FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 36-40 | e-ISSN: 2347-470X

Enhanced Recognition of Human Activity using Hybrid Deep Learning Techniques

Abinaya S[1*], Rajasenbagam T[2], Indira K[3], Uttej Kumar K[4], and Potti Sai Pavan Guru Jayanth[5]

[1] *School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India; s.abinaya@vit.ac.in*
[2]*Department of Computer Science and Engineering, Government College of Technology, Coimbatore 641013,India; trajasenbagam@gct.ac.in*

[3] *Department of Computer Science and Engineering, Thiagarajar College of Engineering, Madurai 625015, India; kiit@tce.edu*

[4]*School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India; kandagatlauttej.kumar2019@vitstudent.ac.in*

[5]*School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India; sai.pavangurujayanth2019@vitstudent.ac.in*

***Correspondence:** Abinaya S; s.abinaya@vit.ac.in

ABSTRACT- In the domain of deep learning, Human Activity Recognition (HAR) models stand out, surpassing conventional methods. These cutting-edge models excel in autonomously extracting vital data features and managing complex sensor data. However, the evolving nature of HAR demands costly and frequent retraining due to subjects, sensors, and sampling rate variations. To address this challenge, we introduce Cross-Domain Activities Analysis (CDAA) combined with a clustering-based Gated Recurrent Unit (GRU) model. CDAA reimagines motion clusters, merging origin and destination movements while quantifying domain disparities. Expanding our horizons, we incorporate image datasets, leveraging Convolutional Neural Networks (CNNs). The innovative aspects of the proposed hybrid GRU_CNN model, showcasing its superiority in addressing specific challenges in human activity recognition, such as subject and sensor variations. This approach consistently achieves 98.5% accuracy across image, UCI-HAR, and PAMAP2 datasets. It excels in distinguishing activities with similar postures. Our research not only pushes boundaries but also reshapes the landscape of HAR, opening doors to innovative applications in healthcare, fitness tracking, and beyond.

Keywords: Human Activity Recognition (HAR); Convolutional Neural Networks (CNNs); Sensor Data Analysis, Activity Classification, Multi-Modal Data Fusion.

ARTICLE INFORMATION

Author(s): Abinaya S, Rajasenbagam T, Indira K, Uttej Kumar K, Potti Sai Pavan Guru Jayanth;
Received: 06/11/2023; **Accepted**: 13/01/2024; **Published**: 20/01/2024;
e-ISSN: 2347-470X;
Paper Id: IJEER 0611-10;
Citation: 10.37391/IJEER.120106
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-12/ijeer-120106.html

Publisher's Note: FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

In the exciting realm of computer vision, human activity recognition (HAR) is the star, with its applications spanning healthcare, surveillance, and sports. Deep learning has elevated HAR accuracy, and two key players, Gated Recurrent Units (GRUs) and Convolutional Neural Networks (CNNs), have taken the stage [2-5]. CNNs excel at image recognition and feature extraction, while GRUs are experts in sequential data, making them ideal for HAR [8]. The magic happens when these two join forces, producing promising results in recognizing human activities with precision.

In our study, we unveil a hybrid model fusing CNNs and GRUs to decipher human activities from both image and sensory datasets [4]. We've given the image dataset a makeover with data augmentation techniques, enhancing its resilience to various image variations. The sensory datasets, drawn from UCI HAR and PAMAP2, are stalwarts in the HAR domain. UCI HAR comprises accelerometer and gyroscope data from thirty subjects engaging in six activities, while PAMAP2 features data from nine subjects performing 18 diverse activities.

Our hybrid model is a three-step marvel. We prep the image dataset, extract high-level features with CNNs, and let GRUs unravel the temporal patterns within the sensory data. The goal is to prove the model's mettle in recognizing human activities using a multi-modal approach. These results have potential applications [7] in sports, healthcare, and surveillance. While CNNs and GRUs have made leaps in HAR accuracy, limitations persist [9]. The models can be sensitive to data fluctuations, impacting their real-time performance. Our contributions aim to address these challenges by introducing a novel hybrid GRU_CNN model, leveraging image pre-processing and Wasserstein-based clustering to enhance accuracy and adapt to evolving HAR landscapes [1]. Our research fills a vital gap by uniting the powers of GRUs and CNNs to transform human activity recognition, with applications in various domains. The model's ability to effectively oversee 3-S conflicts and cross-

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 36-40 | e-ISSN: 2347-470X

domain activities analysis, and incorporate image pre-processing techniques and Wasserstein clustering, has shown promising results in improving human activity recognition technology. Following are some of the major contributions:

(1) A unique hybrid GRU_CNN model was proposed for identification of human activities that combines image and sensory datasets to achieve high accuracy.
(2) Used Image Pre-processing techniques and Image Segmentation to enhance the effectiveness of the model by improving the quality and clarity of the image data.
(3) Incorporated Wasserstein-based clustering to cluster similar activities and enhance the accuracy of the model by reducing the impact of the 3-S conflicts.
(4) Created the CDAA approach to tackle the difficulty of training the model when the HAR community changes by hybrid GRU_CNN model that can effectively capture spatiotemporal and spatial features.
(5) Achieved high accuracy in detecting various activities, including walking, jogging, standing, and sitting that is experimented with UCI-HAR and PAMAP2, and image dataset.

2. MATERIALS AND METHODS

This section unveils the essence of our work: data preparation, the finesse of data preprocessing, and the core of the hybrid GRU_CNN model.

2.1 Overall Workflow

Figure 1 illustrates our human activity recognition method, driven by two datasets: Images and Sensors. Images go through enhancements like Contrast Limited Adaptive Histogram Equalization (CLAHE) and data augmentation, expanding dataset size and improving contrast. The Sensory dataset records movements using accelerometers and gyroscopes, and Wasserstein-based clustering groups similar activities [6]. This approach identifies akin activities with diverse movement patterns.

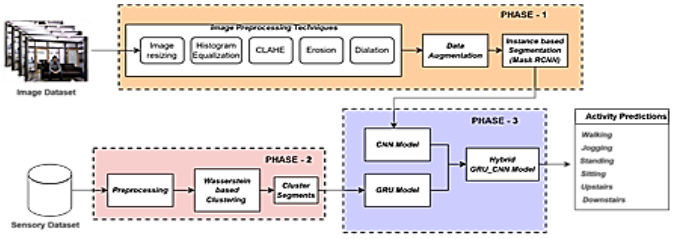


Figure 1. Overall Proposed hybrid GRU_CNN Workflow

2.2. Image Segmentation using Mask R-CNN (Phase-1)

2.2.1. Image Processing Techniques

In our pursuit of enhanced recognition, we employ a two-fold approach: image enhancement and data augmentation. For image enhancement, we convert images to grayscale and utilize Histogram Equalization and Contrast Limited Adaptive Histogram Equalization (CLAHE) techniques to redistribute intensity levels intelligently [3] as shown in the *figure 2*. This process accentuates crucial features. Simultaneously, data augmentation is applied to prevent overfitting during training. Employing geometric transformations, such as cropping, rotating, shifting, shearing, zooming, and flipping, we create additional images from each input, effectively doubling the standard class dataset, thereby enhancing its robustness.

(a) (b) (c)

Figure 2. (a) Original Image; (b) Applied Histogram Equalization Technique for original image; (c) Applying Contrast Limited Adaptive Histogram Equalization (CLAHE) Technique for original image.

2.2.2. Mask R-CNN-Based Segmentation

Mask R-CNN excels in image segmentation for human activity recognition. It uses a convolutional neural network (CNN) [10] to extract visual patterns and incorporates a Region Proposal Network (RPN) and a Mask Head. The RPN identifies Regions of Interest (RoIs), and the Mask Head refines bounding boxes and generates pixel-level masks. The model minimizes a combined loss function for precise localization and instance mask creation. The RoIAlign layer enhances accuracy by preventing alignment issues, and the network architecture comprises a backbone (ResNet or ResNeXt) and a network head that handles RoIs' classification and regression, with a focus on efficient convolutional mask prediction.

Figure 3. Sample results of Mask RCNN on Activities Images

2.3. Wasserstein-Based Clustering of Sensory Data (Phase-2)

2.3.1. Cross-Domain Activity Distance Estimation & Wasserstein Distance

Wasserstein distance measures the separation between probability distributions. In our context, it quantifies the dissimilarity between activities in different domains. Equation (1) outlines the distance estimation, with the Wasserstein distance (f) helping form activity clusters by grouping similar activities based on their distributional similarities.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 36-40 | e-ISSN: 2347-470X

$$f\left(D_{src,}D_{trg}\right)=\gamma\epsilon\pi\left(D_{src},D_{trg}\right)\inf\ E_{\left(t_{src,}t_{trg}\right)\sim\gamma}\left\|t_{src}-t_{trg}\right\| \quad (1)$$

where $\pi\left(D_{src,}D_{trg}\right)$ is a set of all joint distributions $\gamma\left(t_{src,}t_{trg}\right)$ whose marginal are D_{src} and D_{trg} respectively. The $\gamma\left(t_{src,}t_{trg}\right)$demonstrates the mass that must be transported from the originating to the destination domain in order to transfer the distribution of D_{src} into D_{trg} .

To measure the separation between probability distributions and quantify the mass required transporting the source domain's distribution into the target domain's distribution, we utilize the Wasserstein distance *equation (f).* This clustering technique groups activities based on their similarities and differences, ensuring that each cluster houses closely related yet distinct activities. *Figure 4* illustrates the process of calculating distances between cross-domain activities, aligning source domain activities $\left(O_n\right)$ with their adjacent target activities $\left(T_n\right)$ to form activity groups $C\left(C_1,C_2,\cdots,C_n\right)$ based on proximity.

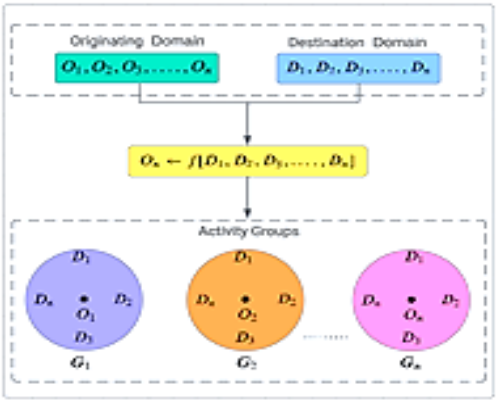


Figure 4. Measurement of cross-domain activities using a clustering technique

2.4. Hybrid GRU_CNN Model Architecture

The hybrid GRU_CNN model focuses on recognizing human activities that ingeniously combines image and sensory data as illustrated in Figure 5, is structured into three main components:

i. CNN Feature Extraction: The CNN module extracts feature from input image data via a deep convolutional neural network. This component effectively handles spatial variations in image data, making it ideal for feature extraction. The output is a feature map.

ii. GRU Temporal Modeling: The GRU module manages the temporal characteristics of sensory data by employing gated recurrent units (GRUs) to capture temporal relationships. It takes the feature map from the CNN module as input, generating a sequence of hidden states. The final hidden state embodies the temporal information.

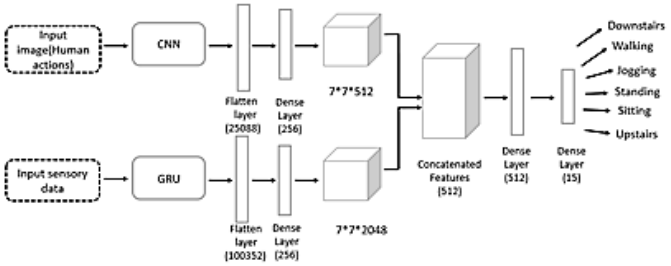


Figure 5. Hybrid GRU_CNN Model Architecture

iii. Fusion Module: This module combines the CNN output with the final GRU hidden state to produce the ultimate prediction. Initially, the GRU output is dimensionally reduced via a fully connected layer. This reduced vector is then merged with the compressed CNN feature map and traverses through a sequence of fully connected layers to generate the final forecast. The architecture employs separate loss functions for classification and regression, and the model is trained using the Adam optimizer with a decaying learning rate.

The general equation for a hybrid GRU_CNN model is expressed as in *eq. (2).*

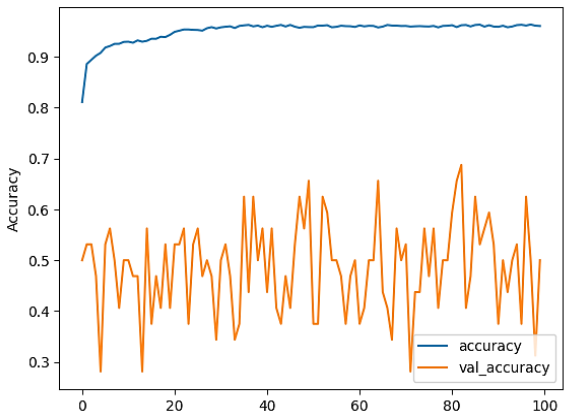
$$y_t = f\left(W_{cnn}x_t + U_{gru}h_{t-1} + b\right) \quad (2)$$

Where x_t represents the input at time step t, h_{t-1} is the previous GRU hidden state, W_{cnn} and U_{gru} are weight matrices, and b is the bias term. The function f denotes an activation function, such as the sigmoid. This hybrid model integrates the strengths of both CNN and GRU components to effectively model spatial and temporal features for activity recognition.

3. EXPERIMENTAL RESULTS

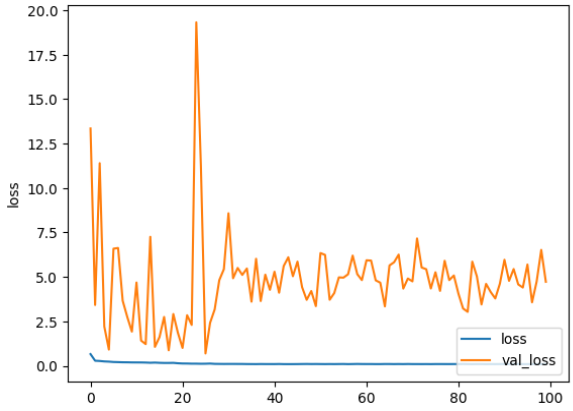
3.1. Hybrid GRU_CNN Algorithm Performance

This section displays the simulation results obtained utilizing the CNN-GRU algorithm (Fig. 6). Using grid search capability, the hyper parameters of the model are set to utilize the Adam optimization technique, the loss function employed is categorical cross-entropy using a batch size of 96 and over 100 epochs with a learning rate of 0.01 and momentum is 0.9.



(a)

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 36-40 | e-ISSN: 2347-470X



(b)

Figure 6. (a) Accuracy curve of proposed Model; (b) Loss curve of proposed Model

3.2. Performance on Image Dataset

Table 1 presents the results of applying CNN and VGG16 models to our image dataset, which is comprised of images representing activities carried out by humans. In order to get the highest level of accuracy possible, we use image pre-processing methods and image segmentation.

Table 1. Performance on Image Dataset

Model	Accuracy	Precision	Recall	F1-score
CNN	96.44	0.93	0.93	0.96
VGG16	94.56	0.92	0.92	0.95

3.3. Performance on Combination of PAMAP2 and UCI HAR Dataset

Table 2 presents the results of combining UCI-HAR and PAMPA2 datasets in terms of performance. The combination consists of using Wasserstein distance clustering to lessen the effects of negative learning brought on by the 3's challenge. The GRU's success can be attributed to its ability to capture temporal dependencies in the sensory data. In activities where the sequence of movements is crucial for recognition, the GRU excels, providing superior accuracy compared to models like LSTM and CNN-LSTM.

Table 2. Performance on Combination of PAMAP2 and UCI HAR Dataset

Model	Accuracy	Precision	Recall	F1-score
LSTM	93.95	0.93	0.93	0.92
CNN-LSTM	96.73	0.95	0.96	0.97
GRU	98.68	0.98	0.98	0.97

3.4. Performance on Combination of Image Dataset and Sensory (PAMAP2 & UCI-HAR) Dataset

Table 3 displays the accuracy of the suggested model, which may be applied to picture datasets in addition to sensory datasets. Integrating image and sensory datasets posed challenges in handling disparate data modalities. The advantage lies in the hybrid model's ability to effectively extract both spatial and temporal features from the respective datasets, creating a more comprehensive representation for accurate activity recognition. The challenge was ensuring a seamless fusion of information from diverse sources.

Table 3. Performance on Combination of Image and Sensory (PAMAP2 & UCI-HAR) Datasets

Model	Accuracy	Precision	Recall	F1-score
RNN	81.33	78.35	71.49	79.65
LSTM	78.65	84.80	63.94	86.90
Bi-LSTM	71.19	84.81	65.80	79.10
GRU-CNN	98.96	0.99	0.99	0.98

4. CONCLUSION AND FUTURE WORK

Human Activity Recognition (HAR) is a pivotal technology with applications in healthcare, sports, and security. In this research, we introduced a novel hybrid GRU_CNN model tailored for HAR, integrating both image and sensory data. Our model utilizes convolutional neural networks (CNN) to extract features from images and gated recurrent units (GRU) for processing sensory data, enhanced by Wasserstein clustering for dimensionality reduction. Experimental results showcased the model's superior accuracy compared to established techniques, marking it as a promising approach for HAR. This would involve acknowledging potential ethical implications, especially in surveillance applications, and outlining strategies or considerations to mitigate these concerns. It adds a layer of responsibility to the research findings. Future research directions include exploring more complex models, alternative clustering techniques for sensory data, and extending the model's capabilities to recognize intricate activities by incorporating additional data sources like audio or text.

REFERENCES

[1] Paul, A., Dey, N., & Chakraborty, S. (2020). Hybrid deep learning model for human activity recognition using smartphone sensors. Multimedia Tools and Applications, 77(14), 18023-18043.
[2] Abinaya, S., & Rajasenbagam, T. (2022). Enhanced Visual Analytics Technique for Content-Based Medical Image Retrieval. IJEER, 10(2), 93-99.
[3] Zhang, Y., Xu, B., Yang, L., & Liu, F. (2019). Multimodal deep learning for human activity recognition: A survey. Neurocomputing, 335, 27-49.
[4] Kumar, P., & Suresh, S. (2023). DeepTransHAR: a novel clustering-based transfer learning approach for recognizing the cross-domain human activities using GRUs (Gated Recurrent Units) Networks. Internet of Things, 21, 100681.
[5] Bishoy, M., Bahaa, E. K., & Khattab, T. M. (2019). Human activity recognition using a hybrid model combining 3D-CNNs and LSTM networks. Multimedia Tools and Applications, 78(5), 5565-5584.
[6] Khan, Z., Gao, Y., Khan, I. U., Ali, M., & Rehman, A. (2021). Multi-head convolutional neural networks with attention-based fusion for human activity recognition. Applied Soft Computing, 100, 107035.

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 36-40 | e-ISSN: 2347-470X

[7] Islam, M. M., Sharif, M. H., Idris, M. Y. I., & Kamal, N. A. M. (2021). Recent advances in deep learning-based human activity recognition: a comprehensive review. Sensors, 21(17), 5926.
[8] Nguyen, H. N., Nguyen, N. T., & Dang, T. N. (2019). Multi-modal deep learning for human activity recognition using RGB-D images and inertial sensors. Sensors, 19(13), 2958.
[9] Lu, L., Zhang, C., Cao, K., Deng, T., & Yang, Q. (2022). A multichannel CNN-GRU model for human activity recognition. IEEE Access, 10, 66797-66810.
[10] Dua, N., Singh, S. N., & Semwal, V. B. (2021). Multi-input CNN-GRU based human activity recognition using wearable sensors. Computing, 103, 1461-1478.