International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(22s) 2026 pp. 377-387
https://doi.org/10.70917/ijcisim-2026-5437
Article

A Lightweight CNN-GRU Model for Human Activity Recognition with Efficient Edge Deployment Using TFLite

Ganesha H S[1], Harsha K G[2], Roshini P[3], Shabana Sultana[4], Sharada Guptha M N[5], Manu Y M[6]

[1]Department of Electronics & Communication Engineering, Vidyavardhaka College of Engineering, Mysuru 570002, Karnataka, India
[2]Department of Computer Science & Engineering, Vidyavardhaka College of Engineering, Mysuru 570002, Karnataka, India
[3]Department of Computer Science & Engineering (Data Science), Maharaja Institute of Technology Mysore 571477, Karnataka, India
[4]Department of Computer Science & Engineering, The National Institute of Engineering, Mysore- 570008, Karnataka, India
[5]Department of Electronics & Communication Engineering, Sri Siddhartha Institute of Technology, Sri Siddhartha Academy of Higher Education, Tumakuru, Karnataka, India
[6]Department of Computer Science & Engineering, BGS Institute of Technology, Adichunchanagiri University, B. G. Nagara
ganesha.harve@gmail.com, harshakg.official@gmail.com, roshini.prem@gmail.com, shabana@nie.ac.in
sharadagupthamn@ssit.edu.in, manugowdaym3@gmail.com

Abstract: Wearable sensor-based human activity recognition (HAR) has become increasingly popular for applications in health monitoring, fitness, and smart living. But the use of deep learning models on edge devices is still challenging due to limited memory and computational power. In this paper, we develop a resource-constrained CNN-GRU hybrid model for HAR on the WISDM dataset. This architecture uses convolutional layers for spatial learning and gated recurrent units (GRU) for sequence learning. For deployment on the edge, the model is quantized to TensorFlow Lite (TFLite) using float16. Our experiments show that the model achieves an accuracy of 94.07%, while the size of the model is substantially smaller and suitable for deployment on edge devices. Importantly, the TFLite model maintains the same accuracy as the original model, ensuring its suitability for real-time deployment. The extensive assessment through confusion matrices, ROC curves and classification metrics confirms the effectiveness of the model across various activities. The proposed approach offers a balance between accuracy and computational efficiency, enabling real-time HAR on edge devices.
Keywords: human activity recognition, deep learning, lightweight model, hybrid model

1. INTRODUCTION

Human Activity Recognition (HAR) is becoming an important research focus because it can be applied in many areas such as health monitoring, rehabilitation, fitness monitoring, security and human computer interaction systems [1, 2]. The advancements in wearable technologies and embedded sensors have allowed the collection of large amounts of physiological and motion data in real time through technologies such as inertial measurement units (IMUs), accelerometers, electrocardiography (ECG), and other biosensors [1]. This has allowed real-time tracking of human activities, allowing better lifestyle monitoring, early diagnosis of medical problems and better user-friendly intelligent systems [3,4].

HAR has historically been tackled as a supervised learning task for multivariate time-series classification. Traditional machine learning methods, including support vector machines (SVMs), decision trees, and k-nearest neighbors (KNN), heavily depend on manual feature engineering, requiring expert knowledge, and potentially restricting model adaptation [1]. By contrast, deep learning (DL) models, such as convolutional neural networks

(CNNs) and recurrent neural networks (RNNs) such as long short-term memory (LSTM) have been shown to perform better by learning hierarchical and temporal features from raw data without relying solely on manually extracted features [1]. These approaches reduce the dependency on manual feature engineering and improve classification accuracy across diverse HAR tasks.

However, most state-of-the-art deep learning models for HAR are high-computational and have many parameters to learn, making them unsuitable for edge devices with limited resources [5]. Additionally, the limited availability of labeled sensor datasets and the abstract nature of sensor signals often lead to overfitting in complex models, thereby affecting their generalization capability [4, 6]. This presents a major challenge in designing fast and accurate models. Hence, compact models with smaller sizes and faster inference times are crucial for their practical usage in real-time applications, especially for wearable and edge computing devices [6, 7].

Moreover, as the use of smart wearable devices and smartphones with built-in sensors becomes more common, it is becoming increasingly important to perform activity recognition locally, rather than exclusively in the cloud [3]. Sending raw sensor data to the cloud for processing results in high latency, energy consumption, and network traffic. Local inference on the edge devices can reduce the communication cost, latency, and privacy concerns [3,7]. This calls for lightweight HAR models that can operate with limited computational resources and memory. Here, we propose a lightweight hybrid CNN-GRU model to perform human activity recognition from a wearable sensor dataset - the WISDM. The proposed model uses convolutional layers for spatial feature extraction and gated recurrent units (GRU) for temporal feature extraction from time-series data. For efficient inference on edge devices, we convert the model to TensorFlow Lite (TFLite) format with float16 optimization, reducing its size without compromising accuracy. Through experimental evaluation, we show that the proposed model offers high accuracy and efficiency, and can be used for processing real-time HAR data on edge devices.

The rest of this paper is structured as follows. Section II reviews the existing research on HAR and light-weight deep learning models. Section III provides an overview of the dataset, data preprocessing, and the model architecture. The results and evaluation are presented in Section IV. Finally, Section V provides a conclusion and suggestions for future work.

2. REVIEWS THE EXISTING RESEARCH

The emergence of wearable sensors and computing technologies has seen a significant focus on HAR. Initially, HAR focused on conventional machine learning approaches coupled with manual feature extraction. But these methods may not be robust enough to work effectively across different datasets and settings. Thanks to deep learning, it is now possible to automatically extract features from sensor data, leading to more accurate recognition. In-depth reviews in this field show that deep learning models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have emerged as the most popular due to their capacity to learn spatial and temporal features in sensor data [8], [9], [10].

A number of studies have investigated hybrid deep learning models to improve HAR accuracy by leveraging the capabilities of multiple models. For example, CNNs excel in capturing spatial patterns, while RNNs (such as LSTM and GRU) are adept at modeling temporal relationships in sequential data. Combining CNN and RNN-based models has shown promising results in complex HAR tasks [11], [12]. These models benefit from hierarchical feature extraction and temporal modeling, leading to improved accuracy over single models. But these models tend to have larger model sizes and increased computational costs.

While deep learning models have achieved remarkable results, their deployment on edge devices is challenging. Current state-of-the-art HAR models are often complex and demand high computational resources, which limits their deployment on edge devices. There have been recent efforts to develop lightweight and efficient models with high accuracy and low computational requirements. Lightweight models like TinyHAR have shown that models can be greatly reduced in size while maintaining accuracy [13], [14]. Model compression, pruning and quantization have also been extensively studied to facilitate deployment on edge and embedded devices [15].

Further, the incorporation of edge computing in HAR systems has become more popular because of its potential to minimise latency, bandwidth, and energy requirements. Efficient deep learning models for edge devices have demonstrated satisfactory performance in real-time activity recognition [16], [17]. These models are designed to balance performance and efficiency, allowing them to be deployed on resource constrained devices such as mobile phones, wearables and microcontrollers. However, issues such as lack of labeled data, inter-user variability and variability in the real world are still open challenges [8], [13]. To address these issues, this paper proposes the design of a lightweight CNN-GRU model with efficient deployment at the edge using TFLite, for improved performance and low computational complexity.

3. PROPOSED LIGHTWEIGHT MODEL ARCHITECTURE

This section presents the proposed methodology for human activity recognition using lightweight deep learning models. The overall workflow consists of dataset acquisition, data pre-processing, sequence generation, model design, training, and deployment. The complete pipeline of the proposed system is illustrated in Fig. 1.

A. System Overview

The proposed framework follows a structured pipeline starting from raw sensor data acquisition to edge deployment. Initially, raw accelerometer data is collected from the WISDM dataset, followed by preprocessing and segmentation into fixed-length sequences. These sequences are then fed into deep learning models for classification. Finally, the trained model is converted into a lightweight TensorFlow Lite (TFLite) format for efficient deployment on edge devices.

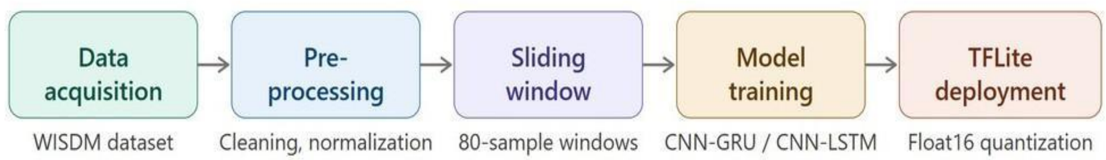

Fig. 1. Block diagram of the proposed HAR system (Data acquisition → Preprocessing → Segmentation → Model training→ TFLite deployment).

B. Dataset Description

The dataset used in this work is the publicly available WISDM (Wireless Sensor Data Mining) dataset, which contains accelerometer data collected from smartphone sensors [18]. The dataset includes tri-axial accelerometer readings (x, y, z) collected from multiple users performing daily activities such as walking, jogging, sitting, standing, upstairs, and downstairs listed in Table I.

TABLE I. DATASET DETAILS

Parameter	**Description**
Data Type	Time-series
Source	Smartphone-based sensing
Dataset Name	WISDM Dataset
Sensor Type	Accelerometer (3-axis)
Activities	6 (Walking, Jogging, Sitting, etc.)
Features	x, y, z acceleration

The raw WISDM dataset contains noise, missing values, and inconsistencies that must be addressed before model training. Initially, invalid records and missing entries are removed to ensure data integrity. Corrupted values, particularly in the z-axis where extraneous characters such as semicolons are present, are cleaned and converted into proper numerical format. Subsequently, all sensor readings are structured into a consistent numeric representation suitable for processing. The activity labels, originally in categorical form, are then transformed into integer-encoded values using label encoding techniques. This preprocessing pipeline ensures that the dataset is clean, consistent, and well-structured, thereby making it appropriate for effective training of deep learning models.

Since HAR is inherently a time-series classification problem, the continuous sensor data is segmented into fixed-length windows to capture temporal patterns effectively. In this work, a sliding window approach is adopted with a sequence length of 80 samples, where each segment consists of three sensor features corresponding to the x, y, and z axes. This results in an input shape of (N, 80, 3), where N represents the number of generated sequences. To assign labels to each segment, two different strategies are employed based on the model design. In the CNN-GRU (W1) approach, the label is taken from the last sample of each sequence, capturing the most recent activity state. In contrast, the CNN-LSTM (W2) approach uses a majority voting mechanism, where the most frequent label within the window is assigned to the entire segment, thereby improving robustness against noisy transitions. This segmentation process transforms raw continuous data into structured sequences suitable for deep learning models, as illustrated in Fig. 2.

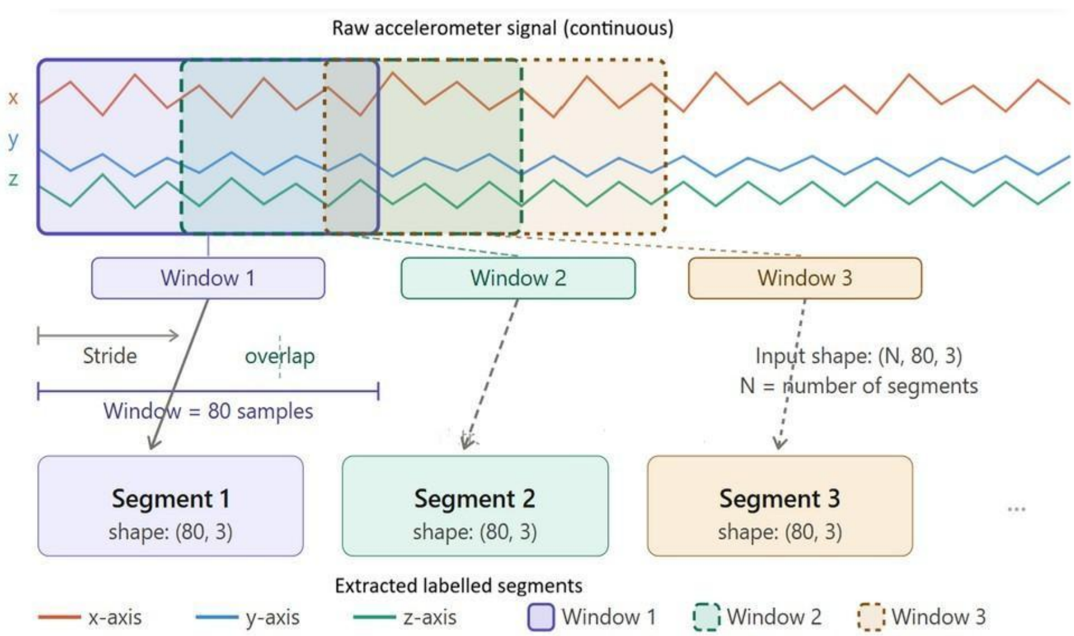

Fig. 2. Sliding window segmentation of time-series sensor data.

To improve convergence and model stability, normalization is applied:

$$Xnorm = (X - \mu) / \sigma \quad (1)$$

where μ and σ represent the mean and standard deviation of the training data.

C. Proposed Lightweight Model Architectures

In this work, two lightweight hybrid deep learning models are designed and comparatively analyzed for HAR. The first model, referred to as the CNN-GRU, integrates convolutional layers with a Gated Recurrent Unit (GRU) to effectively capture both spatial and temporal features from sensor data. Initially, a Conv1D layer with 64 filters and kernel size 5 is applied, followed by batch normalization and max pooling to extract low-level features and reduce dimensionality. This is followed by a second Conv1D layer with 128 filters, again combined with batch normalization and max pooling for deeper feature extraction. The temporal dependencies in the sequence are then learned using a GRU layer with 128 units. Finally, a fully connected dense layer with 128 neurons and dropout is used for regularization, followed by a softmax output layer for classification into six activity classes. The overall architecture of this model is illustrated in Fig. 3.

The second model, referred to as the CNN-LSTM, follows a similar structure but replaces the GRU layer with a Long Short-Term Memory (LSTM) unit for temporal modeling. In this architecture, the first Conv1D layer consists of 64 filters followed by batch normalization and max pooling. The second convolutional layer uses 96 filters, making it slightly lighter than the proposed model. After feature extraction, an LSTM layer with 64 units is employed to learn sequential dependencies. This is followed by a dense layer with 96 neurons and dropout, and finally a softmax output layer for multi-class classification. The detailed structure of this model is shown in Fig. 4.

Both models are trained under the same experimental settings to ensure a fair comparison. The sequence length is fixed at 80 samples, with a batch size of 128 and training conducted over 40 epochs. The Adam optimizer is used for efficient learning, while sparse categorical cross-entropy serves as the loss function. ReLU activation is applied in intermediate layers, and softmax is used in the output layer for classification. These hyperparameters are summarized in Table II.

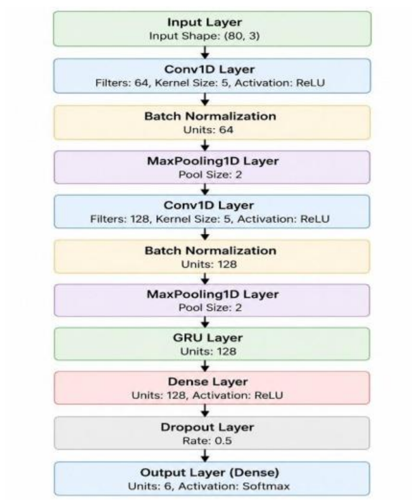

Fig. 3. CNN-GRU model architecture.

behavior, detection of overfitting, and comparison of learning efficiency between the two models.

TABLE II. MODEL HYPERPARAMETERS

Parameter	**Value**
Sequence Length	80
Batch Size	128
Epochs	40
Optimizer	Adam
Loss Function	Sparse Categorical CE
Activation	ReLU, Softmax

To facilitate deployment on resource-constrained edge devices, the trained deep learning models are converted into TensorFlow Lite (TFLite) format using float16 quantization. Initially, the trained Keras model is loaded, after which optimization techniques are applied using tf.lite.Optimize.DEFAULT to enhance efficiency. Float16 compression is then enabled to reduce the precision of model weights, thereby significantly decreasing the model size without causing substantial loss in accuracy. Finally, the optimized model is converted and saved in .tflite format. This conversion process ensures that the model becomes lightweight, faster, and suitable for real-time inference on edge and embedded systems while maintaining comparable performance to the original model.

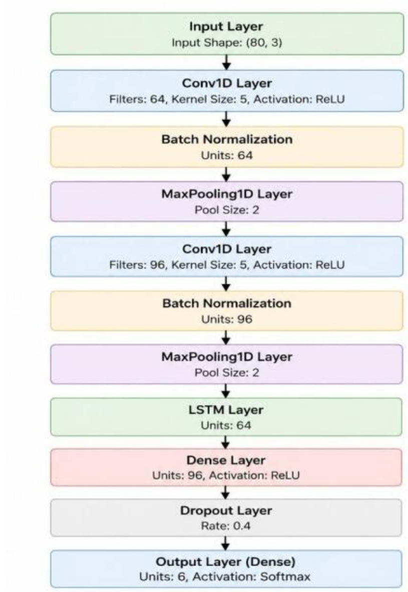

Fig. 4. CNN-LSTM model architecture.

The prepared dataset is divided into training and testing sets using an 80:20 ratio to ensure proper evaluation of model generalization. Both the CNN-GRU and CNN-LSTM models are trained for 40 epochs using the Adam optimizer, which provides efficient and adaptive learning during optimization. During training, model performance is continuously monitored using validation accuracy and validation loss computed on the test set. This approach enables tracking of convergence

4. RESULTS AND EVALUATION

The performance of the proposed models is evaluated using validation accuracy and loss obtained during training. The learning curves demonstrate that both models converge effectively within 40 epochs. The training and validation accuracy curves (Fig. 5) indicate a steady improvement in classification performance, with the CNN-LSTM model achieving slightly higher validation accuracy compared to the CNN-GRU model. Similarly, the training and validation loss curves (Fig. 6) show a consistent decrease in loss values, confirming stable learning behavior with minimal overfitting in both models.

To provide a comprehensive evaluation, multiple performance metrics are considered, including accuracy, precision, recall, F1-score, confusion matrix, and ROC analysis. The CNN-LSTM model achieves an overall accuracy of 96.17%, outperforming the CNN-GRU model, which achieves 94.07%. From the class-wise evaluation, CNN-GRU demonstrates superior performance across most activities, particularly in “Jogging,” “Walking,” and “Standing,” with F1- scores exceeding 0.97. In contrast, CNN-LSTM shows slightly lower performance, especially in challenging classes such as “Downstairs” and “Upstairs,” indicating difficulty in distinguishing similar motion patterns.

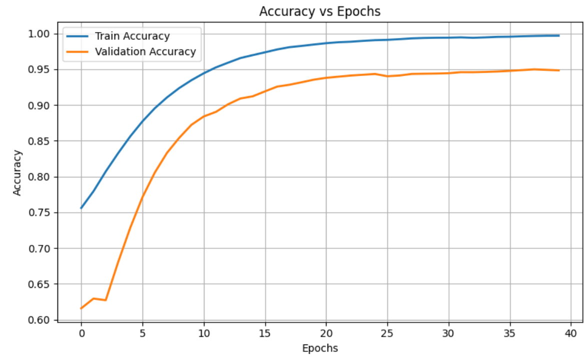

Fig. 5a. Training and validation accuracy – CNN-GRU.

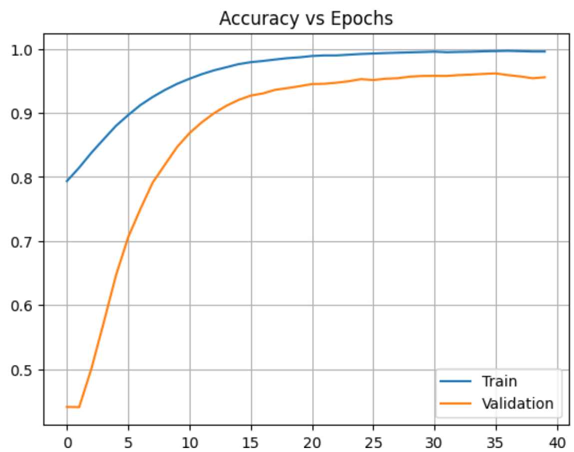

Fig. 5b. Training and validation accuracy – CNN-LSTM.

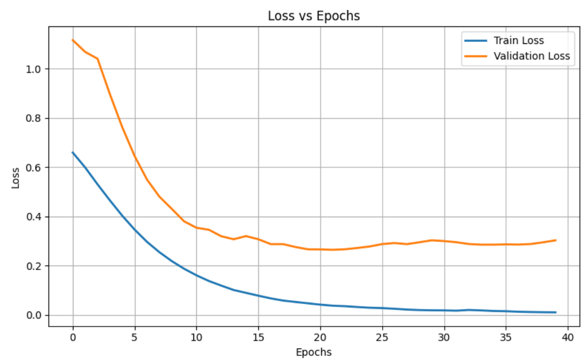

Fig. 6a. Training and validation loss – CNN-GRU.

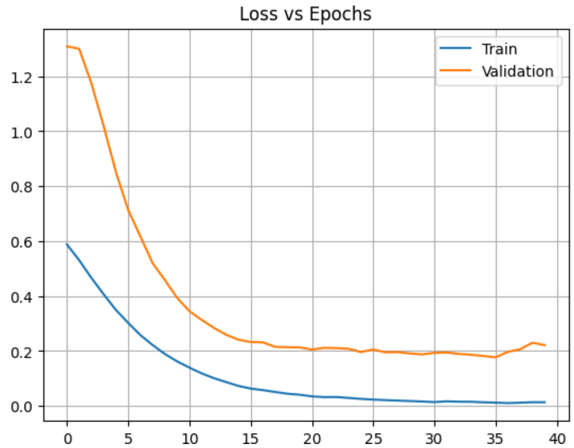

Fig. 6b. Training and validation loss – CNN-LSTM.

The confusion matrix of the proposed model (Fig. 7) highlights that most activities are correctly classified, with minor misclassifications occurring between similar activities such as "Upstairs" and "Downstairs." This behavior is common in sensor-based HAR due to overlapping motion characteristics. The ROC curve for multi-class classification (Fig. 8) further validates the effectiveness of both models. The curves for all activity classes are close to the top-left corner, indicating excellent discriminative capability. The Area Under Curve (AUC) values for both models are consistently high, ranging between 0.99 and 1.00, confirming strong classification performance.

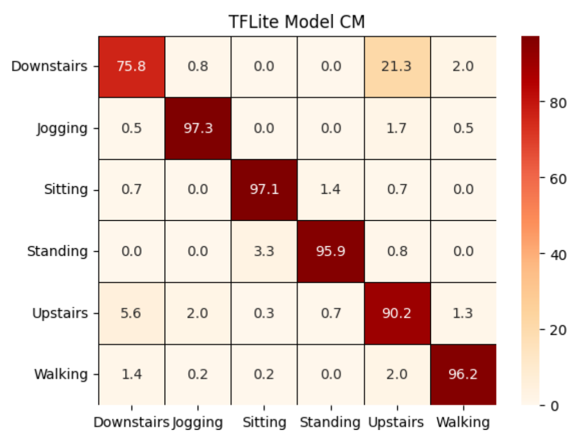

Fig. 7a. Confusion matrix – CNN-GRU (TFLite).

Overall, the results demonstrate that while both models are effective for human activity recognition, the CNN-LSTM model provides slightly better accuracy and robustness, whereas the CNN-GRU model offers a good trade-off between performance and computational efficiency, making it more suitable for lightweight edge deployment scenarios.

The results indicate that the CNN-LSTM model achieves higher classification accuracy compared to the CNN-GRU model, demonstrating better learning capability for temporal patterns. Additionally, TFLite conversion significantly reduces the model size for CNN-GRU, while CNN-LSTM shows a slight increase in size after conversion, provided in Table III. The CNN-LSTM model provides superior performance in terms of accuracy, whereas the CNN-GRU model is more efficient for edge deployment due to its smaller compressed size.

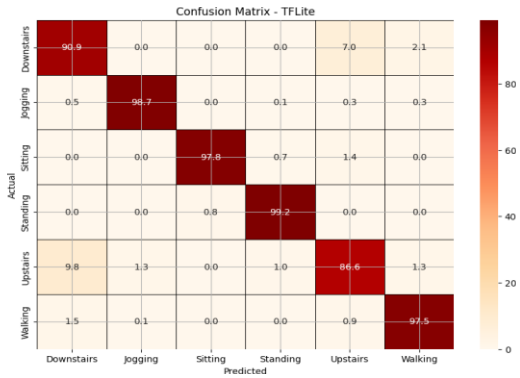

Fig. 7b. Confusion matrix – CNN-LSTM (TFLite).

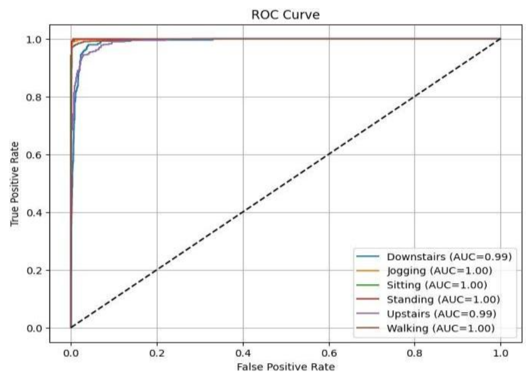

Fig. 8a. ROC curve for multi-class classification – CNN-GRU.

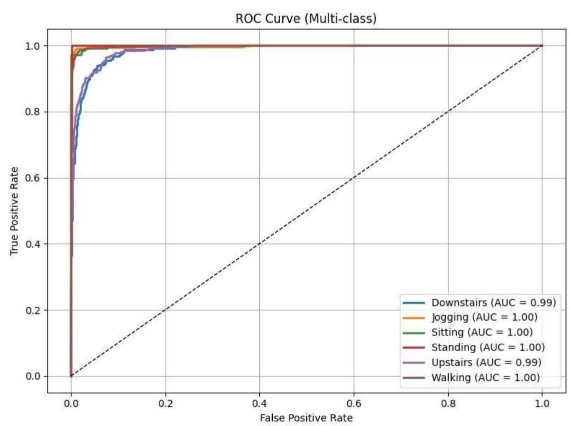

Fig. 8b. ROC curve for multi-class classification – CNN-LSTM.

TABLE III. MODEL EFFICIENCY COMPARISON

Model	**Accuracy (%)**	**Size (MB)**
CNN-GRU	94.07	1.874
TFLite	94.07	0.322
CNN-LSTM	96.17	0.429
TFLite	96.17	0.746

5. CONCLUSION AND SUGGESTIONS FOR FUTURE WORK

In this paper, we proposed a lightweight hybrid CNN-GRU model for HAR. The hybrid model integrates convolutional layers for feature extraction with gated recurrent units for capturing temporal dynamics in multivariate time-series data. The proposed model was evaluated on the WISDM dataset and achieved a classification accuracy of 94.07%, suggesting the model can effectively classify various human activities. This research has also developed an efficient way of deploying the trained model on edge devices by converting it to TFLite format with float16 precision. The experiments demonstrate that the TFLite model achieves the same accuracy as the original model, with a substantially smaller model size (0.322 MB), enabling real-time deployment on low-power devices. Detailed analysis, including confusion matrices, ROC curves and classification metrics, also confirms the effectiveness of the proposed method.

However, there are some limitations. The current research uses a single dataset and a random partition of the subjects, which may not cover all possible variability. Moreover, although the model is very small, its inference time and energy usage can be further improved to better suit applications on ultra-low-power edge devices.

There are a few avenues for future research building on this study. First, the model can be tested using Leave-One-Subject-Out cross-validation to improve understanding of the model's generalization ability. Second, benchmarking against baseline models, including CNN, LSTM, and other lightweight models can offer more insights into model performance. Finally, other model compression strategies such as pruning and quantization-aware training can be employed to achieve lower computational and memory demands.

In summary, the lightweight CNN-GRU model achieves a good balance of classification performance and computational efficiency, and is a feasible approach for real-time human activity recognition on edge devices.

REFERENCES

1. H. S. Ganesha, R. Gupta and S. H. Gupta, "Multi-Layer Heterogeneous Ensemble Model for Human Activity Recognition," 2021 7th International Conference on Signal Processing and Communication (ICSC), Noida, India, 2021, pp. 210-215, doi: 10.1109/ICSC53193.2021.9673341.
2. G. H. S, R. Gupta, S. H. Gupta and S. Rajan, "Deep Learning Ensemble for Recognising Lower Limb Activity," 2023 IEEE Sensors Applications Symposium (SAS), Ottawa, ON, Canada, 2023, pp. 1-5, doi: 10.1109/SAS58821.2023.10254188.
3. B. Adel, A. Badran, N. E. Elshami, A. Salah, A. Fathalla, M. Bekhit, "A Survey on Deep Learning Architectures in Human Activities Recognition Application in Sports Science, Healthcare, and Security," In Proc. ICR'22 Int. Conf. on Innovations in Computing Research, pp. 121–134, Aug. 2022, Springer, doi: 10.1007/978–3-031-14054-9_13.
4. H. S. Ganesha, R. Gupta, S. H. Gupta, and S. Rajan, "Few-shot transfer learning for wearable IMU-based human activity recognition," Neural Computing and Applications, vol. 36, no. 18, pp. 10811–10823, Jun. 2024, doi: 10.1007/s00521-024-09645-7.
5. P. Agarwal and M. Alam, "A lightweight deep learning model for human activity recognition on edge devices," Procedia Computer Science, vol. 167, pp. 2364–2373, 2020, doi:10.1016/j.procs.2020.03.289.
6. Ankita, S. Rani, H. Babbar, S. Coleman, A. Singh, and H. M. Aljahdali, "An efficient and lightweight deep learning model for human activity recognition using smartphones," Sensors, vol. 21, no. 11, p. 3845, 2021, doi: 10.3390/s21113845.
7. H. S. G., R. Gupta, and S. H. Gupta, "Energy efficient LoRa-based AIoT setup for human movement classification," Transactions on Emerging Telecommunications Technologies, vol. 36, p. e70107, 2025, doi: 10.1002/ett.70107.
8. J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," Pattern Recognition Letters, vol. 119, pp. 3–11, 2019, doi: 10.1016/j.patrec.2018.02.010. F. Gu, M.-H. Chung, M. Chignell, S. Valaee, B. Zhou, and X. Liu, "A survey on deep learning for human activity recognition," ACM Computing Surveys, vol. 54, no. 8, Art. no. 177, pp. 1–34, Nov. 2022, doi: 10.1145/3472290.
9. O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," in IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1192-1209, Third Quarter 2013, doi: 10.1109/SURV.2012.110112.00192.
10. N. Jaouedi, N. Boujnah, and M. S. Bouhlel, "A new hybrid deep learning model for human action recognition," Journal of King Saud University - Computer and Information Sciences, vol. 32, no. 4, pp. 447–453, 2020, doi: 10.1016/j.jksuci.2019.09.004.
11. Ankita, S. Rani, H. Babbar, S. Coleman, A. Singh, and H. M. Aljahdali, "An efficient and lightweight deep learning model for human activity recognition using smartphones," Sensors, vol. 21, no. 11, p. 3845, 2021, doi: 10.3390/s21113845.

12. Y. Zhou, H. Zhao, Y. Huang, T. Riedel, M. Hefenbrock, and M. Beigl, "TinyHAR: A lightweight deep learning model designed for human activity recognition," in Proceedings of the 2022 International Symposium on Wearable Computers (ISWC '22), Cambridge, United Kingdom, Sep. 2022, pp. 1–5, doi: 10.1145/3544794.3558467.
13. Y. L. Coelho, F. d. A. S. d. Santos, A. Frizera-Neto and T. F. Bastos-Filho, "A Lightweight Framework for Human Activity Recognition on Wearable Devices," in IEEE Sensors Journal, vol. 21, no. 21, pp. 24471-24481, 1 Nov.1, 2021, doi: 10.1109/JSEN.2021.3113908.
14. C. -H. Wang, K. -Y. Huang, Y. Yao, J. -C. Chen, H. -H. Shuai and W. -H. Cheng, "Lightweight Deep Learning: An Overview," in IEEE Consumer Electronics Magazine, vol. 13, no. 4, pp. 51-64, July 2024, doi: 10.1109/MCE.2022.3181759.
15. P. Agarwal and M. Alam, "A lightweight deep learning model for human activity recognition on edge devices," Procedia Computer Science, vol. 167, pp. 2364–2373, 2020, doi: 10.1016/j.procs.2020.03.289.
16. D. Ravi, C. Wong, B. Lo and G. -Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN), San Francisco, CA, USA, 2016, pp. 71-76, doi: 10.1109/BSN.2016.7516235.
17. Jennifer R. Kwapisz, Gary M. Weiss and Samuel A. Moore (2010). Activity Recognition using Cell Phone Accelerometers, Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10), Washington DC.