Multimedia Tools and Applications (2024) 83:89635–89664
https://doi.org/10.1007/s11042-024-18935-0

Ensemble of deep learning techniques to human activity recognition using smart phone signals

Soodabeh Imanzadeh[1] · **Jafar Tanha**[1] · **Mahdi Jalili**[2]

Received: 6 August 2023 / Revised: 12 February 2024 / Accepted: 13 March 2024 /
Published online: 1 April 2024

Abstract

Human Activity Recognition (HAR) has become a significant area of study in the fields of health, human behavior analysis, the Internet of Things, and human–machine interaction in recent years. Smartphones are a popular choice for HAR as they are common devices used in daily life. However, most available HAR datasets are gathered in laboratory settings, which do not reflect real-world scenarios. To address this issue, a real-world dataset using smartphone inertial sensors, involving 62 individuals, is collected. The collected dataset is noisy, small, and has variable frequency. On the other hand, in the context of HAR, algorithms face additional challenges due to intra-class diversity (which refers to differences in the characteristics of performing an activity by different people or by the same individual under different conditions) and inter-class similarity (which refers to different activities that are highly similar). Consequently, it is essential to extract features accurately from the dataset. Ensemble learning, which combines multiple models, is an effective approach to improve generalization performance. In this paper, a weighted ensemble of hybrid deep models for HAR using smartphone sensors is proposed. The proposed ensemble approach demonstrates superior performance compared to current methods, achieving impressive results across multiple evaluation metrics. Specifically, the experimental analysis demonstrates an accuracy of 97.15%, precision of 96.41%, recall of 95.62%, and an F1-score of 96.01%. These results demonstrate the effectiveness of our ensemble approach in addressing the challenges of HAR in real-world scenarios.

Keywords Human Activity Recognition · Ensemble learning · Deep Learning · Time series classification · Real-world dataset · Smartphone inertial sensors

✉ Jafar Tanha
tanha@tabrizu.ac.ir

Soodabeh Imanzadeh
sdb_imanzadeh@tabrizu.ac.ir

Mahdi Jalili
mahdi.jalili@rmit.edu.au

1 Electrical and Computer Engineering Department, University of Tabriz, Tabriz, Iran

2 School of Engineering, RMIT University, Melbourne, Australia

1 Introduction

1.1 Motivations

HAR is an elegant research field that has made remarkable contributions to ubiquitous computing [1–3], human behavior analysis, and human–computer interaction [4, 5]. Long-term monitoring of the activities of a person provides valuable insights into various diseases, like cardiovascular diseases [6], abnormal behaviors, and mental health [7, 8]. Furthermore, HAR systems have a wide range of applications, such as context-aware computing [1], user-based recommendations [4], elderly support [8, 9], fall detection, climate monitoring, traffic detection, training, security monitoring [4], military [3], internal leadership [3], time management, employee monitoring in various industries [10], entertainment [3], gaming [4], and the Internet of Things [11].

HAR has been extensively investigated based on ambient and wearable sensors [12]. Ambient sensors include various types of sensors such as motion, proximity, microphone, video [13], RFID [14], and infrared. Video sensors are increasingly used in HAR. However, they have infrastructure requirements [15], a limited coverage area [16], and privacy concerns [1, 14]. In contrast, wearable sensors, such as inertial sensors, get around these concerns, which makes them useful for HAR in smart homes. Wearable sensors are worn on the body and allow for continuous data collection and processing [16]. However, the high cost of sensors and their constant wear are disadvantages of using them [1, 10, 17, 18]. Smartphones, on the other hand, are equipped with sensors capable of recognizing human activities; They are not limited to a specific area or infrastructure, like external tools [19]. They are one of the portable devices that people increasingly interact with while performing daily tasks. There are two major concerns to address when using smartphones for HAR: the data collection setup and the activity recognition procedure.

The first key challenge is defining the experimental setup for data collection. Most available datasets are collected under controlled conditions in laboratory settings to ensure data accuracy. These limitations in the data collection step are: 1) the data is collected in the presence of a supervisor. 2) Most of the datasets are obtained through a specific path in a laboratory. 3) in most datasets, a special smartphone is used to collect the data. 4) many datasets employ multiple smartphones, smart watches, and wearable sensors at the same time. 5) the small number of people who participated in the data collection represents another limitation for the available datasets. To address these limitations, careful design of the data collection setup is necessary to ensure that the collected data is representative of real-world scenarios.

HAR faces challenges that set it apart from other supervised machine learning applications. The issues of the recognition procedure are: 1) intra-class diversity, which refers to differences in the characteristics of performing an activity by different people or by the same individual under different conditions; 2) inter-class similarity, which refers to different activities that are highly similar; 3) small datasets, which are a common issue for time series. Collecting and annotating time series data is a costly and time-consuming process; 4) sensory data are illegible and it is difficult to interpret relevant information from raw sensor data; and 5) sensory data often contain a large quantity of noise due to the use of inaccurate sensors and the inherent defects in the sensors themselves. In addition, each additional user movement introduces extra noise into the data. As a result, novel approaches should be developed to address the challenges.

Given the challenges in the HAR procedure, it is crucial to focus on learning inter-modality correlations and capturing intra-modality information. Sensing modality fusion can be achieved through two strategies: Feature Fusion, which combines different modalities to generate single feature vectors for classification, and Classifier Ensemble, which blends the outputs of classifiers operating solely on features from one modality [20].

Deep learning has demonstrated promising performance in various research areas like computer vision, speech recognition, and natural language processing. Ensemble learning combines multiple individual models to improve the overall generalization performance of a system. Deep ensemble learning combines the benefits of both deep learning and ensemble learning; indeed, the final model has superior generalization performance.

As mentioned earlier, recognizing human activities in real-world settings poses numerous challenges in the field of HAR. However, there is a noticeable scarcity of research that offers solutions specifically designed for the recognition of small and noisy datasets collected in such settings. This research gap has served as a strong motivation for us to undertake this study. In the next section, we provide a detailed description of our proposed approach, which aims to address the unique challenges associated with small and noisy datasets in order to provide effective solutions for accurate activity recognition in real-world scenarios.

1.2 Contribution

In this study, we aim to collect data in real-world settings to overcome the limitations of existing HAR datasets. To this end, we have developed an application to collect the data from inertial sensors on Android smartphones at the highest possible frequency. The application collects data from three smartphone sensors: the accelerometer, magnetometer, and gyroscope. We have recruited 62 men and women between the ages of 17 and 35 to participate in the experiments. Participants carry out various activities while holding their smartphones in their hands. Due to the use of several types of smartphones, the data frequency is variable and accompanied by noise. As a result, recognizing activity in this data brings up new issues.

Developing an accurate HAR model from the collected data presents several major challenges. These challenges include: 1) the primary challenge of this study is to identify effective strategies for incorporating sensor data into deep models. In other words, whether satisfactory performance is achieved by modeling each sensor independently and calculating an ensemble of the models, or if it is necessary to utilize hybrid models. 2) the dataset is gathered from a wide number of smartphones, and the maximum frequency at which smartphones store sensor output varies. As a result, the scale of the data differs widely among smartphones. Furthermore, sensor data frequency on a smartphone varies at different times due to the lack of computational power. 3) the data is accompanied by noise as a result of data collection settings such as a. collecting data in real-world conditions, b. utilizing different smartphones, c. not all smartphones being equipped with all sensors, and d. a large number of users engaging in the collection process. 4) the dataset is small, and a huge amount of data is required to develop a powerful deep model.

We have proposed an ensemble of hybrid deep models that simultaneously extract pertinent features from the output of accelerometers, gyroscopes, and magnetometers. We have conducted an experiment on the collected data using a collection of deep models. In this approach, the strategy outlined in [20] is employed, where the data from each sensor is inputted into an independent deep model. Subsequently, the ensemble of the sensor outputs is computed to obtain the final prediction. The findings have revealed that the magnetometer and gyroscope sensors alone are insufficient to produce a classification model with a suitable level of accuracy. Furthermore, the performance of all networks has shown a significant improvement when utilizing augmented data. However, the gyroscope and magnetometer models still do not support the ensemble model. The current ensemble model has not shown an improvement in the accuracy of models. Therefore, the next important challenge is to develop effective strategies for integrating sensor data into these models.

To address this challenge, we have proposed an ensemble of hybrid deep learning methods. There are three deep models for each classifier, which comprise three sub-models to simultaneously extract relevant features from accelerometer, gyroscope, and magnetometer data. Experimental results have demonstrated the effectiveness of the proposed approach, which yields a respectable recognition accuracy and outperforms conventional approaches in real-world scenarios.

The main contributions in this paper are presented as follows:

1. A new dataset is collected in a real-world setting using the inertial sensors (accelerometer, gyroscope, and magnetometer) on smartphones. The dataset comprises data from 62 participants and covers 7 activities.
2. A hybrid deep network is proposed for classifying the collected tiny and noisy dataset, achieving state-of-the-art recognition accuracy.
3. A weighted ensemble of hybrid deep models is proposed for HAR, utilizing the frequency of time series data to enhance the classification accuracy of the small and noisy time series dataset.
4. Experimental results demonstrate the effectiveness of the proposed approach with a recognition accuracy of 97.15%. These findings provide valuable insights for the development of more reliable and accurate HAR systems that can effectively operate in real-world scenarios.

1.3 Organization

The remaining of the paper is organized as follows: Section 2 explores reviews of the literature on inertial sensor datasets. Section 3 describes the proposed method. Section 4 addresses experimental results. Finally, Section 5 concludes with a summary of the exploration.

2 Literature review

2.1 Human activity recognition

HAR has recently gained the attention of many researchers all over the world. These systems recognize user activity using a variety of sensor measurements, like accelerometers. Bao and Intille [21] have offered one of the first HAR systems for the recognition of 20 activities of daily living using five wearable biaxial accelerometers. They are able to attain a classification accuracy of up to 84%, which is a respectable result considering the number of tasks involved. According to [22], the authors use two worn accelerometers and three microphones to identify repetitive actions like filling, drilling, and sanding. The authors attempt in [23] to detect and avoid falls in elderly individuals in smart homes. The majority of papers have used several accelerometers that are fixed in various locations across the human body [22, 24, 25]. Due to the numerous sensors attached to the human body and cable connections, this approach does not appear to be suited to the long-term study of daily life. Gyroscopes have also been used for HAR and have been shown to enhance the performance of recognition when combined with accelerometers [26, 27].

The smartphone is an alternative to wearable sensors because it supports a variety of sensors. Smartphones are a highly helpful tool for activity monitoring in smart homes due to their ability to handle sensors like accelerometers and gyroscopes. They are capable of handling wireless transmission and data processing [28]. Furthermore, they are widely used and practically never need a static infrastructure to function. Smartphones have recently been the focus of numerous activity recognition researchers due to their fast processing speeds and ease of deployment [29, 30]. The researchers in [31] found that relying on extracted features from deep models is not always suitable for distinguishing similar activities. To improve the accuracy of deep classifiers, they extracted handcrafted features. Multiple robust features are extracted from smartphone sensor signals by the authors [32]. They use KPCA to reduce the dimensions of the features and then use the Deep Belief Network for recognition. In [29], the authors collect user data from a chest unit made up of an accelerometer and vital sign sensors using wirelessly connected smartphones. Different machine learning techniques are then used to process and analyze the data. In [30], the authors create a HAR system using smartphone inertial sensors to recognize five transportation activities. The authors of [33] propose an offline HAR system that makes use of a smartphone with a built-in triaxial accelerometer sensor. Throughout the tests, the smartphone is kept in the pocket.

Significant assumptions are made about the production of noise-free data in the available HAR datasets. The following are some examples of limitations imposed on the collection of the HAR datasets:

- Some datasets only use one smartphone for the data collection phase. It necessitates the development of distinct models for each smartphone.
- Several datasets use multiple smartphones or other external sensors simultaneously [15, 34–38]. In the real world, it is not common to carry several smartphones at the same time.
- In some datasets, to ensure that the sensors move less and provide less noise, they are firmly worn to the body. The pockets [33, 35, 37, 39, 40], the belt [28, 34, 36, 38], the arm [34, 38], the waist [34, 37, 38, 41], and the head [41] are a few examples where sensors are placed.
- In the majority of datasets, users usually follow a predetermined path [28, 34–39].
- In some datasets, all applications are terminated to allocate all smartphone resources to the data collection application.

The characteristics of human activity recognition datasets using smartphones are shown in Table 1. The accuracy of activity classification in various publications is provided in the last column. While these datasets are useful for research purposes in controlled environments, they do not accurately represent the complexities of human activity in real-world situations. As such, there is a need for more diverse and representative datasets that can capture the activities performed by individuals in their daily lives.

Table 1 The publicly available HAR datasets using smartphone and smartwatch

	Sensors	Classes	Achieved Performance
SAD[34]	- Acc, Gyro, Mag* - Sp**(4) - arm, belt, waist and pocket	- Walking (8950), Standing (8950), Jogging (8950), Sitting (8950), upstairs (8950), downstairs (8900) - 6 Classes, 4 Users	[42, 43]: f1-score: 0.9(arm), 0.86(belt), 0.97(pocket), 0.89(wrist)
UCI HAR [28]	- Acc, Gyro - Sp - on the left side of the belt - Sample Rate: 50 Hz	- Walking (1722), Running, walking downstairs (1406), walking upstairs (1544), Standing (1506), Sitting (1777), Lying down (1944) - 7 Classes, 30 Users	[44]: accuracy: 97.5(DBN) and 94.12(SVM), [17, 45]: accuracy: 97.62(CNN + sharpen) vs 96.74(TSCHMM) vs 95.75(fft + CNN) vs 97.59(DCNN) vs 96.37(SVM) [46], [47]: accuracy: 93.5(Residual-BiLSTM), [48]: Accuracy: 92.93(CNN), [49]: accuracy: 84 (using 1 label for each user and activity), [50]: accuracy: 96.31(CNN-BiLSTM), [51]: accuracy: 96.9 (light weight model using Lego filters)
UniMiB-SHAR[35]	- Acc - Sp & wearable - their front gym trouser pockets - Sample Rate: 50 Hz	- Standing up from sitting (1.3%), standing up from lying (1.83%), walking (14.77%), running (16.86%), going up (7.82%), jumping (6.34%), going down (11.25%), lying down from standing (2.51%), sitting down (1.7%), falling forward (4.49%), falling left (4.54%), falling right (4.34%), falling backward (4.47%), falling backward sitting-chair (3.69%), falling with protection strategies (4.11%), falling and hitting obstacle (5.62%), syncope (4.36%) - 17 classes (9ADL, 8F), 24F and 6 M (aged between 18 and 60 years)	[52]: accuracy: 74.66, weighted f1: 74.16, average f1: 62.73

Table 1 (continued)

	Sensors	Classes	Achieved Performance
WISDM [33]	- Acc - Sp - (The Nexus One, HTC Hero, and Motorola Backflip. Not all in one time) - front pant pocket - Sample Rate: 20Hz	- Walking: 424,400 (38.6%), Jogging: 342,177 (31.2%), Upstairs:122,869 (11.2%), Downstairs: 100,427 (9.1%), Sitting:59,939 (5.5%), Standing: 48,395 (4.4%) - 6 Classes, 29 Users	[53]: f1-score: 0.98, [49]: accuracy: 71.34(using 1 label for each user and activity) accuracy: 98.84(light weight model using Lego filters)
HHAR [36]	- Acc - Sp(8),Sw**(4) - Sample Rate: max	- Biking (17,650), Sitting (19,169), Standing (17,751), Walking (20,385), Stair Up (16,905), Stair Down (15,199) - 6 Classes, 9 Users	[43]: f1-score: 0.95(accNexusS4), 0.84(accS3), 0.97(accS3mini), 0.95(accSamsungGold), 0.95(gyroNexusS4), 0.82(gyroS3), 0.92(gyroS3mini)
Fusion of SPs [38]	- Acc, LAcc, Gyro, Mag - Sp(5) in right jeans pocket, in left jeans pocket, on the belt position towards the right leg, on the right upper arm, on the right wrist - Sample Rate:50 Hz	- walking, running, sitting, standing, jogging, biking, walking upstairs and walking downstairs - 8 classes, 10 Users (male 25–30)	[38]: accuracy:97 (K Nearest Neighbors) (for the walking downstairs activity) Accuracy:96 (K Nearest Neighbors) (for the walking upstairs activity)
Complex HAR [37]	- Acc, LAcc, Gyro - Sp(2) in right pocket and on right wrist - Sample Rate:50 Hz	- Walking, standing, jogging, sitting, biking, upstairs, downstairs, typing, writing, drinking coffee, talking, smoking, eating - 13 classes, 10 Users(male)	[37]: accuracy of jogging: 96%, accuracy of biking: 93%, accuracy of typing: 96%
MobiAct [39]	- Acc, Gyro, orie - SP(S3) - Sample rate:max - in a pocket of the subject, in any random orientation	- ADLs:Standing, Walking, Jogging, Jumping, Stairs up, Stairs down, Sit chair, Car step in, Car step out Falls: - Fall Forward from standing, use of hands to dampen fall, fall forward from standing, first impact on knees, fall sideward from standing, bending legs, fall backward while trying to sit on a chair - 13 classes (9 ADL and 4 fall),57 Users (42 men and 15 women) (20 and 47 years)	[39]: accuracy of 99% for the involved ADLs

Table 1 (continued)

	Sensors	Classes	Achieved Performance
Motions Sense [40]	- Acc,Gyro, attitude - Sp(iPhone 6 s) front pocket - Sample rate: 20 Hz	- downstairs, upstairs, walking, jogging, sitting, and standing - 6ADLs, Users (16male 10female)	[40]: accuracy: 95.8 (multitask CNN)
On body [41]	- Acc, GPS, Gyro, light, Mag, sound level - Sp and sw (Samsung Galaxy S4 and LG G Watch R) - Chest, forearm, head, shin, thigh, upper arm, and waist - Sample rate:50 Hz	- climbing stairs down and up, jumping, lying, standing, sitting, running/jogging, and walking. (1065 min) - 15 users (age 31.9 ± 12.4) (8male 7female)	[41]: F-Measure: 89%

* Acc: accelerometer LAcc: Linear Accelerometer Gyro: gyroscope Mag: magnetometer

*** Sp: smartphone Sw: smartwatch*

2.2 Ensemble deep learning for HAR

Ensemble learning combines multiple individual models to enhance the overall generalization performance of a system. [54] introduced a novel ensemble Extreme learning machine algorithm specifically designed for human activity recognition using smartphone sensors. The experimental results indicate that this approach achieves recognition accuracies of 97.35% on the UCI-HAR dataset [28]. Deep learning has emerged as a highly successful approach in various domains, including computer vision [55–58], speech recognition [59, 60], and natural language processing [61, 62]. Additionally, it has found applications in autonomous vehicles, industrial robotics, medical diagnostics, and smart farming [63, 64]. Deep models are highly flexible, able to learn the complex relationships between variables and approximate any mapping function. However, the high flexibility of deep neural network models can lead to higher variance and overfitting. To address this challenge, ensemble deep learning has been proposed, which involves training multiple deep models and combining their predictions to improve generalization performance. Ensemble deep learning can mitigate the limitations of any one model and provide more robust and reliable predictions by leveraging the complementary strengths of individual models. Therefore, it is a promising approach for improving the generalization performance of deep learning models. Deep boosting [65], multiclass Deep Boosting [66], incremental Boosting CNN [67], and snapshot Boosting [68] incorporate boosting into the deep models to improve their performance. Moreover, Stacking Ensemble Deep Learning [69–71] and Negative Correlation Based Deep Ensemble Methods [72] are used in several publications. [73–76] have combined ensemble learning and deep learning to improve the accuracy of HAR.

Several studies have demonstrated the effectiveness of combining ensemble learning and deep learning to enhance the accuracy and robustness of HAR systems. For instance, [77] employs an ensemble of various CNN models, achieving an accuracy of 94% on the WISDM dataset [33]. Another study [73] combines a gated recurrent unit (GRU), a CNN stacked on the GRU, and a deep neural network, achieving an accuracy of 96.7% on the UCI HAR dataset. The study [74] proposes an easy ensemble approach for HAR that outperforms traditional ensemble techniques on multiple datasets, including UCI-HAR, WISDM, UniMiB SHAR, and PAMAP2 [78]. Study [75] proposes an ensemble model of CNN, achieving a classification accuracy of 96.11% on their collected dataset. [76] Introduces a fuzzy ensemble of three deep neural networks for HAR using on-body smart sensors. This model adaptively penalized activity classes in cases of assumed incorrect classification and employed a rewarding technique to extract the correct class in adverse situations. The proposed model demonstrates state-of-the-art accuracy on four publicly available wearable sensor datasets. In their study, Guo et al. [79] introduced a methodology for integrating multiple sensing modalities in HAR. They employed Multilayer Perceptron as the base classifier for each sensing modality and combined them through ensemble weights at the classifier level.

Additionally, Study [80] applies an ensemble of auto-encoders, associating each auto-encoder with a specific class. The experimental findings demonstrate the effectiveness, robustness, and competitiveness of this approach. In a different approach, Study [81] employs a personal area network where a smartphone serves as the main node, accompanied by supporting sensor nodes that provide supplementary data to improve recognition accuracy. The proposed method involves aggregating an ensemble of deep classifiers using RNNs. Moreover, Study [82] introduces an ensemble of four deep classification models, including ’CNN-net’, ’CNNLSTM-net’, ’ConvLSTM-net’, and ’StackedLSTM-net’. The evaluation of the proposed model is conducted on the WISDM, PAMAP2, and UCI-HAR datasets.

Lastly, [83] develops a customized HAR model that incorporates CNN and signal decomposition techniques. This model utilizes various signal processing methods, such as Ensemble EMD, Empirical Mode Decomposition (EMD), and Stationary Wavelet Transform, for feature extraction from multi-modal sensor data. The subsequent categorization and information fusion steps are performed using CNN. To personalize the model, the most suitable trained version of CNN is selected for the target subject by analyzing a few seconds of their data. Khan et al. [84] focused on the fall detection problem and proposed a novel approach that goes beyond the traditional classifier ensemble. They introduced an ensemble method based on the reconstruction error from the autoencoder for each sensing modality. These studies collectively demonstrate the effectiveness of ensemble approaches in improving the accuracy of HAR systems, utilizing various deep learning architectures and datasets.

Ref	The objective of the paper	Model	datasets	result
[73]	to monitor individuals at risk of COVID-19 virus infection and manage their activity status, especially considering the widespread isolation and quarantine measures due to the pandemic	a gated recurrent unit (GRU), a convolutional neural network (CNN) stacked on top of the GRU, and a deep neural network (DNN)	UCI-HAR	Accuracy: 96.7 F1-score: 96.8
			WISDM	F1-score: 91.7
			Opportunity	F1-score: 87.4
[74]	EASY ENSEMBLE: the implementation of deep ensemble learning within a single model	VGG	HASC	Accuracy ~ 84
[75]	To Collect a new HAR dataset to address the challenges associated with HAR using smartphone inertial sensors and improve the accuracy of activity classification	Ensemble CNN	Collected dataset	Accuracy:96.11
[76]	To address the limitation of existing models that lack the ability to correct wrong classifications made by a base classifier	Fuzzy Ensemble	MHealth	Accuracy: 100 F1-score: 100
			USC-HAD	Accuracy:96.52 F1-score: 95.30
			WHARF	Accuracy:91.93 F1-score: 88.62
			OPPORTUNITY	Accuracy:89.39 F1-score:90.74
[77]	improve HAR by utilizing ensemble learning	CNN	WISDM	Accuracy: 94
[79]	Multimodal Activity Recognition with Ensemble Classifier	neural network	PAMAP2	Accuracy: 84.8
			MHealth	Accuracy: 92.3
[80]	To develop an efficient and robust approach that can automatically extract complex features from sensor data	ensemble of auto-encoders	WISDM	Accuracy: 82
			MHealth	Accuracy: 82
			PAMAP2	Accuracy: 63

Ref	The objective of the paper	Model	datasets	result
[81]	To improve the accuracy of HAR, particularly training routines like squats, jumps, or arm swings	RNN	Collected dataset	Accuracy:99.5
[83]	to develop a personalized HAR model	combines signal processing techniques (Stationary Wavelet Transform, Empirical Mode Decomposition, and Ensemble EMD) for feature extraction, followed by the use of CNN for information fusion and final classification	MHealth	Accuracy:72.13
[82]	To accurately HAR for applications in medical care, fitness trackers, senior care, and archiving patient information	CNN, CNNLSTM, ConvLSTM, and StackedLSTM	WISDM PAMAP2 UCI-HAR	Accuracy: 98.7 Accuracy:97.45 Accuracy:95.05

As highlighted in the literature, there is a scarcity of research on the application of ensemble methods to HAR in small, noisy, and real-world datasets. This paper introduces an innovative ensemble approach, utilizing a combination of hybrid deep models for HAR. The proposed ensemble approach achieves impressive results across multiple evaluation metrics to address this gap. Impressive results are obtained across several assessment metrics using the proposed ensemble approach.

3 Proposed algorithm to human activity recognition

HAR is the process of interpreting human activities using machine learning technology. In this section, a new HAR dataset is collected in a real-world setting. Then, a novel approach is proposed to classify the noisy data.

3.1 Data collection

As mentioned earlier, existing HAR datasets have major constraints on producing noise-free data. Therefore, datasets are less useful in real-world situations. In this research, a data collection application for Android smartphones has been developed. This application collects data from the accelerometer, magnetometer, and gyroscope at the highest feasible sample rate on the smartphone. Each participant is asked to perform a sequence of activities, including walking, standing, running, walking up and down stairs, driving, and resting the smartphone on a flat surface (still).

The dataset is collected under the following conditions: participants are asked to hold their Android smartphones in their hands rather than wear them. The participants use Android smartphones manufactured by various companies, such as Samsung, Xiaomi, LG, and Honor. The smartphones did not necessarily have all the required sensors. Participants perform the experiments either individually or in groups outside of the laboratory, on arbitrary paths, for a few minutes. They are not forced to take part in all activities. The data collection application operates alongside other open applications. A total of 32 men and 30

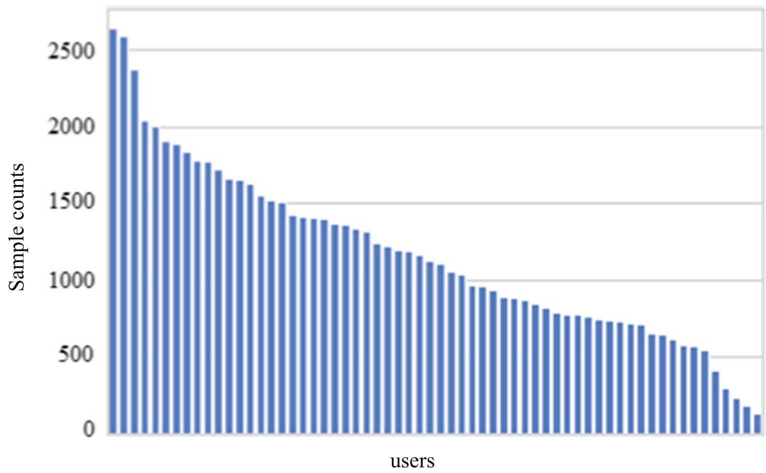

Fig. 1 Distribution of samples for each user

Table 2 The number and percentage of each class in the collected dataset

Class	#samples	Percentage
Walking	435,300	34.773
Static	168,675	13.474
Still	162,320	12.967
Running	158,665	12.675
Walking up stairs	117,690	9.402
Walking down stairs	109,829	8.774
Driving	99,337	7.935

women, ranging in age from 17 to 37, take part in the data gathering. Figure 1 displays the distribution of samples collected for each user. Participants self-report their activity labels within the application without direct supervision. It is important to note that there are short transitional walks (2–4 steps) between floors in the stairwells. These are labeled as walking up and down stairs.

The sensor signal reflects the output of sensors along the three Cartesian axes, resulting in a triplet of values (x, y, and z). Table 2 and Fig. 2 display the distribution of the collected dataset among the various classes.

3.2 Problem formulation

The dataset is collected under real-world circumstances, and it presents several issues that need to be addressed:

1. The final data frequency is not uniform. This is because different smartphones have varying maximum frequencies for recording sensor data.

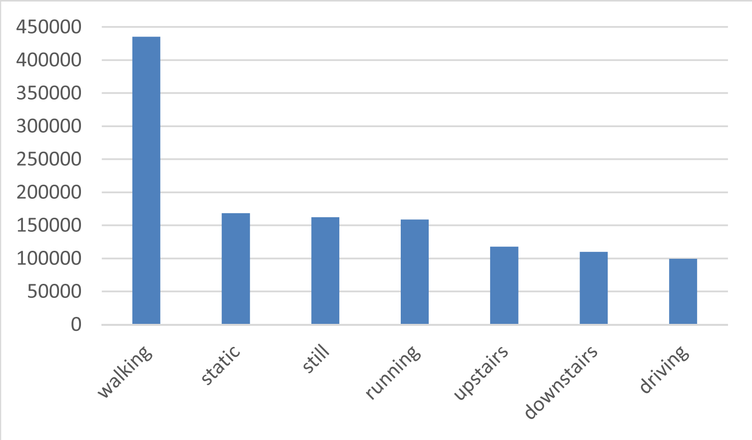

Fig. 2 Distribution of Samples across Different Classes in the Collected Dataset

2. All resources on an Android smartphone, including memory and CPU, are shared among all applications. This means that if resources are limited, the operating system cannot guarantee that a task, such as recording sensor data, will be completed at a specified time. Consequently, the data frequency of a smartphone varies. Figure 3 shows the variation in the number of samples taken by a sensor per second. Furthermore, using multiple smartphones leads to more missing data. Since most smartphones do not have all of the aforementioned sensors.
3. Since the data is collected from real-world settings, the dataset is inherently noisy.

A multivariate time series $X = \{X_1 3, X_2, \ldots, X_t, \ldots, X_T\}$ is denoted as a sequence of T observations. The t-th observation $X_t \epsilon \mathbb{R}^D$ consists of D features $\{x_t^1, x_t^2, \ldots, x_t^D\}$, and is observed at timestamp s_t. As mentioned, the time gap between different timestamps may not be the same, and the X has missing values.

The primary objective of the research is to develop an accurate HAR model using the inertial sensors of a smartphone in real world settings. Initially, an experiment is conducted using an ensemble of deep models. The accelerometer, gyroscope, and magnetometer sensors on the smartphone are employed in this approach. Features are extracted separately from each sensor using three CNN-based networks. The ensemble of these three networks is then computed. The results of this experiment are presented in Table 3. The results

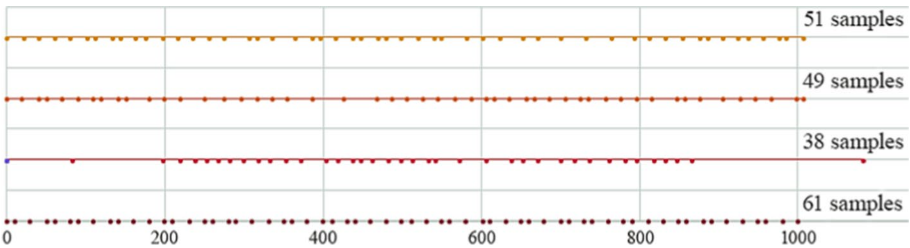

Fig. 3 Frequency Variation of Smartphone Accelerometer Data within a 1000-Millisecond Interval

Table 3 results of the Basic Ensemble of Deep Models

Data	Accuracy
Accelerometer	87.04
Gyroscope	65.39
Magnetometer	60.34
ensemble	82.24

indicate that the ensemble model does not achieve higher accuracy than the best individual learner.

Data augmentation is known to increase the classification accuracy of machine learning models by reducing the risk of overfitting. Accordingly, in the next step, the raw data is augmented and fed into the model. The results are presented in Table 4, which shows that augmenting the dataset has improved the accuracy of the accelerometer, gyroscope, and magnetometer models by 6, 19, and 12 percent, respectively. Because the dataset is noisy and relatively small, augmenting the dataset significantly improved the overall accuracy of each model. Finally, the accuracy of the ensemble model has increased by 6%. However, it is still below the maximum accuracy of the models. Hence, the ensemble model does not effectively integrate the information from each sensor.

To address this issue, a deep hybrid ensemble model is proposed, leveraging the strengths of each individual model while mitigating their weaknesses. The hybrid deep model is designed to incorporate both the raw sensor data and the augmented data, which could improve the accuracy of the model by reducing the impact of noise in the signals.

3.3 Preprocess

Preprocessing is an essential step in preparing the data for neural networks during the classification process. Since the data gathering step takes place in a real-world environment, the raw data contains noise, incorrect data, and missing samples. The data for each activity of every participant is recorded as a sub-dataset and then sent to a server. There are five columns in each sub-dataset: timestamp, sensor name, x, y, and z values. The rows represent the output of a sensor at a special timestamp.

The activities are performed without any supervision, and participants annotate the ground truth. Therefore, the sub-datasets that take less than 15 s are eliminated. These samples are considered noise since they are not related to any of the activities. Additionally, the first and last five seconds of each sub-dataset are excluded since participants need time to touch the start and stop recording buttons.

Table 4 Results of the Basic Ensemble of Deep Models with Augmented input

Data	Accuracy
Accelerometer	93.61
Gyroscope	84.7
Magnetometer	72.36
ensemble	88.18

The following step involves modifying the dataset structure. Each row in the sub-datasets contains the output of a sensor for a specific timestamp. The timestamp column displays the current time on the smartphone in nanoseconds. The redundant timestamps are then removed. The rows with the same time stamp are combined into a single row. The start time stamp of sub-datasets is set to zero using Eq. 1 to standardize the timestamps across all of them. To ensure that the timestamps are sequential, the time stamp of each sub-dataset is combined with the timestamp of the preceding one (Eq. 2). t_j^i is the observation in the i-th sub-dataset at the t-th time stamp.

$$t_j^i = t_j^i - t_0^i \quad \textit{for all } t_{j>0}^i \tag{1}$$

$$t_0^i = t_{last\ row}^{i-1} + 20 \tag{2}$$

The timestamp intervals between continuous rows vary because the data frequencies are different. Additionally, there are instances where the sensor output is not saved for more than one second due to factors like a shortage of operating system resources. If the time stamp interval between the consecutive rows is greater than one second, it is reduced to 20 ms using Eq. 3.

$$\textit{if } (t_j^i - t_{j-1}^i > 1000\ ms)\ \textit{then } (t_j^i = t_{j-1}^i + 20)\ \textit{for } (t_j^i > t_{j-1}^i) \tag{3}$$

Next, the sub-datasets are concatenated into a single dataset. Consequently, a dataset with 25 columns is obtained, which includes the following data: time stamp, accelerometer data (in the direction of x, y, and z), gyroscope data (in the direction of x, y, and z), magnetometer data (in the direction of x, y, and z), and activity label. Figure 5 depicts the feature space of the dataset.

3.4 Time based sliding window

A single data point from a sensor provides a brief position of the user, similar to an image snapshot in a video. The activities consist of a series of sensor outputs across time, similar to images in a video. The sliding window technique involves selecting a sequence of data points from a time series, as illustrated in Fig. 6. Since the windows overlap, some data points are included in multiple windows.

3.5 Proposed ensemble of hybrid models

This paper proposes an ensemble of hybrid deep models to enhance the accuracy of HAR in noisy and small datasets. The ensemble approach combines the predictions of the hybrid

	Time stamp	Acc(x)	Acc(y)	Acc(z)	…	Mag(x)	Mag(y)	Mag(z)	Label
$X_1 \rightarrow$	S_1	x_1^1	x_1^2	x_1^3	…	x_1^7	x_1^8	x_1^9	y_1
$X_2 \rightarrow$	S_2	x_2^1	x_2^2	x_2^3	…	x_2^7	x_2^8	x_2^9	y_1
$X_3 \rightarrow$	S_3	x_3^1	x_3^2	x_3^3	…	x_3^7	x_3^8	x_3^9	y_1
	…	…	…	…	…	…	…	…	
$X_{1.3Million} \rightarrow$	$S_{1.3M}$	$x_{1.3M}^1$	$x_{1.3M}^2$	$x_{1.3M}^3$	…	$x_{1.3M}^7$	$x_{1.3M}^8$	$x_{1.3M}^9$	y_7

Fig. 5 The structure of the raw multivariate time series dataset

Fig. 6 Sliding Window with a Window Size of 90 Rows and 50% Overlap

deep models to improve overall accuracy. The hybrid deep model extracts features from the accelerometer, gyroscope, and magnetometer. Figure 7 provides a general overview of the proposed hybrid deep approach.

Feature extraction using the proposed deep hybrid model: As mentioned earlier, the collected raw dataset is completely noisy and small. It does not have a constant frequency and contains missing data. In the first stage, the raw data is interpolated to N frequencies. The number of frequencies and their values are hyperparameters that depend on the dataset. Next, trainset_0 is created by concatenating the raw data and all of its frequencies. Subsequently, trainset_0 is modified in M steps by incorporating various augmentations. The number of steps and types of augmentations are hyperparameters. As a result of this step, M + 1 trainsets are obtained, as depicted in Fig. 8. In the third phase, M + 1 deep models are selected to train these trainsets. The models and their characteristics are

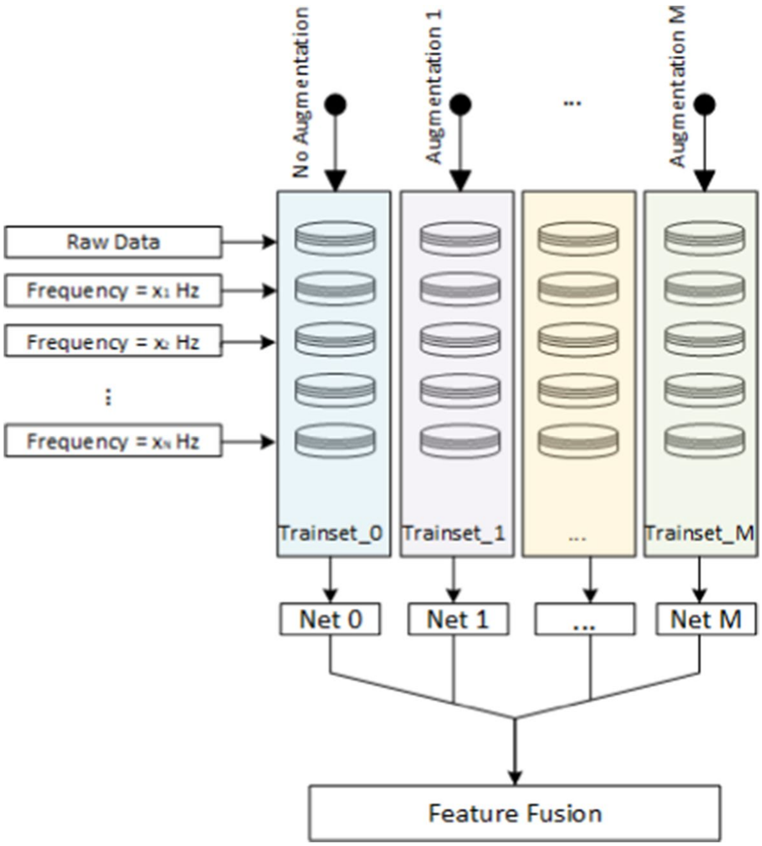

Fig. 7 General overview of the proposed hybrid deep model

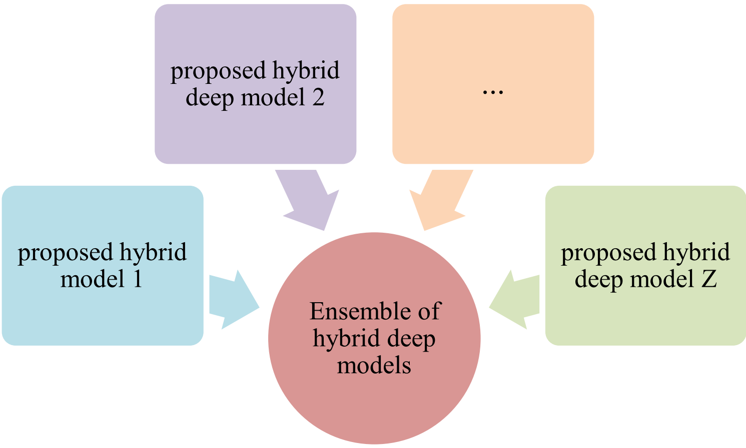

Fig. 8 Proposed Ensemble of Hybrid Deep models

hyperparameters that can be tuned to achieve optimal performance. Finally, the high-level features extracted from each model are concatenated.

Proposed ensemble of hybrid deep models: The first hybrid deep model is created by training the proposed model. Subsequently, the proposed model is trained with various permutations of trainsets to obtain additional models. Finally, a weighted ensemble of deep hybrid models is computed.

The Logcosh loss function (Eq. 4) is used to calculate the difference between the predicted and actual output of the model. In this loss function, n represents the number of data points, y denotes the actual label of data points, and y ̂ represents the predicted value of data points returned by the model. The Logcosh loss function has been found to outperform other loss functions for the entirely noisy dataset. It is less sensitive to outliers and can handle noisy data more effectively, resulting in improved accuracy.

$$logcosh = \frac{1}{n}\sum_{i=1}^{n}\log(\cosh(\hat{y}_i - y_i)) \tag{4}$$

Overall, the proposed ensemble approach combines the strengths of multiple hybrid deep models to improve the accuracy of HAR in noisy small datasets and provides a more reliable method for it. The proposed approach is outlined in Algorithm 1.

Algorithm 1 The proposed Ensemble of hybrid deep models

1. **Split** the dataset into a training set (**x_train**) and a test set (**x_test**) randomly.
2. Define the **hyperparameters**:
   - **Frequencies** for **data interpolation**: f0, f1, ..., fN.
   - **Augmentation** methods: aug[0:M]. (aug[0] returns the data as is)
   - **Sub_models**: net[0:M].
   - **Number of hybrid deep classifiers** in the ensemble model: classifier_count.
3. **Interpolate** the training set (x_train) to different frequencies:
   - Interpolate x_train to f0 frequency: x_train_f[0]
   - Interpolate x_train to f1 frequency: x_train_f[1]
   - ...
   - Interpolate x_train to fN frequency: x_train_f[N]
4. Apply sliding window (w) on the x_train, x_train_f[:], and test_set.
5. Feature extraction:

```
for i in range(classifier-count):
  shuffle(aug)
  for m in range(M+1):
    features[m] = net[m](aug[m](concatenate(x_train, x_train_f[:N])))
  end for
  classifier[i] = concatenate(features[:M+1])
end for
```

6. **compute ensemble** of classifier[:classifier-count]

4 Experiments and results

As mentioned earlier, we train three individual CNN-based learners for the accelerometer, gyroscope, and magnetometer. Then we compute an ensemble of them using the majority voting method. However, the results show that the ensemble model has incredibly poor accuracy. Indeed, the accuracy of the ensemble model has been lower compared to the models on the aforementioned noisy dataset. Therefore, a deep hybrid ensemble model is proposed to increase classification accuracy. In this section, we evaluate the effectiveness of the proposed model through experiments on the noisy dataset.

We reserve 20% of the dataset as the test set before applying the sliding window procedure with a 50% overlap. This ensures that the training set and test set do not have any overlap or intersection. It allows for a reliable evaluation of the proposed model on unseen data. We allocate 20% of the data as the validation set, which is used for hyperparameter tuning.

The first step is to prepare the raw data. We augment the raw data by interpolating it to 20, 25, 40, and 50 Hz. The optimal frequencies for HAR in this dataset have been obtained through the cross-validation method. To develop the hybrid deep model, we select four sub-models: a custom ResNet, a CNN with two layers, a custom VGGNet, and a CNN-LSTM network. Figure 9 illustrates the layers of the proposed hybrid model. The sub-models are fed with the raw data and its various frequencies. The first sub-model receives the pure data, while the other sub-models are fed with augmented data. In this paper, we employ jittering, magnitude warping, and permutation. The extracted high-level features are then concatenated.

In this study, the proposed hybrid deep model is trained with a maximum of 300 epochs, implementing an early stop mechanism to prevent overfitting. The learning rate and batch size are set to 0.0001 and 128, respectively. The ReLU activation function is employed to introduce non-linearity and capture complex patterns in the data. Dropout regularization is incorporated with a dropout rate of 0.5 in the 2-layer CNN sub-model and a rate of 0.7 in the CNN-LSTM sub-model. The Adam optimization algorithm is utilized for efficient parameter updates and adaptive learning rate adjustment. This hybrid model is trained three times using different input permutations for each sub-model. Finally, the ensemble of models is calculated using the averaging method. The results of four deep hybrid models and the ensemble of them are presented in Table 5. Figures 10 and 11 present the training and validation loss plots and the training and validation accuracy plots, respectively, for four hybrid deep models.

In the next step, we employ the weighted ensemble of deep hybrid models. Table 6 provides the weights and the outcomes. Figure 12 summarises the proposed approach performance on the collected dataset. According to experiments, the weighted ensemble performs better than the best individual models and the other ensemble model. Figure 5 illustrates that the best model has a maximum accuracy of 96.16 percent. However, the proposed ensemble method and the weighted ensemble achieve 96.9 and 97.15 percent, respectively.

Figure 13 demonstrates the confusion matrices of four models and the proposed ensemble approaches. The results of a Confusion Matrix are categorized into four groups. A True Positive (TP) occurs when the model accurately predicts the positive class of a window. A False Positive (FP) is produced when the model predicts the positive class of an activity inaccurately. A True Negative (TN) is when the model predicts the negative class of a window accurately. False Negatives (FN) are produced when the model incorrectly predicts the negative class.

In multiclass problems, the true label is considered the positive class, while the other labels are considered the negative class. Several performance measures have been

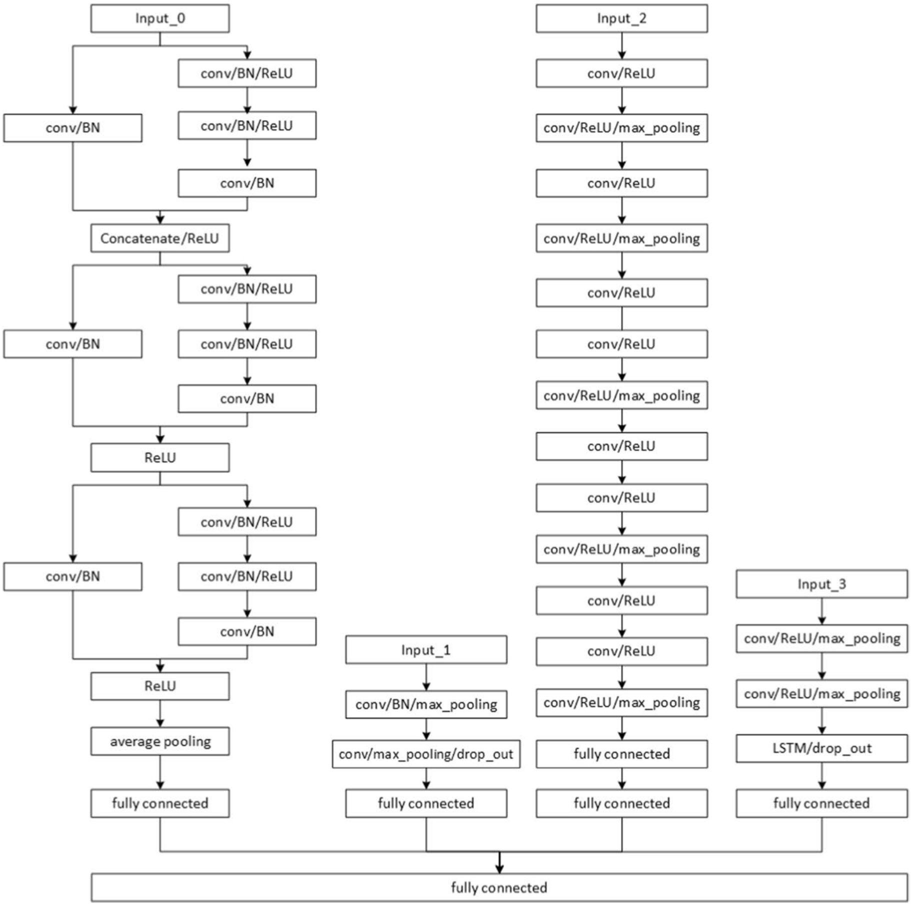

Fig. 9 Architecture of proposed hybrid deep model

Table 5 Performance comparison among various classifiers and the proposed ensemble of hybrid Deep model over collected dataset

Input order for the proposed model	Accuracy
The raw data, jittering*, permutation**, magnitude***	96.17
Magnitude, the raw data, jittering, permutation	94.24
Jittering, permutation, magnitude, the raw data	95.99
Permutation, magnitude, the raw data, jittering	95.85
Ensemble of the models	96.9

* jittering: jittering applied to the raw data

** permutation: permutation applied to the raw data,

*** magnitude: magnitude warping applied to raw data

developed for the confusion matrix. The evaluation indicators employed in this research are accuracy, precision, recall, and F1-score. Recall is the ability of the model to predict the positives. The precision of a class is defined as the ratio of true positives to total

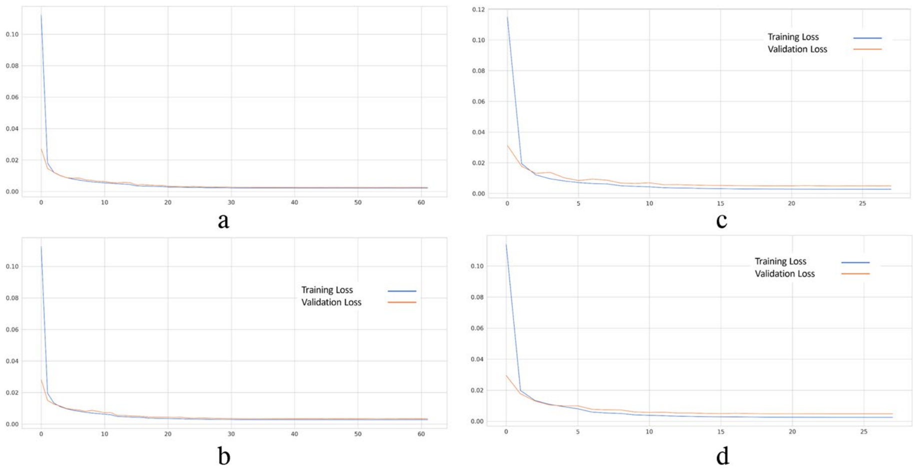

Fig. 10 Training and validation loss of **a**) first model, **b**) second model, **c**) third model, **d**) forth model

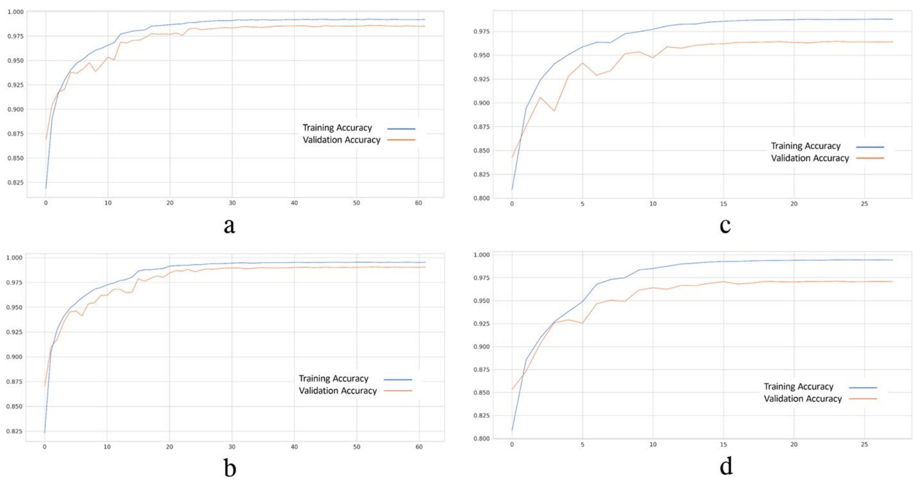

Fig. 11 Training and validation accuracy of **a**. first model, **b**. second model, **c**. third model, **d**. forth model

positive predictions. The harmonic mean of recall and precision is the F1-score. Table 7 is employed to calculate the recall, precision, and f1-score. Table 8 presents the per-class precision, recall, and f1-score for the hybrid deep models and the proposed ensemble models. These experimental findings are analyzed to evaluate the characteristics, advantages, and disadvantages of the model.

Based on these evaluations, the proposed approach outperforms the previous models in terms of accuracy, precision, recall, and f1-score. The proposed ensemble approach yields an accuracy, precision, recall, and f1-score of 96.9%, 96.89%, 96.89%, and 96.89%, respectively. While the proposed weighted ensemble approach surpasses this with an accuracy, precision, recall, and f1-score of 97.15%, 97.14%, 97.15%, and 97.14%, respectively.

Based on the confusion matrix analysis, it is observed that the 'walking down stairs and up stairs' classes are frequently misclassified as the 'walking' class in all models.

Table 6 Performance comparison among various classifiers and the proposed Weighted Ensemble of hybrid Deep model over collected dataset

Input order for the proposed model	weight	Accuracy
The raw data, jittering*, permutation**, magnitude***	1.1	96.17
Magnitude, the raw data, jittering, permutation	1	94.24
Jjittering, permutation, magnitude, the raw data	1.1	95.99
Permutation, magnitude, the raw data, jittering	1	95.85
Ensemble of models		**97.15**

* jittering: jittering applied to the raw data

** permutation: permutation applied to the raw data,

*** magnitude: magnitude warping applied to raw data

However, the proposed models demonstrate improved performance in this aspect. The proposed model makes 29 and 18 incorrect predictions out of 185 and 201 instances in the "walking up stairs" and "walking down stairs" classes, respectively. In contrast, the average misclassification instances of the hybrid deep models for these classes are 34 and 24, respectively.

Figure 14 illustrates the per-class f1-score comparison among all models. The results indicate that the "walking up stairs" class has the lowest F1-score across all models. Therefore, classifying this activity is particularly challenging. In contrast, the "driving" and "running" classes have the highest F1-scores. While the proposed models show promising results, there is still room for further improvement in accurately classifying the 'walking up stairs' class. Overall, the proposed ensemble models demonstrate improved performance compared to the base models. These results indicate that the proposed approach has potential for practical applications in activity recognition.

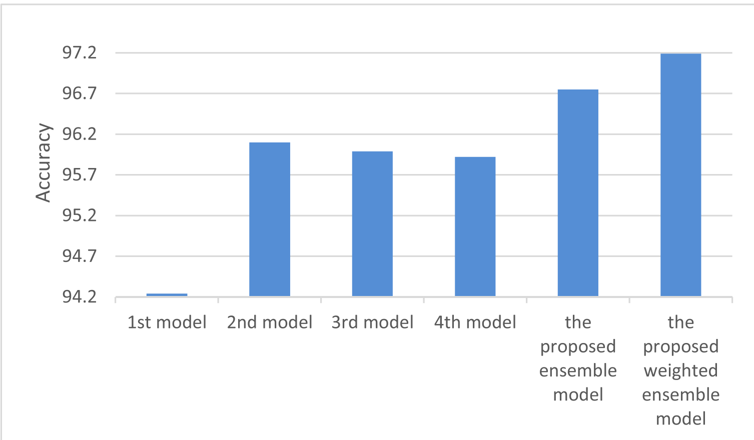

Fig. 12 Performance comparison among various classifiers and the proposed Ensemble of hybrid Deep model models over collected dataset

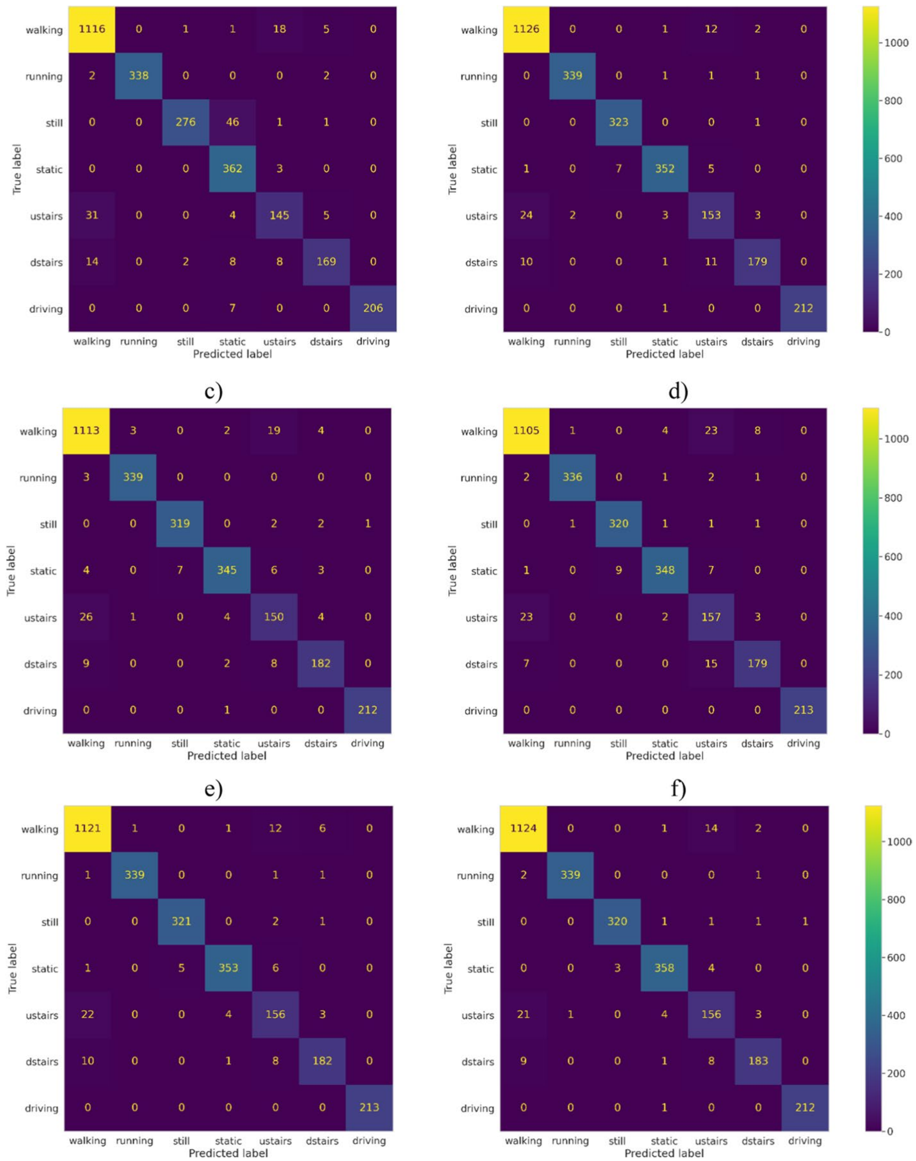

Fig. 13 Confusion matrix of **a**) 1st model, **b**) second model, **c**) third model, **d**) forth model, **e**) the proposed Ensemble of hybrid Deep models, and **f**) the proposed Weighted Ensemble of hybrid Deep models

Table 7 Definition of the classification evaluation metrics

metrics	formula
Accuracy	$\frac{(TP+TN)}{(TP+FP+TN+FN)}$
Recall	$\frac{TP}{(TP+FN)}$
Precision	$\frac{TP}{(TP+FP)}$
F1-score	$2 * \frac{Precision*Recall}{Precision+Recall}$

Table 8 Classification results of a) 1st model, b) second model, c) third model, d)forth model, e) the proposed Ensemble of hybrid Deep models, and f) the proposed Weighted Ensemble of hybrid Deep models

a)			
	precision	recall	f1-score
1	0.9680	0.9798	0.9739
2	0.9971	0.9883	0.9927
3	0.9755	0.9815	0.9785
4	0.9746	0.9479	0.9611
5	0.8148	0.8324	0.8235
6	0.9471	0.8905	0.9179
7	0.9725	0.9953	0.9838
macro avg	0.9499	0.9451	0.9473
weighted avg	0.9619	0.9617	0.9617
b			
model	precision	recall	f1-score
1	0.9596	0.9781	0.9688
2	1.0000	0.9883	0.9941
3	0.9892	0.8519	0.9154
4	0.8458	0.9918	0.9130
5	0.8286	0.7838	0.8056
6	0.9286	0.8408	0.8825
7	1.0000	0.9671	0.9833
macro avg	0.9360	0.9145	0.9232
weighted avg	0.9452	0.9426	0.9423
c			
	precision	recall	f1-score
1	0.9636	0.9755	0.9695
2	0.9883	0.9912	0.9898
3	0.9785	0.9846	0.9815
4	0.9746	0.9452	0.9597
5	0.8108	0.8108	0.8108
6	0.9333	0.9055	0.9192
7	0.9953	0.9953	0.9953
macro avg	0.9492	0.9440	0.9465
weighted avg	0.9599	0.9599	0.9599
d			
	precision	recall	f1-score
1	0.9702	0.9693	0.9698
2	0.9826	0.9883	0.9854
3	0.9755	0.9815	0.9785
4	0.9719	0.9479	0.9598
5	0.7833	0.8595	0.8196
6	0.9372	0.8905	0.9133
7	0.9953	0.9859	0.9906
macro avg	0.9451	0.9461	0.9453
weighted avg	0.9596	0.9585	0.9589
e			
	precision	recall	f1-score
1	0.9706	0.9825	0.9765

Table 8 (continued)

2	0.9971	0.9912	0.9941
3	0.9847	0.9907	0.9877
4	0.9833	0.9671	0.9751
5	0.8432	0.8432	0.8432
6	0.9430	0.9055	0.9239
7	1.0000	1.0000	1.0000
macro avg	0.9603	0.9543	0.9572
weighted avg	0.9689	0.9690	0.9689
f			
	precision	recall	f1-score
1	0.9723	0.9851	0.9787
2	0.9971	0.9912	0.9941
3	0.9907	0.9877	0.9892
4	0.9781	0.9808	0.9795
5	0.8525	0.8432	0.8478
6	0.9632	0.9104	0.9361
7	0.9953	0.9953	0.9953
macro avg	0.9642	0.9563	0.9601
weighted avg	0.9714	0.9715	0.9714

5 Conclusion

HAR via wearable sensors and smartphones has become an increasingly important area of research in recent years. This system has the potential to be applied in various fields such as e-health, human behavior analysis, and context-aware computing. The majority of HAR datasets are collected in laboratory environments, and they are inappropriate for use

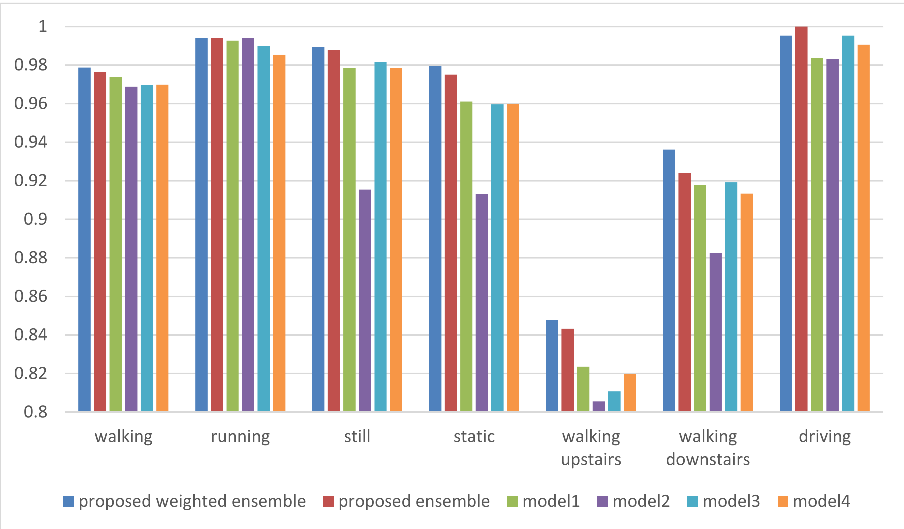

Fig. 14 F1-score of each activity across the models

in the real world. We have gathered a new HAR dataset with the assistance of 62 men and women. Participants perform the activities (walking, static, still, running, walking up and down stairs, and driving) using their own Android smartphones. The values recorded by the accelerometer, gyroscope, and magnetometer are saved at the highest frequency a smartphone can handle. Due to the use of a large number of smartphones, the frequency of the dataset is not constant. Additionally, the dataset is small and noisy. We propose a novel ensemble of deep hybrid models to classify this dataset. The accuracy of the proposed approach is 97.15 percent. The new dataset and proposed approach represent a significant step towards improving the reliability and accuracy of HAR in real-world applications.

In the following, we recommend five strategies for future work:

1. The collected dataset is imbalanced. For instance, the "walking" class accounts for 34% of the data, whereas the "driving" class accounts for roughly 8% of the data. To address this, we recommend balancing the dataset before classification to avoid potential biases.
2. Since the dataset is unreliable, some misclassified samples may belong to a class other than the declared label. Consequently, modifying the approach does not lead to improved model performance. We propose to consider this dataset as unlabeled. On the other hand, collect a reliable, labeled dataset. Therefore, we can solve the issue in a semi-supervised manner.
3. While the current data collection involves holding the phone by the user, we acknowledge the importance of capturing variations in phone placement scenarios. We propose collecting a new dataset that includes diverse scenarios, such as placing the phone in a bag or pocket. This would enhance the diversity of the dataset and improve the generalization capabilities of the models.
4. We propose an end-to-end HAR system for continuous monitoring. Such systems find applications in various domains, including monitoring patient rehabilitation progress, assessing movements and activities to detect abnormalities, ensuring safety in industrial settings by monitoring worker activities, and controlling game characters in virtual environments through body movements.
5. Finally, we propose collecting a new dataset for complex activities via smartphone in a real-world environment. This dataset would enable the development and evaluation of HAR models in more challenging scenarios.

Data availability The datasets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Conflicts of interest The authors declare that there is no conflict of interest.

References

1. Chen Z, Zhang L, Cao Z, Guo J (2018) Distilling the knowledge from handcrafted features for human activity recognition. In: IEEE Trans Industr Inform 14(10):4334–4342
2. Ronao CA, Cho S-B (2016) Human activity recognition with smartphone sensors using deep learning neural networks. Expert Syst Appl 59:235–244

3. Murad A, Pyun J-Y (2017) Deep recurrent neural networks for human activity recognition. Sensors 17(11):2556
4. Hussain Z, Sheng QZ, Zhang WE (2020) A review and categorization of techniques on device-free human activity recognition. J Netw Comput Appl 167:102738
5. Cornacchia M, Ozcan K, Zheng Y, Velipasalar S (2017) A survey on activity detection and classification using wearable sensors. IEEE Sens J 17(2):386–403
6. Barengo NC, Antikainen R, Borodulin K, Harald K, Jousilahti P (2017) Leisure-time physical activity reduces total and cardiovascular mortality and cardiovascular disease incidence in older adults. J Am Geriatr Soc 65(3):504–510
7. Lubans D et al (2016) Physical activity for cognitive and mental health in youth: a systematic review of mechanisms. Pediatrics 138(3)
8. Ihianle IK, Nwajana AO, Ebenuwa SH, Otuka RI, Owa K, Orisatoki MO (2020) A deep learning approach for human activities recognition from multimodal sensing devices. IEEE Access 8:179028–179038
9. Cvetković B, Szeklicki R, Janko V, Lutomski P, Luštrek M (2018) Real-time activity monitoring with a wristband and a smartphone. Information Fusion 43:77–93
10. Matsui S, Inoue N, Akagi Y, Nagino G, Shinoda K (2017) User adaptation of convolutional neural network for human activity recognition. In: 25th IEEE European Signal Processing Conference (EUSIPCO). 753–757
11. Li Q, Gravina R, Li Y, Alsamhi SH, Sun F, Fortino G (2020) Multi-user activity recognition: Challenges and opportunities. Information Fusion 63:121–135
12. Chen Y, Shen C (2017) Performance analysis of smartphone-sensor behavior for human activity recognition. IEEE Access 5:3095–3110
13. Ma S, Sigal L, Sclaroff S (2016) Learning activity progression in lstms for activity detection and early detection. In: Proc IEEE Conf Comput Vis Pattern Recognit pp 1942–1950
14. Yao L, Sheng QZ, Li X, Wang S, Gu T, Ruan W, & Zou W (2015) Freedom: online activity recognition via dictionary-based sparse representation of rfid sensing data. In: 2015 IEEE international conference on data mining. IEEE, pp 1087-1092
15. Vaizman Y, Ellis K, Lanckriet G (2017) Recognizing detailed human context in the wild from smartphones and smartwatches. IEEE Pervasive Comput 16(4):62–74
16. Cui W, Li B, Zhang L, Chen Z (2021) Device-free single-user activity recognition using diversified deep ensemble learning. Appl Soft Comput 102:107066
17. Deng W-Y, Zheng Q-H, Wang Z-M (2014) Cross-person activity recognition using reduced kernel extreme learning machine. Neural Netw 53:1–7
18. Kwon Y, Kang K, Bae C (2014) Unsupervised learning for human activity recognition using smartphone sensors. Expert Syst Appl 41(14):6067–6074
19. Catal C, Tufekci S, Pirmit E, Kocabag G (2015) On the use of ensemble of classifiers for accelerometer-based activity recognition. Appl Soft Comput 37:1018–1022
20. Chen K, Zhang D, Yao L, Guo B, Yu Z, Liu Y (2021) Deep learning for sensor-based human activity recognition: overview, challenges, and opportunities. ACM Comput Surv (CSUR) 54(4):1–40
21. Bao L, Intille SS (2004) Activity recognition from user-annotated acceleration data. In: International conference on pervasive computing. Springer Berlin Heidelberg, pp 1–17
22. Minnen D, Starner T, Ward JA, Lukowicz P, Tröster G (2005) Recognizing and discovering human actions from on-body sensor data. IEEE Int Conf Multimed Expo, ICME 2005:1545–1548
23. Kurban OC, Yildirim T (2019) Daily motion recognition system by a triaxial accelerometer usable in different positions. IEEE Sens J 19(17):7543–7552
24. Khan AM, Lee YK, Lee SY, Kim TS (2010) A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. IEEE Trans Inf Technol Biomed 14(5):1166–1172
25. Maurer U, Smailagic A, Siewiorek DP, Deisher M (2006) Activity recognition and monitoring using multiple sensors on different body positions. In: International workshop on wearable and implantable body sensor networks (BSN'06), p 4
26. Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ (2012) Classification accuracies of physical activities using smartphone motion sensors. J Med Internet Res 14(5):e2208
27. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL (2013) Training computationally efficient smartphone–based human activity recognition models. In: Artificial neural networks and machine learning–ICANN 2013: 23rd International conference on artificial neural networks Sofia, Bulgaria. Proceedings 23 Springer Berlin Heidelberg. pp 426–433
28. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL (2013) A public domain dataset for human activity recognition using smartphones. In Esann 3:3

29. Lara ÓD, Prez AJ, Labrador MA, Posada JD (2012) Centinela: a human activity recognition system based on acceleration and vital sign data. Pervasive Mob Comput 8(5):717–729
30. Lee YS, Cho SB (2014) Activity recognition with android phone using mixture-of-experts co-trained with labeled and unlabeled data. Neurocomputing 126:106–115
31. Chen Z, Xiang S, Ding J, Li X (2020) Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori. IEEE Trans Instrum Meas 69(7):3992–4001
32. Hassan MM, Uddin MZ, Mohamed A, Almogren A (2018) A robust human activity recognition system using smartphone sensors and deep learning. Futur Gener Comput Syst 81:307–313
33. Kwapisz JR, Weiss GM, Moore SA (2011) Activity recognition using cell phone accelerometers. ACM SIGKDD Explorations Newsl 12(2):74–82
34. Shoaib M, Scholten H, Havinga PJ (2013) Towards physical activity recognition using smartphone sensors. In: 2013 IEEE 10th international conference on ubiquitous intelligence and computing and 2013 IEEE 10th international conference on autonomic and trusted computing. pp 80–87
35. Micucci D, Mobilio M, Napoletano P (2017) UniMiB SHAR: a dataset for human activity recognition using acceleration data from smartphones. Appl Sci 7:1101
36. Stisen A, Blunck H, Bhattacharya S, Prentow TS, Kjærgaard MB, Dey A, Sonne T, Jensen MM (2015) Smart devices are different: a SciTePress ssessing and mitigatingmobile sensing heterogeneities for activity recognition. In: Proceedings of the 13th ACM conference on embedded networked sensor systems. pp 127–140
37 Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJM (2016) Complex human activity recognition using smartphone and wrist-worn motion sensors. Sensors 16(4):426
38. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJM (2014) Fusion of smartphone motion sensors for physical activity recognition. Sensors 14(6):10146–10176
39. Vavoulas G, Chatzaki C, Malliotakis T, Pediaditis M, Tsiknakis M (2016) The mobiact dataset: Recognition of activities of daily living using smartphones. In: International conference on information and communication technologies for ageing well and e-health. SciTePress 2:143–151.
40. Malekzadeh M, Clegg RG, Cavallaro A, Haddadi H (2018l) Protecting sensory data against sensitive inferences. In: Proceedings of the 1st workshop on privacy by design in distributed systems. pp 1–6
41. Sztyler T, Stuckenschmidt H (2016) On-body localization of wearable devices: an investigation of position-aware activity recognition. In: 2016 IEEE International conference on pervasive computing and communications, (PerCom) pp 1–9
42. Wen J, Wang Z (2016) Sensor-based adaptive activity recognition with dynamically available sensors. Neurocomputing 218:307–317
43. Wen J, Wang Z (2017) Learning general model for activity recognition with limited labelled data. Expert Syst Appl 74:19–28
44. Hassan MM, Huda S, Uddin MZ, Almogren A, Alrubaian M (2018) Human activity recognition from body sensor data using deep learning. J Med Syst 42(6):99
45. Cho H, Yoon S (2018) Divide and conquer-based 1D CNN human activity recognition using test data sharpening. Sensors 18(4):1055
46. Yurtman A, Barshan B (2017) Activity recognition invariant to sensor orientation with wearable motion sensors. Sensors 17(8):1838
47. Zhao Y, Yang R, Chevalier G, Xu X, Zhang Z (2018) Deep residual bidir-LSTM for human activity recognition using wearable sensors. Math Probl Eng 2018:1–13
48. Wan S, Qi L, Xu X, Tong C, Gu Z (2019) Deep learning models for real-time human activity recognition with smartphones. Mobile Networks Appl 25(2):743–755
49. Yao L, Nie F, Sheng QZ, Gu T, Li X, Wang S (2016) Learning from less for better: semi-supervised activity recognition via shared structure discovery. In: Proceedings of the 2016 ACM International joint conference on pervasive and ubiquitous computing. pp 13–24
50. Challa SK, Kumar A, Semwal VB (2021) A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. Vis Comput 2021:1–15
51. Tang Y, Teng Q, Zhang L, Min F, He J (2021) Layer-wise training convolutional neural networks with smaller filters for human activity recognition using wearable sensors. IEEE Sens J 21(1):581–592
52. Li F, Shirahama K, Nisar M, Köping L, Grzegorzek M (2018) Comparison of feature learning methods for human activity recognition using wearable sensors. Sensors 18(3):679
53. Khan MAAH, Roy N (2018l) Untran: recognizing unseen activities with unlabeled data using transfer learning. In: 2018 IEEE/ACM Third international conference on internet-of-things design and implementation (IoTDI). pp 37–47
54. Chen Z, Jiang C, Xie L (2019) A novel ensemble ELM for human activity recognition using smartphone sensors. IEEE Trans Industr Inform 15(5):2691–2699

55. Tan M, Le Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning PMLR. pp 6105–6114
56. Radosavovic I, Kosaraju RP, Girshick R, He K, Dollár P (2020) Designing network design spaces. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 10428–10436
57. Howard A, Sandler M, Chu G, Chen LC, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, Le QV, Adam H (2019) Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. pp 1314–1324
58. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) Mobilenetv2: inverted residuals and linear bottlenecks. In: Proc IEEE Conf Comput Vis Pattern Recognit pp 4510–4520
59. Lin Y et al (2021) ATCSpeechNet: A multilingual end-to-end speech recognition framework for air traffic control systems. Appl Soft Comput 112:107847
60. Dong, L, Xu S, Xu B (2018) Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 5884–5888
61. Du C, Wang J, Sun H, Qi Q, Liao J (2021) Syntax-type-aware graph convolutional networks for natural language understanding. Appl Soft Comput 102:107080
62. Pellicer LFAO, Ferreira TM, Costa AHR (2023) Data augmentation techniques in natural language processing. Appl Soft Comput 132:109803
63. Akkem Y, Biswas SK, Varanasi A (2023) Smart farming monitoring using ML and MLOps. In: International conference on innovative computing and communication. Springer Nature Singapore, pp 665–675
64. Akkem Y, Biswas SK, Varanasi A (2023) Smart farming using artificial intelligence: a review. Eng Appl Artif Intell 120:105899
65. Cortes C, Mohri M, Syed U (2014) Deep boosting. In: International conference on machine learning. PMLR, pp 1179–1187
66. Kuznetsov V, Mohri M, Syed U (2014) Multi-class deep boosting. Adv Neural Inf Process Syst, 27
67. Han S, Meng Z, Khan AS, Tong Y (2016) Incremental boosting convolutional neural network for facial action unit recognition. Adv Neural Inf Process Syst, 29
68. Zhang W, Jiang J, Shao Y, Cui B (2020) Snapshot boosting: a fast ensemble framework for deep neural networks. SCIENCE CHINA Inf Sci 63(1):1–12
69. Welchowski T, Schmid M (2016) A framework for parameter estimation and model selection in kernel deep stacking networks. Artif Intell Med 70:31–40
70. Sun C, Ma M, Zhao Z, Chen X (2018) Sparse deep stacking network for fault diagnosis of motor. IEEE Trans Industr Inform 14(7):3261–3270
71. Low CY, Park J, Teoh ABJ (2020) Stacking-based deep neural network: Deep analytic network for pattern classification. IEEE Trans Cybern 50(12):5021–5034
72. Zhang L et al (2021) Nonlinear regression via deep negative correlation learning. IEEE Trans Pattern Anal Mach Intell 43(3):982–998
73. Tan TH, Wu JY, Liu SH, Gochoo M (2022) Human activity recognition using an ensemble learning algorithm with smartphone sensor data. Electronics (Basel) 11(3):322
74. Hasegawa T, Kondo K (2022) Easy ensemble: simple deep ensemble learning for sensor-based human activity recognition. IEEE Internet of Things J 10(6):5506–5518
75. Zhu R et al (2019) Efficient human activity recognition solving the confusing activities via deep ensemble learning. IEEE Access 7:75490–75499
76. Ghosal S, Sarkar M, Sarkar R (2022) NoFED-Net: Nonlinear fuzzy ensemble of deep neural networks for human activity recognition. IEEE Internet Things J 9(18):17526–17535
77. Zehra N, Azeem SH, Farhan M (2021) Human activity recognition through ensemble learning of multiple convolutional neural networks. In: 2021 55th Annual conference on information sciences and systems (CISS). IEEE, pp 1–5
78. Reiss A, Stricker D (2012) Creating and benchmarking a new dataset for physical activity monitoring. In: Proceedings of the 5th international conference on pervasive technologies related to assistive environments. pp 1–8
79. Guo H, Chen L, Peng L, Chen G (2016) Wearable sensor based multimodal human activity recognition exploiting the diversity of classifier ensemble. In: Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing. pp 1112–112
80. Garcia KD et al (2021) An ensemble of autonomous auto-encoders for human activity recognition. Neurocomputing 439:271–280
81. Bernaś M, Płaczek B, Lewandowski M (2022) Ensemble of RNN classifiers for activity detection using a smartphone and supporting nodes. Sensors 22(23):9451

82 Bhattacharya D, Sharma D, Kim W, Ijaz MF, Singh PK (2022) Ensem-HAR: An ensemble deep learning model for smartphone sensor-based human activity recognition for measurement of elderly health monitoring. Biosensors 12(6):393
83. Gholamiangonabadi D, Grolinger K (2023) Personalized models for human activity recognition with wearable sensors: deep neural networks and signal processing. Appl Intell 53(5):6041–6061
84. Khan SS, Taati B (2017) Detecting unseen falls from wearable devices using channel-wise ensemble of autoencoders. Expert Syst Appl 87:280–290