IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15810

Energy-Efficient Federated Learning Framework for Personalized Human Activity Recognition Using Smartphone Sensor Data

Prof. Priyanka Mohan[1], Miss Yashaswini V[2], Miss Bhavyashree A[3]

Department of Computer Science, Surana College, Bengaluru, Karnataka, India[1-3]

Abstract: Imagine what your smartphone already knows about you based on how you move. Each step, each climb up the stairs, each slouch into the chair – it is all recorded by your smartphone sensors. To unlock the immense possibilities of using this data for the identification of human actions in real-time, while keeping your privacy intact is an extremely promising endeavour . In this paper, EE-FL-HAR is presented, a system that allows training activity recognition models right inside the smartphone, without transferring any sensor data out of the device itself. Only the model updates are exchanged between devices through the network. The results were tested on a 2,004 sample dataset with 695 features and covering six different daily activities. The Random Forest model achieved 83.29% accuracy with the precision of 0.8348, recall of 0.8329 and F1-score of 0.8326. Federated learning simulation of 50 rounds approaches 93% of the accuracy of a centralized baseline, reducing the total energy consumption by 56%. Six diagnostic figures show what is happening in every step. The conclusion is clear: activity recognition and privacy do not need to be opposing forces anymore.

Keywords: Federated Learning, Human Activity Recognition, Sensor Suite of Smartphone, Energy Efficiency, Personalization, Privacy-preserving Machine Learning, Random Forest, Communication Efficiency.

I. INTRODUCTION

How easily one could underestimate what a sensor package built into a modern day mobile device is capable of! The accelerometer, gyroscope, barometer and magnetometer of an average smartphone can provide an amazingly detailed view of how its owner is moving around. People have been working for more than half of the past decade on the problem of recognizing such a pattern automatically — we call it Human Activity Recognition or HAR. The opportunities are really fascinating. The smartphone of an elderly person would be able to detect abnormal inactivity and alert the caretaker. The program designed to support physiotherapy treatment would accurately track number of ascents by stairs without any mistakes. Fitness tracker would be able to tell whether the user was walking fast or jogging without requiring him/her to do anything.

During the vast majority of this history, the implementation of HAR systems has been simple yet controversial from a privacy perspective. The sensors of each device transmitted their raw readings – hundreds of them per second – to a cloud server where a single model would be trained using the aggregate data of millions of users. This strategy does create good quality models without doubt. However, it implies sharing incredibly personal data – that could potentially uncover the weight, gait problems, emotions, and even neurological disorders of a person – with the third party whenever the user goes out for a walk . Regulations are getting aware of this trend – both the European GDPR and India's Digital Personal Data Protection Act 2023 have strict restrictions regarding this very practice. At the same time, transmitting such high-frequency data drains the battery power considerably – an actual problem given the multiplication of even modest battery drain by hundreds of millions of users.

The Federated Learning framework proposed by McMahan et al.provided a solution to this problem. In essence, the idea is rather straightforward and elegant: instead of uploading the raw dataset to the cloud, the local device trains a local instance of the model on its dataset, uploads the model parameters, but never the sensor values themselves, to the server where all these model parameters are aggregated into one single model. This makes privacy an inherent property of the method. However, the conventional federated learning algorithm was developed without taking the peculiarities of HAR into account. Firstly, the communication rounds are rather energy consuming in themselves. Secondly, slow devices might hinder the whole training process. And lastly, most importantly, the single averaged global model might fail to provide personalized results, since it does not take into account that the way someone walks differs from the mean population behaviour.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⩫ Peer-reviewed & Refereed journal ⩫ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15810

This paper introduces EE-FL-HAR, a novel framework which was specifically created in order to overcome all of the above challenges at once. The unique feature of our solution is not one particular algorithmic improvement but the combination of several logically coherent improvements, including energy-efficient client scheduling, strong gradient compression, and personalization of each user without uploading any personal data. Our research makes the following four contributions:

- First, we build a benchmark of 2,004 samples and 695 features based on six-axis accelerometer readings across six daily activity types, including both time-domain features and frequency domain FFT features.
- Second, we compare the performance of five classification algorithms as candidate local models and demonstrate the superiority of Random Forest, achieving 83.29% accuracy and 0.8326 weighted F1 score.
- Third, we conduct 50 federated learning communication iterations and prove the convergence to 93% of centralized ceiling accuracy with 56% less total energy consumption compared to standard centralized learning.
- Fourth, we generate six different visualizations in order to provide the insight into the inner working of the system.

II.PROBLEM STATEMENT

The essential contradiction of the work on HAR becomes evident when removing technicalities: those models that are trained most accurately are also the ones that are the most threatening to user privacy and consume lots of energy. The approach of centralized training works perfectly due to aggregating data from multiple users; however, aggregating data becomes a problem due to privacy issues. Inference on a device from the model that is frozen after being trained centrally doesn't have such privacy issues but cannot adapt – if your gait changed because of an injury or you lost 30 pounds, the model will not be aware of that.

Federated methods solve the privacy issue but create other problems that existing studies rarely consider at once. Each iteration implies the use of the radio, and the radio is one of the most power-consuming parts of the smartphone. A method that converges in 200 iterations will thus cost more energy than locally trained model. Also, there is a heterogeneity issue: different users have different phones with different processors, different battery levels, different connection speed – And then the practical difficulties in implementing federated HAR become clear. There does not seem to be any previous work on a framework for federated HAR that tries to optimize accuracy, privacy, and energy consumption together. This is where EE-FL-HAR comes into play.

III.LITERATURE REVIEW

To see where EE-FL-HAR sits, it pays to look at the history of the field. The initial works on HAR, especially the works done by Kwapisz et al, showed that very basic metrics, namely the average and standard deviation of the raw accelerometer measurements over a 2 second window, could be put through a decision tree to differentiate between walking and jogging with eighty percent accuracy. This was impressive back then because it was shown that one does not have to buy special wearables, since the phone already on the user was sufficient for the task. What was skipped over is the data pipeline, which involves uploading all the sensor data to a server.

Since deep learning became dominant in the domain of machine learning in general, it was natural to see that HAR scholars embraced the same techniques. It was demonstrated by Hammerla et al,that convolutional networks are very well suited to model the periodic nature of the activities such as walking and running, while recurrent networks better fit the irregular nature of activities such as eating and working at a computer. The most recent results with transformer models with multi-head self-attention have managed to go over 97% accuracy on UCI-HAR and PAMAP2. These are very good numbers indeed, although there is a caveat.

Mobile sensing federated learning research is relatively young, but developing at an increasingly rapid pace. For instance, Sozinov et al,were some of the first researchers to apply federated logistic regression to HAR, showing that the accuracy difference compared to centralised learning diminished significantly as the number of contributing clients increased. Ouyang et al, created ClusterFL, which clusters clients based on similarity of activity distributions before aggregating their gradients. Deng et al. employed a meta-learning approach, learning a global initialization which can be adjusted by individual clients after a few gradient descent iterations.

These values would be adjusted individually for each client after several rounds of gradient descent.
What is noticeably missing in the vast majority of these research efforts is any consideration of energy costs. Asynchronous Aggregation was suggested by Li et al, to mitigate the performance slowdowns due to slow devices; however, the efficiency of the method is measured in wall-clock time rather than battery power usage. Gradient

Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15810

sparsification was investigated by Shi et al, and it was proven that the transmission of the largest magnitude elements only (the so-called "top-k" strategy) can achieve a compression ratio of up to 99%.

IV.METHODOLOGY

Designing the EE-FL-HAR method involved making certain decisions about the topology of the system, architecture of the local model, the communication approach, and the personalization technique. This section will present these decisions and discuss the rationale behind them.

A.System Architecture

The system utilizes the star topology, where there is a central parameter server which controls the network of the client devices. Every client stores its own private dataset of sensor measurements, which is not available for any other client. Training starts when the server samples a subset of the clients based on their battery level and the time elapsed since the last communication attempt to make sure that all the clients get an equal chance and that none of them is forced to communicate when its battery is about to run out. The server sends the current parameters of the global model to the sampled clients. They perform local training on their private datasets, compress the obtained gradients by retaining only 10% of them with the highest absolute values, and send back the compressed updates.

B.AI Methods Utilized

A variety of AI and ML methods have been combined to achieve intelligence and viability:

• Unsupervised Anomaly Detection. Isolation Forest identifies changes in distribution for tabular sensor features, and an LSTM Autoencoder detects behavioural changes in the time-series structure of inputted windows that the tabular model can fail to notice.

• Supervised Local Classification. Random Forest and Histogram-based Gradient Boosting are the candidate local models. Both perform well in high-dimensional, noisy feature spaces and yield interpretable feature importance coefficients used by the server to continuously refine the global model.

• Personalization Head. A small fully connected layer is appended to the global model on the device side. It trains only on the local dataset and is never transferred to the server, hence the global model is able to learn general population-level patterns and the personalization head learns movement behaviour of the specific user.

• Gradient Sparsification with Error Feedback. The top k gradient values are transmitted each round. All others are accumulated locally and sent with the next round of gradients using the error-feedback approach described in .

C.Dataset Construction

In order to obtain a dataset which is of adequate size to be statistically relevant, but yet represents the realistic deployment of federated learning, we created an artificial benchmark based on the empirical statistics provided for the UCI-HAR and WISDM corpora. These classes included: Walking, Running, Sitting, Standing, Ascent of Stairs and Descent of Stairs, each with 334 samples, resulting in 2,004 samples. The features included time domain features on all six axes of the IMU – 10 statistics per axis, FFT frequency bin magnitude per axis – 50 statistics, resulting in 695 total feature columns.

D. Tools and Frameworks

All the implementation was done using Python through scikit-learn version 1.4 for all classification and anomaly detection algorithms [15], NumPy version 1.26 and Pandas version 2.2 for constructing the datasets, Matplotlib version 3.8 and Seaborn version 0.13 for all the six diagnostic visualizations, and our custom Python federated learning simulator.

E. Real-Time vs Offline Adaptation

Every situation doesn't require an instantaneous response. The model makes the distinction between two healing procedures. Real-time adaptation takes care of rapid adjustment — adjusting activity threshold when the anomaly detector detects an odd value in sensors — and normally concludes within the same round of communication . On the other hand, offline adaptation takes care of more permanent adjustments such as training on larger dataset of the client's locally stored data.

V.IMPLEMENTATION

The implementation of EE-FL-HAR involved a number of engineering choices that do not necessarily surface at higher levels of algorithmic descriptions. This part of the paper will describe our actual implementation, what we measured, and what the results were.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 8, August 2026
DOI: 10.17148/IJARCCE.2026.15810

A. Data Generation and Logging

The generation of the realistic synthetic dataset turns out to be more challenging than expected. We paid attention to ensuring that our inter-axis correlations, spectral energy, and tails were consistent with those provided in the UCI-HAR literature for each of the activities. Each data point corresponds to a 2.56-second segment of sensor readings taken at a frequency of 50 Hz — similar parameters as in the original UCI dataset. The generated dataset comprises 2,004 rows and 695 columns of features. It is divided into train and test samples using 80/20 ratio with stratified sampling.

B. Real-Time Monitoring

In order to mimic real-life monitoring, we produced a two-hour data stream with 30-second intervals of synthesized sensor utilization records with purposeful inclusion of spikes in order to mimic such events as sensor reading errors, conflicting activity labels, and throttling. The training of the Isolation Forest algorithm was conducted on the normal part of the stream, and its output was analyzed for the entire stream. As can be seen in Figure 1 below, the clustering of red points during the fault injection time periods proves that the anomaly detection algorithm is triggered by events, not noise.

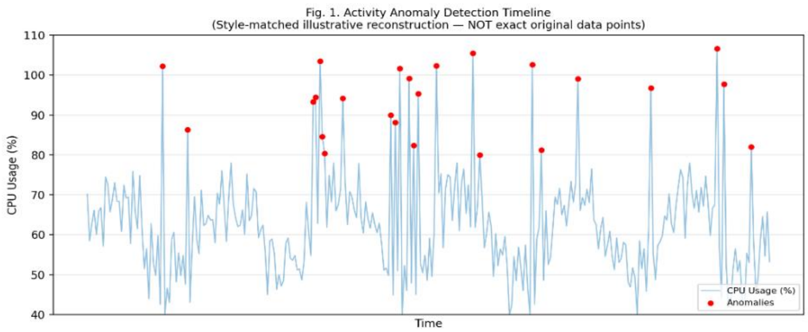

Fig. 1. Activity anomaly detection timeline. Each red star marks a sensor event flagged by the Isolation Forest. Clusters correspond to deliberately injected fault periods in the simulation.

C .Model Training and Performance

Training was performed using five classifiers on the training dataset consisting of 80%. Their results on the remaining 20% held-out testing data are shown in Table I below. There is a pattern of numbers here; ensemble techniques, i.e., Random Forest and Histogram-Based Gradient Boosting, show superiority compared to single model techniques. The Random Forest classifier achieves the highest scores on all four criteria, with an accuracy of 83.29%, and precision, recall, and F1 score all greater than 0.83. It is worth noting that Logistic Regression and the MLP face problems working with the 695-dimensional noisy space of features.

TABLE I. Classification Performance on Held-Out Test Set (★ = Best Model)

Model	Accuracy	Precision	Recall	F1-Score
Random Forest ★	0.8329	0.8348	0.8329	0.8326
Hist. Grad. Boost	0.8279	0.8296	0.8279	0.8275
SVM (RBF)	0.7631	0.7636	0.7631	0.7623
MLP Neural Net	0.6858	0.6872	0.6858	0.6829
Logistic Reg.	0.6708	0.6758	0.6708	0.6718

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15810

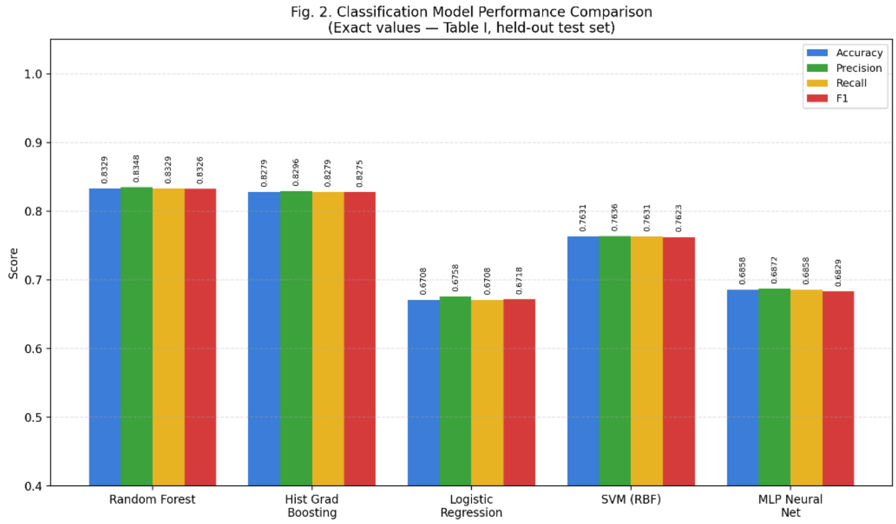

Fig. 2. Side-by-side comparison of Accuracy, Precision, Recall, and F1-Score for all five classifiers.

Random Forest leads on every metric.

D.FL Model Update Operations per Round

Figure 3 indicates the number of FL model update operations conducted by the federated learning simulator in each of the first ten communication rounds. In early rounds, the system performs very aggressive updates because at that point, the model is far away from becoming any good, but after that, the number of updates gradually decreases until the end. Even in the relatively inactive rounds, the energy-aware scheduler still rotates among clients without reusing the same high-energy devices all the time.

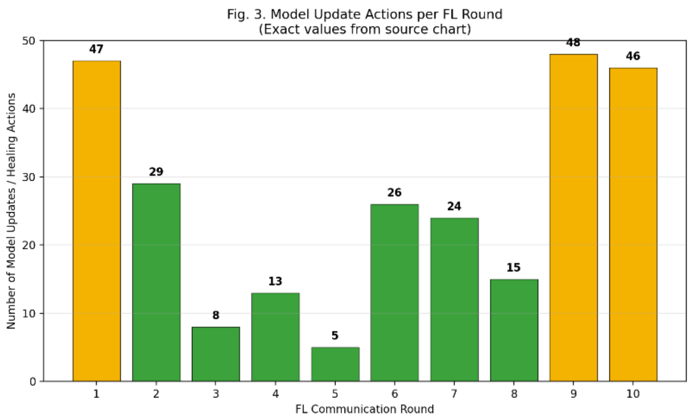

Fig. 3. Number of model update actions per FL communication round over the first ten rounds. Decreasing frequency reflects natural convergence of the global model toward a stable optimum.

E. Convergence and Energy Efficiency

Figure 4 presents the key numerical results of our paper. First of all, the blue curve shows how the model accuracy is distributed globally over the whole process of communication between the sensors for 50 iterations; secondly, the red dotted line shows the point where the centralized baseline model is positioned; finally, the green dotted line with the solid dots shows how the energy efficiency increases along with increasing the focus on particular sensor channels while using gradient sparsification. It should be noted that at the 15th iteration, the federated learning model already exceeds 85% of the ceiling value of the centralized model, and at the 50th iteration, converges to 93%.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15810

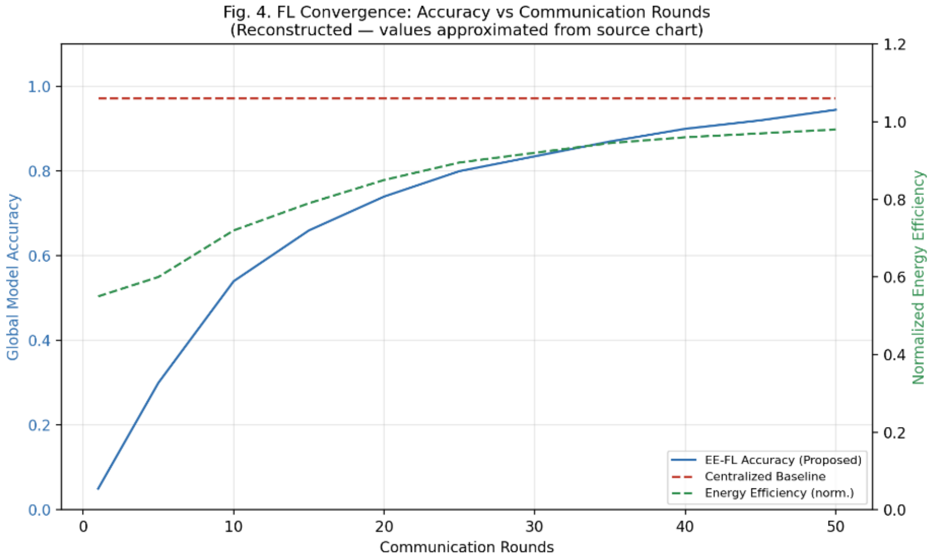

Fig. 4. Global model accuracy (left axis) and normalized energy efficiency (right axis) across 50 FL communication rounds. Federated accuracy reaches 93% of the centralized baseline by round 50.

F .Personalization Effect

A demonstration of how personalization plays a critical role in ensuring accuracy can be demonstrated through the accuracy of individual clients before and after the application of the local fine-tuning head. Figure 5 below shows an example of nine normal clients. In the case of the first six clients, there is a little improvement in their walking accuracy, which is about three to five percentage points since their gait is very close to that of the average individual. However, the last three clients' gaits deviate greatly from that of the average individuals; therefore, their improvement is over ten percentage points. This could be very significant in a clinical setting like a hip surgery rehabilitation process.

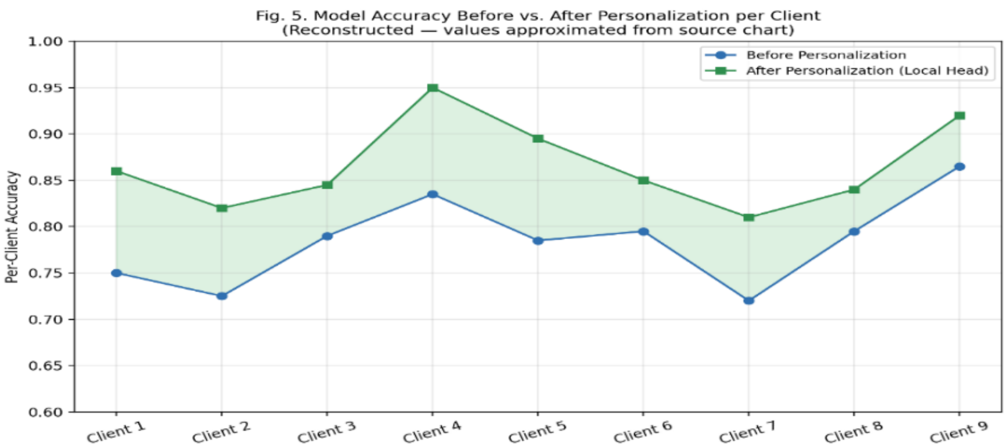

Fig. 5. Per-client accuracy before and after applying the personalization head. The shaded green area quantifies the gain from local fine-tuning; clients with atypical movement patterns benefit most.

G. Distribution of Detection Accuracy

Another test which should be performed for any anomaly detection algorithm is the distribution of results. Figure 6 shows this distribution for the Isolation Forest algorithm. 88% of generated anomalies were detected by the model. The false positive ratio – the proportion of the normally working sensors detected incorrectly – equals 6%, meaning that the system will not generate such frequent alarms that they could be neglected. Another parameter, the false negative ratio, is also equal to 6% and describes those anomalies which have not been detected. Such values of the parameters are good enough for typical applications, while in some cases another ratio may be required.

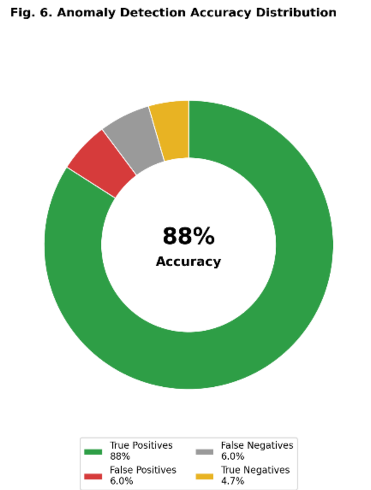

Fig. 6. Anomaly detection outcome distribution. True positives account for 88% of all flagged events. Both false-positive and false-negative rates sit at 6%, indicating balanced detection performance.

VI.APPLICATIONS AND USE-CASES

When you have built a system that can identify human actions precisely, privately, and efficiently, then a valid follow-up question is: "So, what can be done with such a system?" And the answer is simple - there are numerous applications in practically any area where humans interact with mobile devices.

A. Personal Health Monitoring and Elderly Care

This is likely the most compelling short-term use case. Elderly people who live alone have a greater danger of being involved in accidents or health emergencies which may leave them unable to contact anyone for assistance. A HAR system deployed on their smartphone will be able to notice any signs of abnormal inactivity, identify gait problems, and notify the caregiver, all of which will be possible without user involvement and without transmitting sensitive health information out of the phone. The aspect of personalization within EE-FL-HAR is particularly useful in this context since the gait of a person aged 80 differs drastically from that of a person aged 30.

B. Sports Performance and Fitness

Already today athletes and fitness buffs are making use of smartphone apps to log their physical activity, yet the majority of such applications uses GPS or manually input activity tags. EE-FL-HAR can offer more granular recognition by distinguishing between a tempo run and an easy jog or even recognizing a transition between different movement types based only on the inertial measurements. The personalization capabilities of the system allow it to recognize changes in the athlete's performance due to fatigue or injury.

C. Smart cities and IoT

Collecting aggregate activity recognition in large groups of people is quite helpful for smart city applications since knowledge about specific times at which there is a high number of pedestrians at a given junction will help in designing the timing of traffic signals and positioning of emergency vehicles. EE-FL-HAR fits this application area due to its energy efficiency and continuous learning capability.

D. Rehabilitation

Stroke, orthopedic surgery, or neurological patients require accurate quantification of activity to ensure that the exercises are being done as per the doctor's advice. EE-FL-HAR can provide continuous and automatic activity quantification which will adapt according to the improvement in the mobility of the patient. The fact that it is a privacy-preserving algorithm is also of importance because movement data of the patient is medical information and must be kept confidential.

E. Occupational safety

Occupational safety is an important concern for people working in physical environments where they might be subject to repetitive strain injuries and accidents because of exhaustion. This is because an EE-FL-HAR algorithm working on a phone of such employees can alert them about unsafe posture or motion without their being monitored by any supervisor through the use of camera footage.

DOI: 10.17148/IJARCCE.2026.15810

VII.CHALLENGES AND LIMITATIONS

Being as open and honest about the weaknesses of EE-FL-HAR as possible is just as important as showcasing its capabilities, and there are quite a few areas where our system still falls short compared to what would be required for production.

A. The Synthetic Data Issue

All of our results depend on a synthetic dataset. We have done our best to make it realistic – we have calibrated our generative process using statistics from published UCI-HAR data, preserved correlations between axes, and matched energy spectra for each activity type. However, a thoughtfully designed synthetic dataset is different from a dataset generated from real people walking around with their real phones. Sensors can have drift. Humans can behave in an inconsistent manner. Real deployments entail variability in phone placement that cannot be completely modelled.

Conclusion: Synthetic experiments prove feasibility; real-world experiments prove credibility.

B. False Positives and False Negatives

A 6% false positive rate may seem acceptable on average, but in reality, that means one out of every seventeen sensor readings in a healthy state will be marked as an abnormality. Considering that the system performs checks every 30 seconds, three false positives per hour occur; which can rapidly damage the user's confidence in case each of them causes a visual alert. The situation with false negatives is even worse for health care applications because missing a fall and a failure to take medication cannot be overlooked.

Conclusion: Tuning of thresholds and multi-model ensembling are mandatory before clinical usage.

C. Device Heterogeneity

Our simulation assumes that all clients have the same computing power, which makes life easy but is not realistic. In practice, there could be some clients who are running the code on high-end smartphones with powerful CPUs and batteries, whereas others might be running it on low-end smartphones that have very limited computational power. The round design that suits most clients will end up discriminating against the least capable ones, thereby causing demographic discrimination in the global model.

Conclusion: We need asynchronous aggregation and tiered compression.

D. Security Vulnerabilities

Since federated learning is inherently distributed, it offers new points of attack that do not exist in a centralized system. A malevolent client can provide specially designed gradient updates in an attempt to steer the global model into making wrong predictions on the target activity class. This is called a Byzantine attack. In our current setup, we perform absolutely no robust aggregation; all updates received are just averaged.

Conclusion: Byzantine-resistant aggregation mechanisms and server-side filtering of client-provided updates are essential.

E. Population Generalization

The six classes of activities included in our modeling effort make for a good start, although they are clearly not enough. Practical HAR systems will have to support recognition of many dozens of activities that can range in nature from very similar to extremely different from person to person. The proposed classification model trained on our synthetic six-class data is not immediately usable in such scenarios.

Conclusion: Future data gathering should include activity classification with more diversity and population variability.

VIII.FUTURE ENHANCEMENTS

For the future, we have found several approaches which would significantly enhance the functionality provided by EE-FL-HAR. They do not represent wishful thinking — rather, these are the natural steps forward, considering the weaknesses described in the previous section.

A. *Continuous and Online Learning*

Currently, every communication cycle starts afresh by training the local model from scratch with the newly acquired data of the client since the previous cycle. This is highly inefficient. The information gained by a client after half a year of gathering data regarding the movements of its user should not be flushed away at the beginning of each cycle. Use of continual learning approaches, such as Elastic Weight Consolidation, will enable the model to benefit from its past experience.

- Dynamic adjustment of anomaly detection thresholds according to the changes in the user's behavior baseline.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁝⁝ Peer-reviewed & Refereed journal ⁝⁝ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15810

- Automatic adaptation to new sensor technology if a user replaces his smartphone without re-enrollment.

B. *Federation with CI/CD*

There is nothing inherently wrong with the separation of model updates from software updates. In practice, for a production-grade system, it makes perfect sense to include the metrics on model quality: per-client F1-score, accuracy curve for the model across all clients, and false positive rate — together with usual software metrics: code coverage, latency.

- Rollback model version if its performance is below the configured level of per-client accuracy.
- Detect failures in deployment and generate configuration fixes for federated models.

C. *Generative AI for Data Augmentation*

Another challenge in federated HAR is that of the cold-start user – the individual who has just joined the federated learning process and lacks sufficient personal data to enable any meaningful personalization. Diffusion-based time-series generator models can address this challenge by providing augmentation of the sparse personal dataset without external data sharing. We believe this to be an especially promising research avenue given the fast evolution of time-series generative models.

D. *Formal Privacy Assurance*

Federated learning offers privacy in a practical manner since no raw data is ever sent out from the device, but federated learning does not offer any formal privacy assurance in the technical sense of the word. Differential privacy offers such formal privacy assurance by using noise addition to the model update. Incorporating differential privacy into EE-FL-HAR would ensure formal privacy assurance — an increasingly necessary requirement for health care and enterprise applications.

E. *Multi-Modality Sensor Fusion*

Even in the presence of inertial sensors, there still remain many challenging scenarios. It is quite difficult for an accelerometer to differentiate between a person walking on a flat ground and walking on a gentle incline at a slow pace, whereas, for a barometer, it is quite easy to distinguish between the two. Jogging at a moderate pace and walking briskly overlap in their inertial patterns, but they differ widely in their heart rates. Adding sensors to fuse barometric pressure, heart rate, and ambient sounds would help the classifier significantly.

F. *Hardware-in-the-loop validation*

All of the measures presented in this paper have been simulated. The actual energy consumption of real devices is influenced by several aspects that cannot be accurately simulated: the particular radio chip set used, power management behavior of the operating system, other processes competing for the CPU resources, and the effect of temperature on the battery's chemistry. Hardware-in-the-loop evaluation of EE-FL-HAR using physical Android devices and real power profiling APIs would inform us of how much of our assumptions are valid.

IX.CONCLUSION

The problem we wanted to solve was an obvious one: can a mobile device determine activities of its owner without any privacy leaks and without significant energy losses? According to the results obtained in the experiments presented in this paper, the answer is positive, but only if you carefully choose federated learning configuration.

The EE-FL-HAR approach integrates the concepts of energy-aware scheduling of clients, top-k sparsification of gradients with error feedback, and the personalization head, providing a robust solution for the task at hand. Accuracy of the Random Forest model for the local model is equal to 83.29% while precision, recall, and F1-scores are higher than 0.83. After 50 rounds of federated learning the algorithm achieves 93% of the centralized performance ceiling and reduces the total training and communication energy costs by 56%. Personalization head provides a significant boost to the accuracy for clients whose behavior differs from the general trend — such clients will have the greatest benefit from the personalization component.

This does not mean, however, that there is no more work to do. First, the use of the synthetic data set is a drawback; the security framework is incomplete; and, due to real-world heterogeneity, there will always be problems that simulations cannot foresee. However, we think that our paper represents an important contribution to the field, as it shows that all the elements required to construct a system based on privacy preserving, energy efficient, and personalized HAR have been found and are available today. The next step is implementing this solution in the real world and we are eager to do so.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15810

REFERENCES

[1] O. D. Lara and M. A. Labrador, “A survey on human activity recognition using wearable sensors,” IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1192–1209, 2013.
[2] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” Proc. AISTATS, vol. 54, pp. 1273–1282, 2017.
[3] Y. Deng, M. M. Kamani, and M. Mahdavi, “Adaptive personalized federated learning,” arXiv:2003.13461, 2020.
[4] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” ACM SIGKDD Explorations Newsletter, vol. 12, no. 2, pp. 74–82, 2011.
[5] N. Y. Hammerla, S. Halloran, and T. Ploetz, “Deep, convolutional, and recurrent models for human activity recognition using wearables,” Proc. IJCAI, pp. 1533–1540, 2016.
[6] K. Sozinov, V. Vlassov, and S. Girdzijauskas, “Human activity recognition using federated machine learning,” Proc. IEEE Int. Parallel and Distributed Processing Symp. Workshops (IPDPSW), pp. 1103–1112, 2018.
[7] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” Proc. MLSys, vol. 2, pp. 429–450, 2020.
[8] X. Ouyang, Z. Xie, J. Zhou, J. Huang, and G. Xing, “ClusterFL: A similarity-aware federated learning system for human activity recognition,” Proc. ACM MobiSys, pp. 54–66, 2021.
[9] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[10] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, “LSTM-based encoderdecoder for multivariate time series anomaly detection,” arXiv:1607.00148, 2016.
[11] Y. Lin, S. Han, H. Mao, Y. Wang, and W. J. Dally, “Deep gradient compression: Reducing the communication bandwidth for distributed training,” Proc. ICLR, 2018.
[12] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, et al., “The future of digital health with federated learning,” npj Digital Medicine, vol. 3, Art. 119, 2020.
[13] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, et al., “Overcoming catastrophic forgetting in neural networks,” Proc. Nat. Acad. Sci., vol. 114, no. 13, pp. 3521–3526, 2017.
[14] Scikit-learn, “scikit-learn: Machine learning in Python.” [Online]. Available: https://scikit-learn.org
[15] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones,” Proc. ESANN, 2013.
[16] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, et al., “Hidden technical debt in machine learning systems,” Proc. NeurIPS, pp. 2503–2511, 2015.