random forests, and LSTM were considered for activity recognition. Wearable sensor-based recognition systems can use deep learning networks such as convolutional neural networks to achieve better performance [22]. Long short-term memory networks are used to solve temporal dependency issues. Nowadays, their applications are widely used in activity recognition [23,24]. M. M. Hossain Shuvo et al. [25] employed a random forest classifier on accelerometer and gyroscope sensors to identify static and moving activities. S. K. Bashar et al. [26] gathered smartphone sensor recordings for activity classification. They employed neighborhood component analysis feature selection methods. Still, other hybrid feature selection methods need to be explored [27]. K. Chen et al. [28] presented a semi-supervised deep model to recognize imbalanced human activity using multimodal wearable sensor data. Gorji et al. [29] detected walking, sitting, and standing activities using a two-stage classifier. Still, more sophisticated human activities need to be detected. Semwal, V. B. et al. [30] defined gait analysis as the study of human locomotion Semwal, V. B. et al. [31] proposed gait-based person identification. SVM, ANN, and XGBoost algorithms are used to classify the data. P. Patil. et al. [32] employed ELM, KNN, SVM, and LMP algorithms for human gait classification. The performance of ELM was good, with an accuracy of 93.54%. Gupta A. et al. [33] applied ELM, SVM, KNN, and PCA algorithms for human gait classification. Semwal, V. B. et al. [34] applied human gait patterns to recognize human walking activities. A combination of deep and ensemble learning models is used for classification. Dua et al. [35,36] used CNN and GRU hybrid model for activity recognition. The model is validated using three public activity datasets viz. PAMAP2, WISDM, and UCI. Bijalwan et al. [37] proposed an activity recognition model to identify seven different activities. The dataset includes accelerometer, gyroscope,

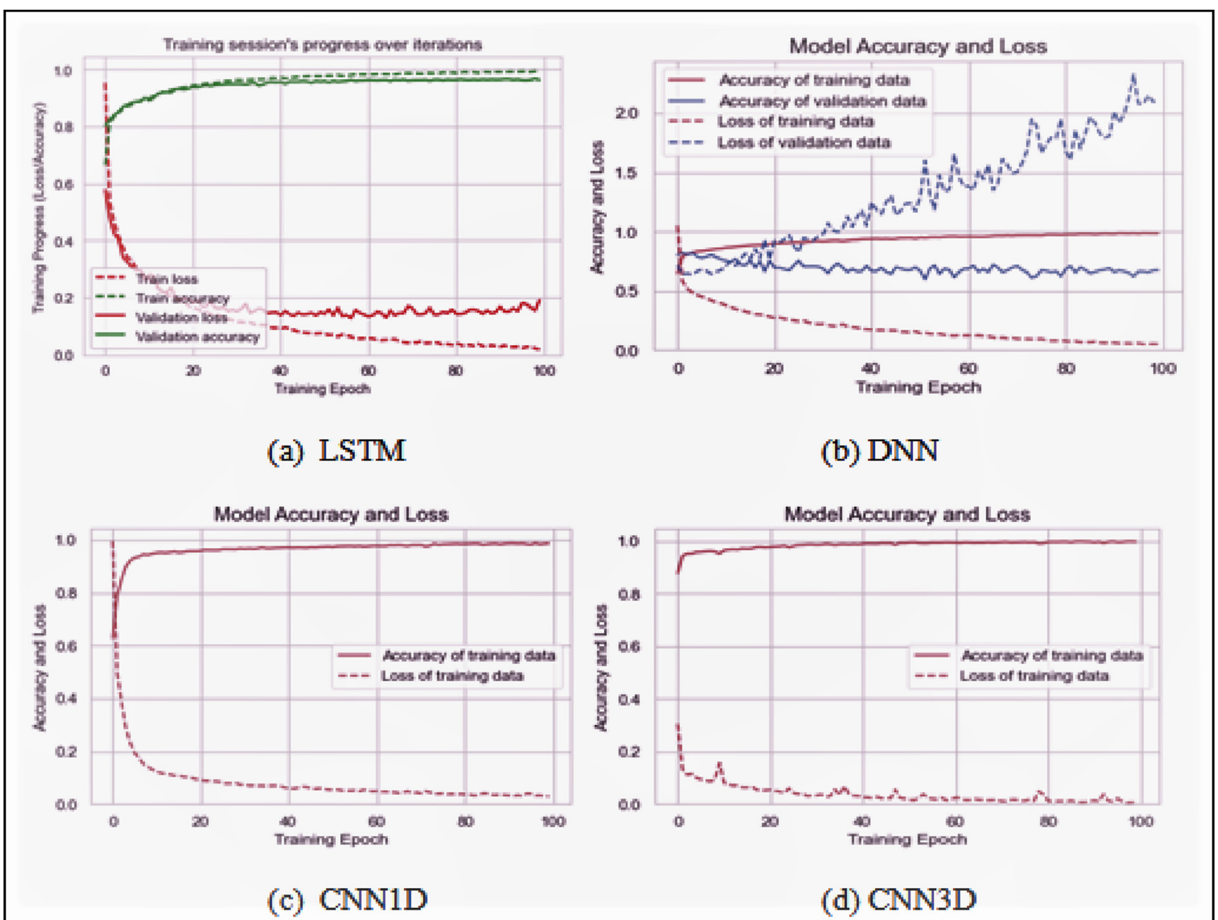

Fig. 6. Comparison of training and validation accuracy of different deep learning models (a) LSTM (b) DNN (c) CNN1D (d) CNN3D to adam optimizer for epoch 100.

and magnetometer sensors. Data are classified using deep learning models.

Due to various health concerns and tremendous applications, HAR is vital, and the research is underway. A few methods are proposed for various activity recognition applications using data fusion techniques. Some of the methods used a single dataset for activity recognition. These methods have limitations such as reduced efficiency and extended computational complexity. In this paper, we employ sensor-level fusion of accelerometer and gyroscope sensors for human activity recognition, which gives greater efficiency. Deep learning models such as DNN, CNN and LSTM are compared to identify the high accuracy model.

3. Proposed approach

The proposed method contains five primary stages, namely: collecting data from different sensor signals, data integration, data preprocessing and feature extraction, building deep learning models, and classifying human activities. Fig. 1 depicts the architecture of the proposed work.

3.1. Collecting data from different sensor signals

Collecting data from different sensor signals is the primary and paramount phase of gathering desired sensor data for human activity recognition. Accelerometer and gyroscope sensors are used to identify human activity. In the training phase, the data whose classification results are already known is used, and in the testing phase, the performances of the machine learning models are observed. The training phase HAR sensors data has been gathered from publicly available standard UCI - HAR machine learning repository [12]. This benchmark data considered a triaxial accelerometer within a smartphone to record linear acceleration along the X, Y, and Z axes, and a 3D gyroscope to measure angular velocity. Six physical activities, namely walking, sitting, standing, walking upstairs, walking downstairs, and jogging, are recorded

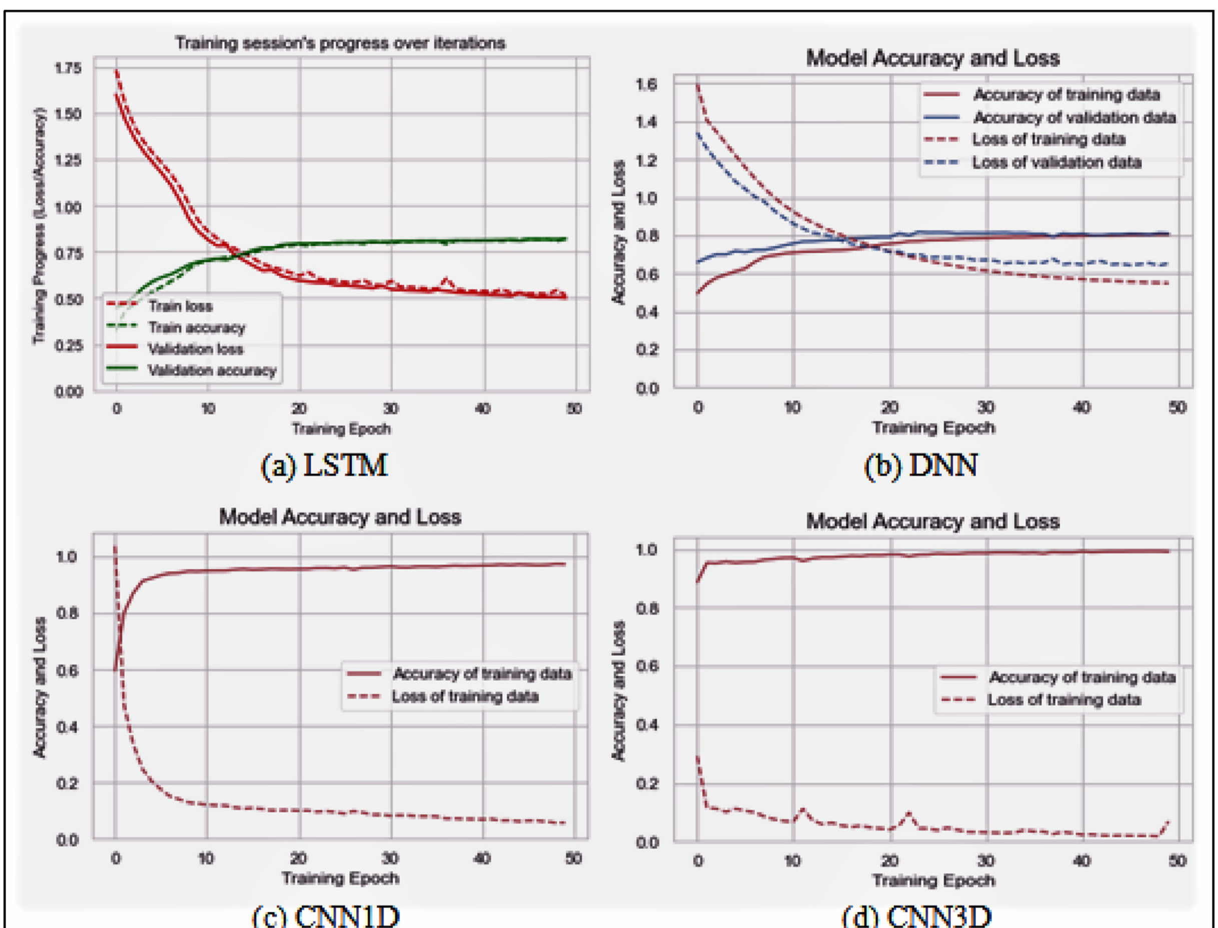

Fig. 7. Comparison of training and validation accuracy of different deep learning models (a) LSTM (b) DNN (c) CNN1D (d) CNN3D for SGD optimizer for epoch 50.

using 36 participants at a sampling rate of 50 Hz. Each participant wore a smartphone around the waist to perform these activities. The sensor data app installed on the smartphone gathers accelerometer and gyroscope data at a sample rate of 50 Hz, which is then preprocessed and used in the classification stage.

3.2. Data integration from different sources

Data from various sources can be merged in the data integration phase to produce more accurate and valuable information than a single source alone could provide. The individual dataset may not produce proper activity recognition results; hence, both accelerometer and gyroscope sensors are fused using sensor-level fusion [1]. Sensor fusion integrates data derived from different modalities to obtain accurate information, which wouldn't be possible when these data sets are used individually. Fusion is applied to the raw data to combine different data sets to produce enhanced data [38,39]. Sensor-level fusion has various advantages, such as robustness, more confidence, less ambiguity, and reduced uncertainty. In sensor fusion, the deep learning architecture integrates data at various stages. This method of fusion separates the data from different sensors and integrates it with a fully connected layer succeeding the convolutional layer [19]. This process allows an independent sensor-specific pipeline to meet the necessities of each method. Since the data integration is happening at the sensor level, it combines raw information that can account for inter- and intra-class and facilitates decision-making based on the fused raw information. Fig. 2 depicts sensor-level fusion using deep learning.

3.3. Data pre-processing

Smartphone and wearable sensor signals are often continuous time series. These sensors record activity data in a time series manner. The method of identifying human activity begins with the

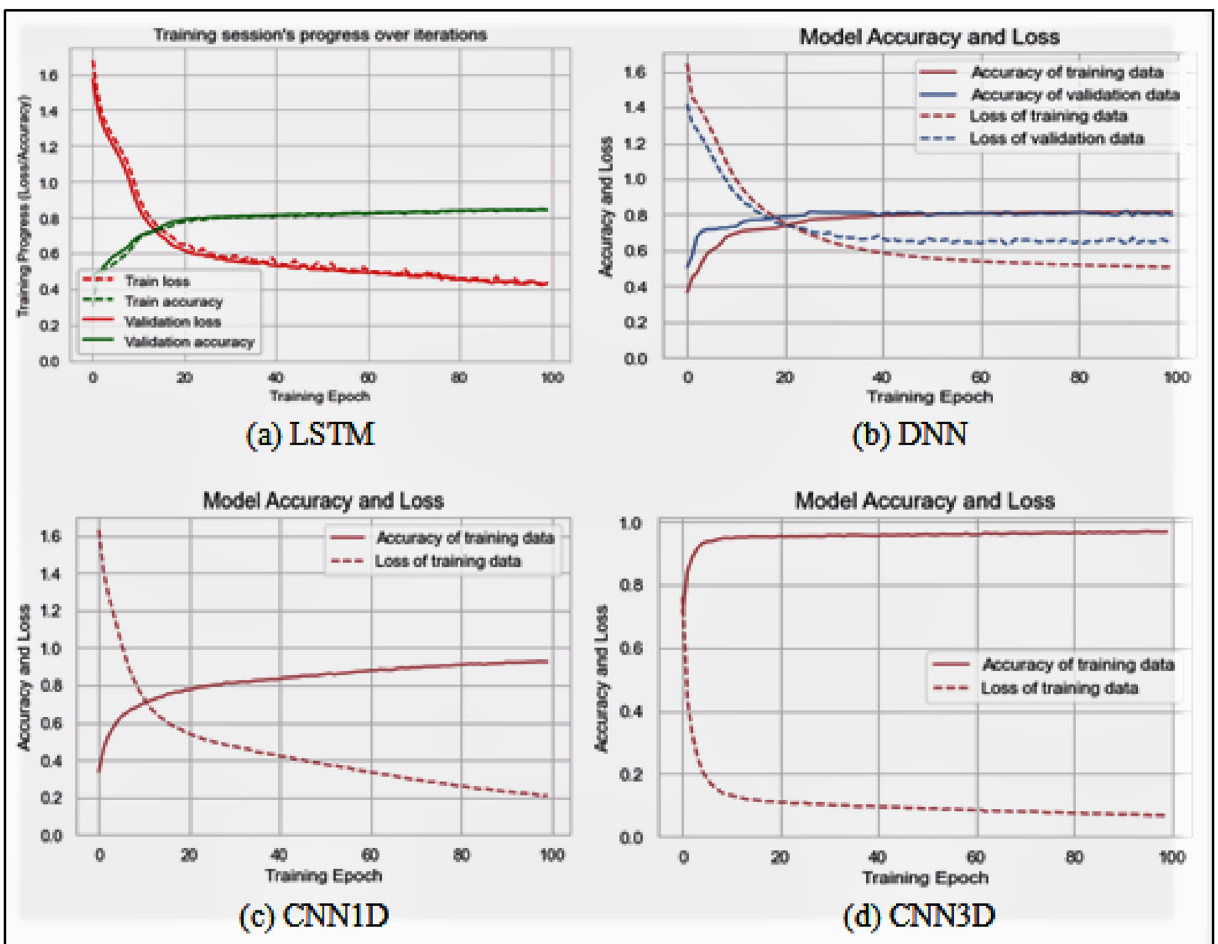

Fig. 8. Comparison of training and validation accuracy of different deep learning models (a) LSTM (b) DNN (c) CNN1D (d) CNN3D for SGD optimizer for epoch 100.

creation of segments from the sensor data. WISDM and UCI-HAR are the datasets utilized in the proposed work. The data preprocessing phase handles missing values in the dataset. This phase eliminates or filters out undesirable and noisy data to obtain an easier and more understandable model. In the beginning, all the null values from the datasets are dropped. Then accelerometer and gyroscope sensors are combined into a single data frame using sensor-level data fusion. Various features from accelerometer and gyroscope sensors are extracted and the dataset has been split into a train: test ratio of 80:20. The sensor data has been standardized to have a zero mean and unity standard deviation using the min–max normalization technique [40]. The main purpose of the min–max normalization method is to normalize the data. The min–max normalization is shown in equation (1).

$$y = x - \frac{average(x)}{max(x) - \min(x)} \tag{1}$$

where, x is the original value and y is the normalized value.

3.4. Building deep learning models and optimization of hyperparameters

Deep learning models such as CNN, LSTM, and DNN and machine learning models such as logistic regression, SGD, and XGBoost are used to classify the data. Deep learning models' hyperparameter values are optimized to get better results.

3.4.1. Convolutional Neural Network (CNN)

CNN is a type of feed-forward neural network consisting of 20–30 layers. It can be applied to extract data from the images by using various hidden layers [41]. The different layers in CNN are convolution, relu, pooling, and fully connected layers. The image is sent through a convolutional layer first and later passed through activation functions such as the sigmoid and relu. The formulas for the sigmoid and the relu functions are given in equations (2) and (3), respectively.

$$s(x) = 1/(1 + e^{-x}) \tag{2}$$

where, s(x) is the sigmoid function, which exists between 0 and 1, and e is Euler's number.

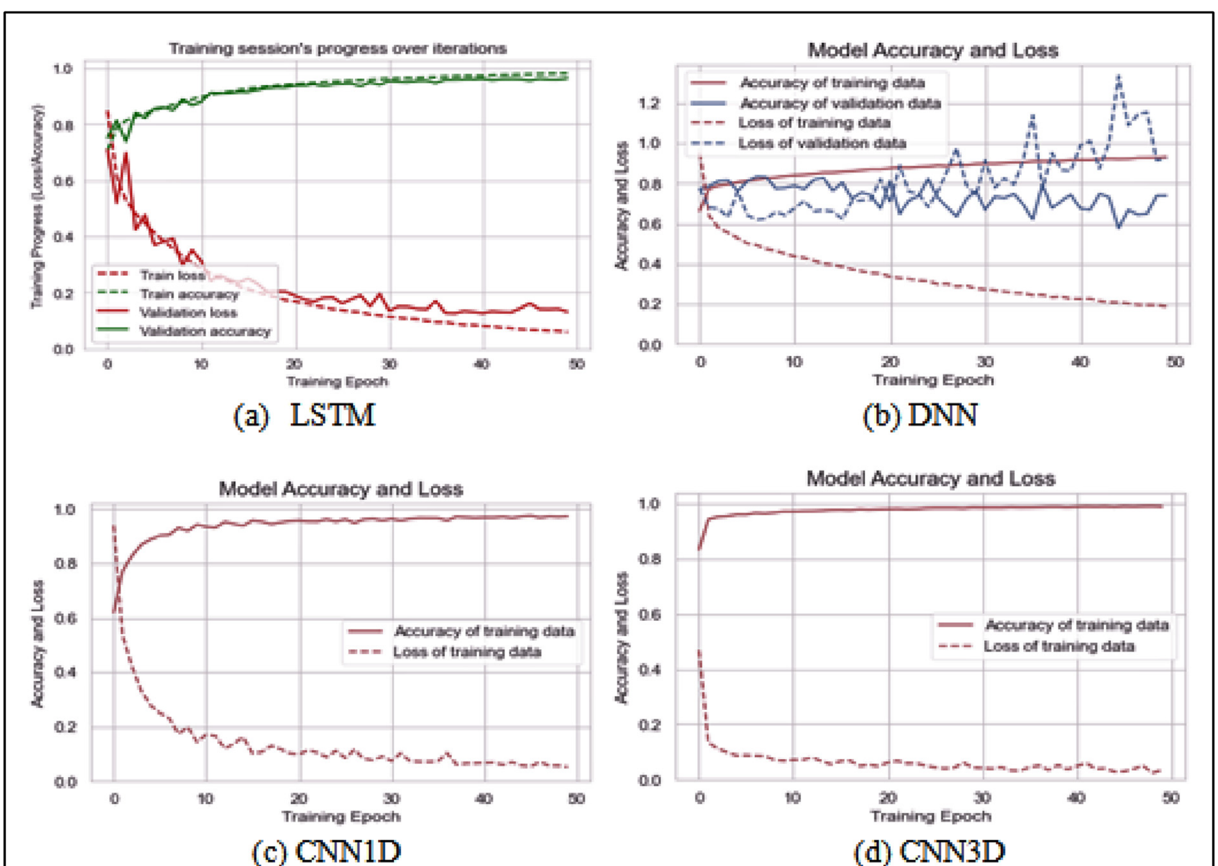

Fig. 9. Comparison of training and validation accuracy of different deep learning models (a) LSTM (b) DNN (c) CNN1D (d) CNN3D for RMSPROP optimizer for epoch 50.

$$f(x) = max(0, x) \quad (3)$$

where, x denotes the input. The function directly returns the input value or the value 0, if the input is 0 or less.

Single-headed CNN: It is a sequential model. It consists of two input layers and uses a convolutional 1D network with a relu activation function, one pooling layer, and two dense layers. One of the dense layers has a rectified linear unit activation function, and another has a softmax activation function for multi-label classification.

Multi-headed model: In the multi-headed model, every head uses a kernel of a different size to read input time steps. For instance, a three-headed CNN can have 3, 5, and 11 kernel sizes to read the information at three resolutions using the relu activation function. A fully linked layer translates and combines the three head interpretations in order to make a prediction.

3.4.2. *Long short-term memory (LSTM)*

The recurrent neural network has a special type called LSTM to determine long-duration dependencies. It recalls previous events and discovers patterns over time. LSTMs can be used for efficient machine interpretation, feedback connections, and language processing. LSTM consists of a cell to recollect random time interval data and three gates called input, output, and forget gates to control data flow to and from the cell.

3.4.3. *Deep neural network (DNN)*

DNN is an important type of artificial neural network. In between the input and output layers, different layers will be present. DNNs are typically

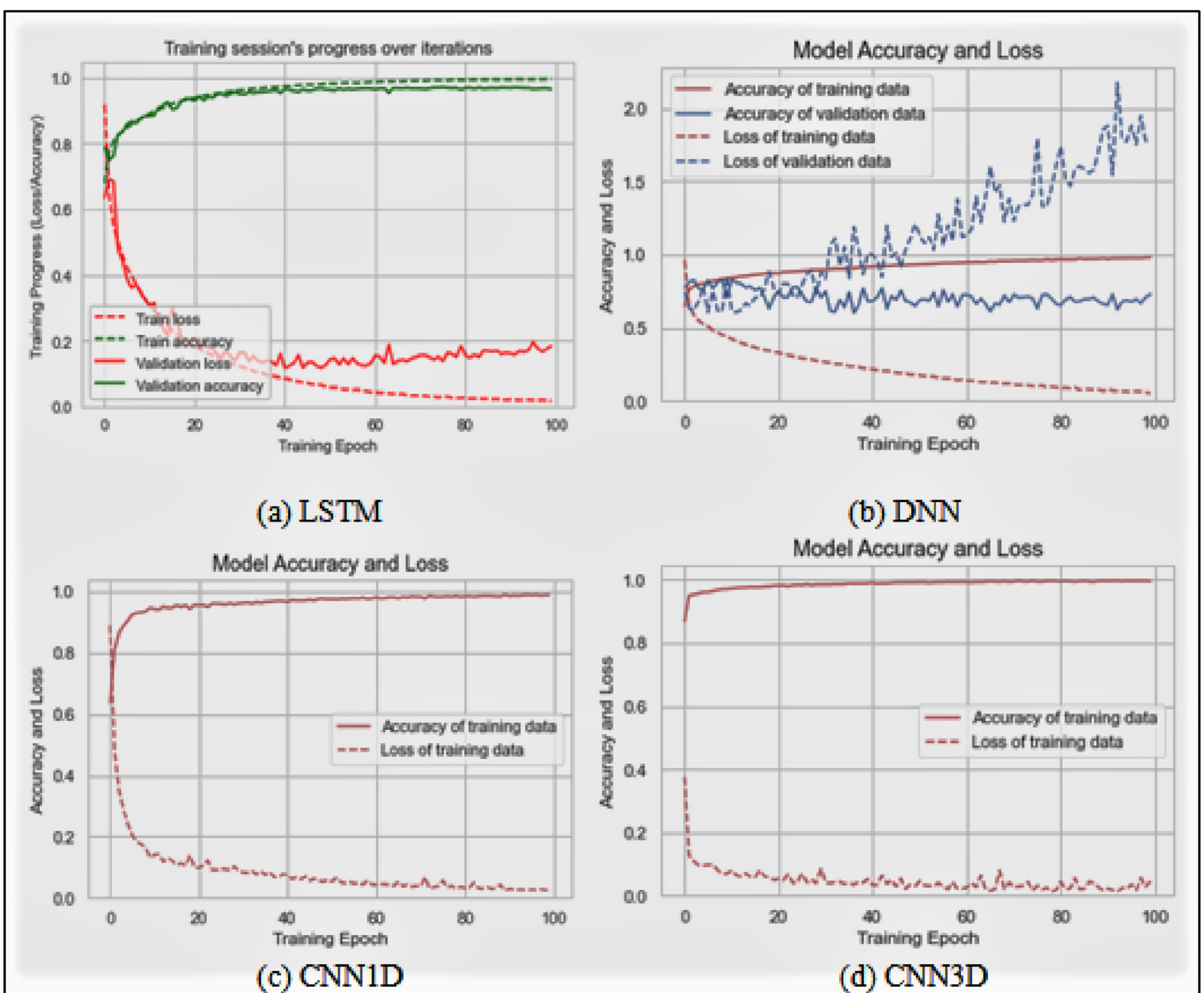

Fig. 10. Comparison of training and validation accuracy of different deep learning models (a) LSTM (b) DNN (c) CNN1D (d) CNN3D for RMSPROP optimizer for epoch 100.

feed-forward networks. The data flow direction is from the input to the output layer but not backward. A one-way link with a forward direction is present between various layers.

3.4.4. Logistic regression

Logistic regression is a type of supervised learning technique. It can predict the target variable's probability and calculate the binary (yes/no) probabilities of the events occurring.

3.4.5. Stochastic gradient descent classifier (SGD)

SGD classifier is a general neural network optimization algorithm that minimizes the cost function. It is used for the optimization of linear support vector machines and logistic regression. This algorithm can be used to optimize linear classifiers such as logistic regression and linear support vector machines. It uses a simple network to calculate the gradient.

3.4.6. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting is a very effective and powerful tree-based algorithm that applies the gradient boosting framework. It includes parameters for missing values, cross-validation, and regularization, as well as a scikit-learn compatible API, user-defined objective functions, and tree parameters. It makes use of ensemble principles and is a sequential technique. It provides better prediction accuracy by integrating a group of weak learners.

Algorithm for generating optimal hyperparameter values for Deep Learning (DL) techniques:

To obtain better accuracy, tunable parameters, namely the optimizer and epoch values, have been varied, and the results are tabulated. The proposed method is detailed in Algorithm1. This novel algorithm generates optimal hyperparameter values for various deep learning models. Evaluation metrics such as accuracy, precision, recall, and F1 score have been used to examine the reliability of human activity recognition models.

4. Performance evaluation and discussion

Various deep learning model evaluation metrics such as accuracy, precision, recall, and F1 score are used to examine the reliability of human activity

---

Algorithm 1: Generating Optimal Hyper Parameter Values for DL Techniques

---

Input: Multimodal data from accelerometer and gyroscope sensors
Output: Human Activity Recognition Model

1. An ← Accelerometer dataset, G ← Gyroscope dataset
2. Let A = {a1, a2, a3,… an} and G = {g1, g2, g3, … gn}.where ai and gi are independent variables and H ← HAR- the dependent variable that has to be predicted
3. Missing values elimination in A:
4. **for** i in n:
5. **do**
6. A={a1,a2,a3,..........an}
7. **if** ai==null **then**
8. Eliminate ai
9. **end if**
10. **end for**
11. Missing values elimination in B:
12. **for** i in n:
13. **do**
14. G={g1,g2,g3,..........gn}
15. **if** gi==null **then**
16. Eliminate gi
17. **end if**
18. **end for**
19. Perform sensor-level data fusion of A and G datasets using CNN:
20. **Input:** layer1 = A, layer2 = G
21. **Conv:** dense layer1 = A, dense layer2 = G
22. Fully connected Layer: A U G → D (Integrated Dataset)
23. #Generating Hyper Parameters Value for Optimization of DL Techniques
24. **HP←Hyper Parameters {HPCNN1D, HPCNN2D, HPLSTM, HPDNN}**
25. **HPCNN1D← {epoch, optimizer}**
26. **HPCNN2D← {epoch, optimizer}**
27. **HPLSTM← {epoch, optimizer}**
28. **HPDNN← {epoch, optimizer}**
29. **Optimizer←{adam, sgd, rmsprop}**
30. **for** split ratio ϵ{80:20} **do**
31. **for** each DLi ϵ {CNN1D, CNN2D, LSTM, DNN} HPi ϵ {HPCNN1D, HPCNN2D, HPLSTM, HPDNN}
32. Generate the HPi value and identify HAR
33. Measure Precision, Recall, F1-score, and Accuracy
34. **end for**
35. **end for**

---

recognition models. Accuracy is calculated as the percentage of correctly classified points out of all the test set points. In order to evaluate the classification model and have a general understanding of it, accuracy is used.

Precision: It is the ratio of true positives and the sum of true positives and false positives as depicted in equation (4)

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

where, TP denotes true positives, and FP denotes false positives.

Recall: It is defined as the ratio of true positives and the sum of true positives and false negatives as depicted in equation (5)

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

where, TP denotes true positives, and FN stands for false negatives.

F1-Score: F1-Score is the harmonic mean of precision and recall that provides the combined result of both precision and recall together and its formula is given in equation (6)

$$F1 - measure = 2*\frac{Precision*Recall}{Precision + Recall} \tag{6}$$

The outcomes of all the classifier models in terms of precision, recall, and F1-score are shown in Tables 1 and 2. It is evident that LSTM outperforms other classifiers.

5. Experiments and results

Implementation is carried out using the Python language, and experiments are performed on deep learning models. These experiments are conducted in a Jupyter notebook using the standard UCI-HAR dataset. The dataset is analyzed using exploratory data analysis.

Fig. 3 depicts the graph drawn to show the six activities done by 36 participants, ranging in age from 23 to 58 years. The X-axis represents the total number of participants, and the Y-axis represents the activity sample count for each individual participant.

Fig. 4 depicts the number of samples for individual activities such as walking, jogging, walking upstairs, walking downstairs, sitting, and standing for overall participants.

As shown in Fig. 5(a), as in the case of the LSTM model, when the training epoch is set to 50, there is a decrease in training loss and validation loss, enhancing the accuracy of the model. The graph in Fig. 5(b) displays the loss and accuracy of the DNN model. Here we can observe that the validation loss is increasing as a result of declining validation accuracy. It is evident that the training loss is declining, which results in an improvement in training accuracy. Fig. 5(c) and (d) and show CNN1D and CNN3D graphs. There is a steep increase in training accuracy and a sharp fall in training loss here until training epoch 30, after which there is a slight variance in either metric. From this, we can observe that the LSTM model outperforms other models, considering the adam optimizer with an accuracy of 96%. The adam optimizer uses weighted momentum, faster convergence, and individual learning rates for greater efficiency.

As shown in Fig. 6(a), for the LSTM model, when the training epoch was set to 100, there is a decrease in training loss and validation loss, resulting in an increase in model accuracy up to epoch 50. Following that, it can be shown that both training accuracy and validation accuracy remain unchanged. The graph in Fig. 6(b) displays the accuracy and loss of the DNN model. Even if the data validation loss is growing in this case, the accuracy of the validation after epoch 20 remains unchanged. After epoch 20 iterations, the training accuracy is steady and the training loss is dropping. The CNN1D and CNN3D graphs are shown in Fig. 6(c) and (d). It can be seen that at training epoch 30, there is a steep decrease in training loss and an increase in training accuracy, followed by a period of slight variation in training loss and accuracy. It leads to the conclusion that the LSTM model outperforms other models with an accuracy of 96%.

For the SGD optimizer and training epoch value of 50, Fig. 7 compares the training accuracy and validation accuracy of LSTM, DNN, CNN1D, and CNN3D, respectively. Rather than selecting from the entire dataset, the SGD optimizer randomly chooses data in batches and in a generalized form. As a result, it can be shown that training and validation accuracy have not improved much even while the validation loss and training loss are declining. The multi-headed CNN outperformed the other models with 91% accuracy.

For the SGD optimizer and training epoch value of 100, Fig. 8 compares the training accuracy and validation accuracy of LSTM, DNN, CNN1D, and CNN3D, respectively. SGD optimizer randomly selects the data in batches and in a generalized form rather than choosing the entire dataset. As a result, it is noted that training and validation accuracy have not significantly improved despite a decrease in validation loss and training loss. Comparatively speaking, the LSTM model outperforms CNN1D and DNN. CNN3D outperforms the LSTM model with an accuracy rate of 91% at epoch 100.

For the RMSPROP optimizer and training epoch values of 50 and 100, Figs. 9 and 10 compare the training accuracy and validation accuracy of LSTM, DNN, CNN1D, and CNN3D. RMSPROP is an extension of the RPPROP optimizer. Large datasets are ideally suited for RMSPROP, which also speeds up the optimization process by reducing the number of function evaluations. Human activity prediction is challenging due to variations in the attributes of sensors such as accelerometers and gyroscopes. The RMSPROP adapts to the changes in the parameters very quickly and requires less tuning. The graph depicts that LSTM outperforms DNN, CNN1D, and CNN3D with greater training accuracy.

For the training epoch value of 50, Table 1 compares deep learning models like LSTM, DNN, single-headed CNN, and multi-headed CNN with optimizers like ADAM, SGD, and RMSPROP. By varying the optimizers and maintaining a constant epoch value of 50 and a training test ratio of 80:20, various performance metrics are evaluated and tabulated for the deep learning models built. This clearly shows that the LSTM model performs better than other models. Since LSTM is based on recurrent networks, the data is divided into small batches and given to a recurring model of the layers that allows them to retain the information. It is also capable of learning long-term dependencies in data.

Table 3. Comparison of DL models for individual activity with epoch 50 and ADAM optimizer.

Optimizers	Human Activity	Model Name	Precision	Recall	F1-Score
Adam	**Walking**	LSTM	0.9	0.84	0.86
		DNN	0.72	0.75	0.73
		CNN1D	0.96	0.93	0.94
		CNN3D	0.98	0.95	0.97
	Jogging	LSTM	0.99	0.98	0.99
		DNN	0.98	0.97	0.98
		CNN1D	0.87	0.88	0.87
		CNN3D	0.93	0.95	0.97
	Upstairs	LSTM	1	0.99	0.99
		DNN	1	1	1
		CNN1D	0.85	1	0.91
		CNN3D	0.9	0.99	0.94
	Downstairs	LSTM	0.98	0.99	0.99
		DNN	0.99	1	1
		CNN1D	0.75	0.85	0.8
		CNN3D	0.84	0.82	0.83
	Sitting	LSTM	0.86	0.92	0.89
		DNN	0.68	0.83	0.75
		CNN1D	0.89	0.71	0.79
		CNN3D	0.89	0.86	0.87
	Standing	LSTM	0.98	98	0.98
		DNN	0.95	0.89	0.92
		CNN1D	1	0.95	0.97
		CNN3D	1	1	1

Table 4. Comparison of DL models for individual activity with training epoch of 50 and SGD optimizer.

Optimizers	Human Activity	Model Name	Precision	Recall	F1-Score
SGD	**Walking**	LSTM	0.55	0.19	0.28
		DNN	0.55	0.12	0.2
		CNN1D	0.64	0.77	0.7
		CNN3D	0.99	0.98	0.98
	Jogging	LSTM	0.96	0.98	0.97
		DNN	0.94	0.97	0.96
		CNN1D	0.65	0.67	0.66
		CNN3D	0.86	0.93	0.9
	Upstairs	LSTM	0.99	0.84	0.91
		DNN	0.97	0.91	0.94
		CNN1D	0.78	0.7	0.74
		CNN3D	0.93	0.98	0.96
	Downstairs	LSTM	0.86	0.99	0.92
		DNN	0.95	0.98	0.97
		CNN1D	0.87	0.73	0.8
		CNN3D	0.89	0.72	0.8
	Sitting	LSTM	0.55	0.33	0.41
		DNN	0.46	0.29	0.36
		CNN1D	0.81	0.87	0.84
		CNN3D	0.81	0.9	0.85
	Standing	LSTM	0.78	0.97	0.86
		DNN	0.75	0.95	0.84
		CNN1D	1	0.95	0.97
		CNN3D	1	0.95	0.97

Table 5. Comparison of DL models for individual activity with training epoch of 50 and RMSPROP optimizer.

Optimizers	Human Activity	Model Name	Precision	Recall	F1-Score
Rmsprop	**Walking**	LSTM	0.91	0.87	0.89
		DNN	0.83	0.48	0.61
		CNN1D	0.98	0.9	0.94
		CNN3D	0.91	0.97	0.94
	Jogging	LSTM	0.99	0.99	0.99
		DNN	0.98	0.98	0.98
		CNN1D	0.86	0.92	0.89
		CNN3D	0.9	0.88	0.89
	Upstairs	LSTM	1	0.99	0.99
		DNN	1	1	1
		CNN1D	0.84	0.99	0.91
		CNN3D	0.94	0.97	0.95
	Downstairs	LSTM	0.98	0.99	0.98
		DNN	0.99	1	0.99
		CNN1D	0.88	0.67	0.76
		CNN3D	0.84	0.88	0.86
	Sitting	LSTM	0.87	0.91	0.89
		DNN	0.65	0.77	0.7
		CNN1D	0.78	0.9	0.83
		CNN3D	0.92	0.84	0.88
	Standing	LSTM	0.98	0.98	0.98
		DNN	0.89	0.93	0.91
		CNN1D	1	0.95	0.97
		CNN3D	0.99	0.95	0.97

Table 6. Comparing results of different ML models.

Model Name	Precision	Recall	F1-Score	Accuracy
Logistic Regression	0.37	0.38	0.35	57%
XGBoost	0.88	0.77	0.80	86%
SGDClassifier	0.37	0.35	0.34	56%

Table 7. Comparing state-of-the-art methods.

State of the art approach	sSensor	Classified activities	. Total no. of subjects	Accuracy (%)
Patil et al. [32]	Accelerometer	Normal, crunch1,-crunch2,MS, stroke	5	93.54
Linag et al. [42]	Accelerometer	Standing still, lying, sitting, motorized transportation, walking, running, bicycling, jumping	24	85
Zeng et al. [43]	GPS	,walking,bicycling motorized transportation	65	76
Martin et al. [44]	Accelerometer	g,Slowly walkin normally walking, fast walking, running, sitting standing still	16	88
Kau and chen [28]	Accelerometer	Falling down	9	92

Table 2 compares deep learning models such as LSTM, DNN, single-headed CNN, and multi-headed CNN with optimizers like ADAM, SGD, and RMSPROP for the training epoch value of 100. The table shows the evaluation of the models by varying the optimizers and keeping a constant epoch value of 100 and a training test ratio of 80:20. Performance metrics such as precision, recall, f1-score, and accuracy are measured and tabulated. For both ADAM and RMSPROP, the LSTM model gives 96% accuracy because both optimizers converge fast and require less tuning of parameters.

Table 3 compares precision, recall, and F1-score values for individual human activities such as walking, jogging, walking upstairs, walking downstairs, sitting, and standing with different optimizers to understand how DL models perform each activity.

Tables 3–5 depict the comparison of precision, recall, and F1-score values with a train-to-test ratio of 80:20 for each human activity such as walking, jogging, walking upstairs, walking downstairs, sitting, and standing with different optimizers to understand how DL models perform for each different activity.

Table 6 shows the different ML models such as logistic regression, XGBoost, and SGDClassifier results compared using precision, recall, F1-score, and accuracy with a train-to-test ratio of 80:20.

From the results, it is observed that our proposed algorithm with optimal hyperparameter values gives better performance than the state-of-the-art models. The model performs well on the UCI-HAR benchmark dataset. The performance of the current study is compared to the results of state-of-the-art approaches. Table 7 provides an overview of recent investigations.

6. Conclusion and future work

This method recognizes human activity by integrating gyroscope and accelerometer sensors using sensor-level fusion. Six different human activities such as walking, jogging, walking upstairs, walking downstairs, sitting, and standing are classified using deep learning models, i.e. CNN, DNN, and LSTM. Machine learning models such as logistic regression, XGBoost, and SGD classifier are also compared. Optimal hyperparameter values for different DL models have been investigated, and the performances of these models are compared. LSTM model has given good results for recognizing human activity compared to other techniques. The accuracy of CNN, DNN, and LSTM is 92%, 93%, and 96%, respectively.

Our future work will explore various ways to identify more relevant sensors for activity recognition. We plan to incorporate different data fusion techniques to further reduce computational complexity. The set of activities can be extended by including complex physical activities and a collection of better features. To improve recognition accuracy, the combination of sensors and wearable cameras for deep learning applications can be analyzed individually.

Acknowledgments

This research was supported by Visvesvaraya Technological University, Jnana Sangama, Belagavi.

References

[1] M. Webber, R.F. Rojas, Human activity recognition with accelerometer and gyroscope: a data fusion approach, in: IEEE Sensor J 21, 2021, pp. 16979−16989.

[2] X. Zhou, W. Liang, K.I.-K. Wang, H. Wang, L.T. Yang, Q. Jin, Deep-learning-enhanced human activity recognition for internet of healthcare things, in: IEEE Internet Things J 7, 2020, pp. 6429−6438.

[3] N.S. Ghosh, R. Majumdar, B. Giri, A. Ghosh, Detection of Human Activity by Widget, 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO. 2020, pp. 1330−1334.

[4] M. Muaaz, A. Chelli, A.A. Abdelgawwad, A.C. Mallofre, M. Patzold, WiWeHAR: multimodal human activity recognition using wi-fi and wearable sensing modalities, in: IEEE Access vol. 8, 2020, pp. 164453−164470.

[5] V. Bijalwan, V.B. Semwal, Wearable sensor-based pattern mining for human activity recognition: deep learning approach, Ind. Robot 48 (2021) 21−33.

[6] V.B. Semwal, P. Lalwani, M.K. Mishra, V. Bijalwan, J.S. Chada, An optimized feature selection using bio-geography optimization technique for human walking activities recognition, Computing 12 (2021) 2893−2914.

[7] G. Ascioglu, Y. Senol, Design of a wearable wireless multi-sensor monitoring system and application for activity recognition using deep learning, in: IEEE Access 8, 2020, pp. 169183−169195.

[8] M. Ehatisham-Ul-Haq, Robust human activity recognition using multimodal feature-level fusion, in: IEEE 7, 2019, pp. 60736−60751.

[9] R. Jain, V.B. Semwal, P. Kaushik, Stride segmentation of inertial sensor data using statistical methods for different walking activities, Robotica 40 (2022) 2567−2580.

[10] U. Alrazzak, B. Alhalabi, A Survey on Human Activity Recognition Using Accelerometer Sensor, Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). 2019, pp. 152−159.

[11] Y. Zheng, Methodologies for cross-domain data fusion: an overview, in: IEEE Transactions on Big Data 1, 2015, pp. 16−34.

[12] Dr Gary Weiss, Computer and Information Sciences Department, Fordham University. 2019 (accessed March 25, 2022), https://archive.ics.uci.edu/ml/datasets/WISDM+Smartphone+and+Smartwatch+Activity+and+Biometrics+Dataset+.

[13] S.K. Challa, A. Kumar, V.B. Semwal, A multi-branch CNN-BiLSTM model for human activity recognition using wearable sensor data, Vis. Comput. 37 (2021) 1−15.

[14] R. Jain, V.B. Semwal, P. Kaushik, Deep ensemble learning approach for lower extremity activity recognition using wearable sensors, Expet. Syst. 39 (2022) e12743.

[15] Y. Wang, S. Cang, H. Yu, A data fusion-based hybrid sensory system for older people's daily activity and daily routine recognition, in: IEEE Sensor J 18, 2018, pp. 6874−6888.

[16] T. Huynh-The, C.-H. Hua, N.A. Tu, D.-S. Kim, Physical activity recognition with statistical-deep fusion model using multiple sensory data for smart health, in: IEEE Internet Things J 8, 2021, pp. 1533−1543.

[17] T. Lv, X. Wang, L. Jin, Y. Xiao, M. Song, A hybrid network based on dense connection and weighted feature aggregation for human activity recognition, in: IEEE Access vol. 8, 2020, pp. 68320−68332.

[18] A. Jain, V. Kanhangad, Human activity classification in smartphones using accelerometer and gyroscope sensors, in: IEEE Sensors Journal vol. 18, 2018, pp. 1169−1177.

[19] S. Munzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, R. Durichen, CNN-Based Sensor Fusion Techniques for Multimodal Human Activity Recognition in Proceedings of the 2017 ACM International Symposium on Wearable Computers (ISWC '17), Association for Computing Machinery, New York, NY, USA. 2017, pp. 158−165.

[20] H.F. Nweke, Y.W. Teh, G. Mujtaba, M.A. Al-garadi, Data fusion, and multiple classifier systems for human activity detection and health monitoring: review and open research directions, Inf. Fusion 46 (2019) 147−170.

[21] Z. Chen, C. Jiang, S. Xiang, J. Ding, M. Wu, X. Li, Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori, in: IEEE Trans Instrum Meas 69, 2020, pp. 3992−4001.

[22] S.-M. Lee, S.M. Yoon, H. Cho, Human activity recognition from accelerometer data using Convolutional Neural Network, IEEE Int. Conf. Big Data Smart Comp. (BigComp) (2017) 131−134.

[23] N. Tufek, O. Ozkaya, A Comparative Research on Human Activity Recognition Using Deep Learning, 27th Signal Processing and Communications Applications Conference, SIU. 2019, pp. 1−4.

[24] M. Devanne, P. Papadakis, S.M. Nguyen, Recognition of activities of daily living via hierarchical long-short term memory networks, IEEE Int. Conf. Syst. Man. Cybern. (2019) 3318−3324.

[25] M.M. Hossain Shuvo, N. Ahmed, K. Nouduri, K. Palaniappan, A hybrid approach for human activity recognition with support vector machine and 1D convolutional neural network, IEEE Appl. Imag. Pattern Recogn. Workshop (AIPR) (2020) 1−5.

[26] S.K. Bashar, A. Al Fahim, K.H. Chon, Smartphone-based human activity recognition with feature selection and dense neural network, in: 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC. 2020, pp. 5888−5891.

[27] Y. Tian, J. Zhang, L. Li, Z. Liu, A novel sensor-based human activity recognition method based on hybrid feature selection and combinational optimization, in: IEEE Access 9, 2021, pp. 107235−107249.

[28] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, F. Nie, A semisupervised recurrent convolutional attention model for human activity recognition, in: IEEE Transact Neural Networks Learn Syst 31, 2020, pp. 1747−1756.

[29] A. Gorji, H.-U.-R. Khalid, A. Bourdoux, H. Sahli, On the generalization and reliability of single radar-based human activity recognition, in: IEEE Access 9, 2021, pp. 85334−85349.

[30] V.B. Semwal, N. Gaud, P. Lalwani, V. Bijalwan, A.K. Alok, Pattern identification of different human joints for different human walking styles using inertial measurement unit (IMU) sensor, Artif. Intell. Rev. 55 (2022) 1149−1169.

[31] V.B. Semwal, A. Mazumdar, A. Jha, N. Gaud, V. Bijalwan, Speed, cloth and pose invariant gait recognition-based person identification, in: M. Pandey, S.S. Rautaray, eds., Machine Learning: Theoretical Foundations and Practical Applications. Studies in Big Data vol. 87, Springer, Singapore. 2021, pp. 1−18.

[32] P. Patil, K.S. Kumar, N. Gaud, V.B. Semwal, Clinical Human Gait Classification: Extreme Learning Machine Approach, 1st International Conference on Advances in Science, Engineering and Robotics Technology, ICASERT. 2019, pp. 1−6.

[33] A. Gupta, V.B. Semwal, Multiple task human gait analysis and identification: ensemble learning approach, in: S.N. Mohanty, eds., Emotion and Information Processing, Springer, Cham. 2020, pp. 172−185. E-Publishing Inc., Switzerland.

[34] V.B. Semwal, A. Gupta, P. Lalwani, An optimized hybrid deep learning model using an ensemble learning approach for human walking activity recognition, J. Supercomput. 77 (2021) 12256−12279.

[35] N. Dua, S. Singh, V.B. Semwal, S.K. Challa, Inception-inspired CNN-GRU hybrid network for human activity recognition, Multimed. Tool Appl. 67 (2022) 1−35.

[36] N. Dua, S.N. Singh, V.B. Semwal, Multi-input CNN-GRU based human activity recognition using wearable sensors, Computing 103 (2021) 1461−1478.

[37] V. Bijalwan, V.B. Semwal, Wearable sensor-based pattern mining for human activity recognition: deep learning approach, Ind. Robot. 49 (2021) 21−33.

[38] B.U. Patil, D.V. Ashoka, B.V. Ajay Prakash, Optimization of hyper parameters in machine learning techniques for air

quality predictive analysis, in: International Journal on Information Technologies & Security 13, 2021, pp. 73−86.
[39] R. Chetan, D.V. Ashoka, B.V. Ajay Prakash, IMLAPC: interfused machine learning approach for prediction of crops, Rev. Intel. Artif. 36 (2022) 169−174.
[40] L. Munkhdalai, T. Munkhdalai, K.H. Park, H.G. Lee, M. Li, K.H. Ryu, Mixture of activation functions with extended min-max normalization for forex market prediction, in: IEEE Access 7, 2019, pp. 183680−183691.
[41] Y.L. Chaitra, R. Dinesh, M.T. Gopalakrishna, B.V. Ajay Prakash, Deep-CNNTL: text localization from natural scene images using deep convolution neural network with transfer learning, Arabian J. Sci. Eng. 47 (2021) 9629−9640.
[42] Y. Liang, X. Zhou, Z. Yu, B. Guo, Energy-efficient motion related activity recognition on mobile devices for pervasive healthcare, Mob. Netw. Appl. 19 (2013) 303−317.
[43] Y. Zheng, Y. Chen, Q. Li, X. Xie, W.-Y. Ma, Understanding transportation modes based on GPS data for web applications, ACM Trans. Web 4 (2010) 1−36.
[44] H. Martin, A.M. Bernardos, J. Iglesias, J.R. Casar, Activity logging using lightweight classification techniques in mobile devices, Person. Ubiquit. Comput. 17 (2012) 675−695.