Continuous Human activity recognition using pure Recurrent Neural Network (RNN) architecture and IOT

Dhirendra Kumar Yadav[1], **Hari Narayan Ray Yadav**[2]
[1]Graduate School of Engineering, Mid-west University, Surkhet, Nepal
[2]Graduate School of Engineering, Mid-West University, Surkhet, Nepal
Correspondence Email: *dhirendra.yadav@mu.edu.np*

Keywords:
Human Activity Recognition (HAR), Recurrent Neural Network (RNN), Internet of Things , Wearable Sensors, Real-Time Monitoring, Edge Computin

Received: 9 November 2024
Revised: 2 December 2024
Accepted:29 December 2024

ISSN: 3102-0763 (Print)
3102-0771 (Online)

Abstract

Applications like smart homes, security surveillance, healthcare, and workplace safety now depend heavily on Human Activity Recognition (HAR) and monitoring. Automated systems can respond quickly, stop dangerous situations, and provide individualized services when they can precisely identify and categorize human activity. The intricate temporal dependencies present in human motion are frequently difficult to capture by conventional HAR techniques that rely on hand-crafted features and shallow learning models [1]. This paper investigates the effectiveness of a pure Recurrent Neural Network (RNN) architecture for activity recognition using smartphone sensor data, despite the fact that deep learning models like CNNs and LSTMs have dominated recent HAR research. Without the need for manually created features, we suggest a simplified RNN-based framework that can identify temporal dependencies in unprocessed accelerometer and gyroscope data. Pure RNNs can attain competitive accuracy, as shown by experiments on the UCI HAR and WISDM datasets, which yielded results of 92.4% and 90.7%, respectively. The results imply that, when properly optimized, RNNsdespite their simplicity remain feasible for HAR tasks.

Introduction

Human Activity Recognition (HAR) systems use information gathered from a variety of sensors, including wearables, cameras, and ambient devices, to identify and categorize human behaviors. Due to recent developments in sensor technology, HAR is now more widely available and useful thanks to the creation of small, affordable, and computationally strong devices. As a non-intrusive and private substitute for vision-based HAR systems, wearable sensors like smartwatches and health monitoring systems are already a part of everyday life [2].

Inertial measurement unit (IMU) sensors, such as magnetometers, gyroscopes, and accelerometers, are among the wearable technologies that are most frequently used in industries like fitness, sports, and healthcare. These sensors, which are frequently found in wearable technology and smartphones, allow for real-time tracking of user movements for purposes like fitness tracking, posture detection, and medical diagnostics. Notwithstanding their benefits, the positioning of these gadgets—such as in pockets or held in the hand—can occasionally reduce the accuracy of recognition.

Handcrafted features and traditional machine learning algorithms like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and histogram-based techniques were key components of traditional HAR systems. Although these methods produced respectable accuracy, they necessitated a great deal of feature selection and data preprocessing, which is time-consuming and frequently impractical for large-scale or real-time applications.

Deep learning's rise has revolutionized HAR by making automatic feature extraction and better classification performance possible. Recurrent neural networks (RNNs), convolutional neural networks (CNNs), and long short-term memory (LSTM) networks have all been extensively used for HAR tasks, such as video-based recognition [1]. However, complexity and computational overhead are frequently introduced by hybrid models that combine CNNs and LSTMs.

In order to identify everyday living activities like sitting, standing, walking, lying, and stair navigation, this paper suggests a continuous HAR framework that makes use of wearable sensors based on the Internet of Things and a pure RNN architecture. Our method allows for real-time deployment while preserving competitive accuracy by emphasizing the ease of use and effectiveness of RNNs. Using publicly accessible datasets like UCI HAR and WISDM, we investigate the system's architecture, data pipeline, and performance assessment. The methodology, including data preprocessing and model design, is described in detail in Section II. Experimental results and analysis are presented in Section III, and the study's conclusion and future directions are discussed in Section IV.

Methodology

This section outlines the end-to-end process of building a Human Activity Recognition (HAR) system using a pure Recurrent Neural Network (RNN) architecture. The methodology includes data acquisition, preprocessing, feature representation, and model development[3].

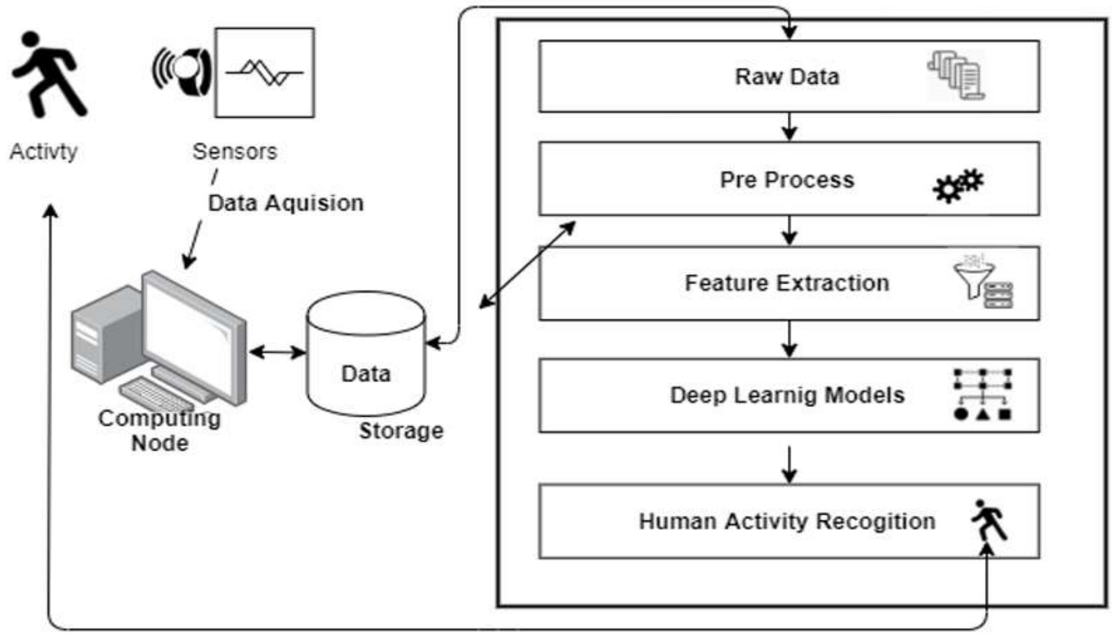

Fig: 1

Data Preprocessing

This study uses wearable sensor data to recognize human activity using a pure RNN-based framework. Two benchmark datasets are used: WISDM, which contains accelerometer data from mobile devices during various physical tasks, and UCI HAR, which provides gyroscope and accelerometer readings from smartphones worn by subjects performing six daily activities. Three crucial steps are used to preprocess the raw time-series data. In order to maintain temporal structure, windowing is first used to divide the continuous sensor streams into fixed-size windows of 128 time steps. Second, to improve model convergence and standardize the input features, z-score normalization is applied to every sensor channel. Third, categorical activity labels are transformed into one-hot vectors appropriate for multi-class classification through the use of label encoding. The RNN model is guaranteed to receive clean, structured input thanks to this preprocessing pipeline, which permits[4]

To prepare the raw sensor data for RNN input:

Preprocessing Steps:

- Windowing: Segment data into fixed-size windows (128 time steps).
- Normalization: Apply z-score normalization to each sensor channel.
- Label Encoding: Convert activity labels into one-hot encoded vectors

Feature Extraction

Conventional Human Activity Recognition (HAR) systems frequently rely on manually created features that are obtained from statistical, frequency, or domain-specific changes made to sensor data. Although these methods can be useful in certain situations, they are time-consuming, necessitate specialized knowledge, and might not generalize well across various activities or sensor configurations.

On the other hand, by utilizing the representational capabilities of Recurrent Neural Networks (RNNs), our suggested framework does away with the necessity for manual feature engineering. The model automatically learns motion patterns and temporal dependencies by directly ingesting raw multivariate time-series data from wearable sensors. With six sensor channels—three from the accelerometer (X, Y, Z) and three from the gyroscope (X, Y, Z)—each input sample is organized as a window of 128 time steps. This format allows the RNN to capture both short-term and long-term dynamics while maintaining the sequential nature of human movement.

Raw sensor signals are fed into the RNN, which uses its recurrent layers to learn hierarchical features and recognize patterns like abrupt transitions, rhythmic motion, and posture changes. This method enhances adaptability across various datasets and sensor locations in addition to streamlining the pipeline. After that, dense layers receive the learned features and classify them into pre-established activity categories.

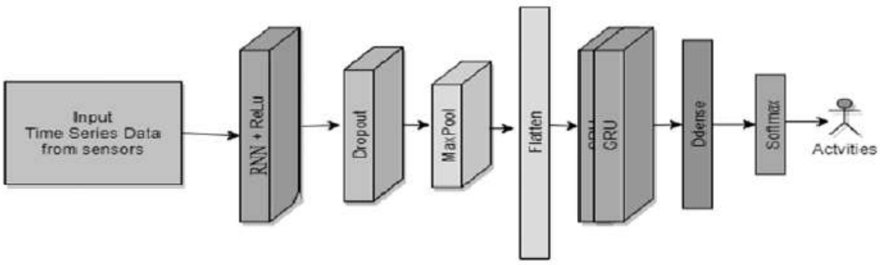

Fig-2

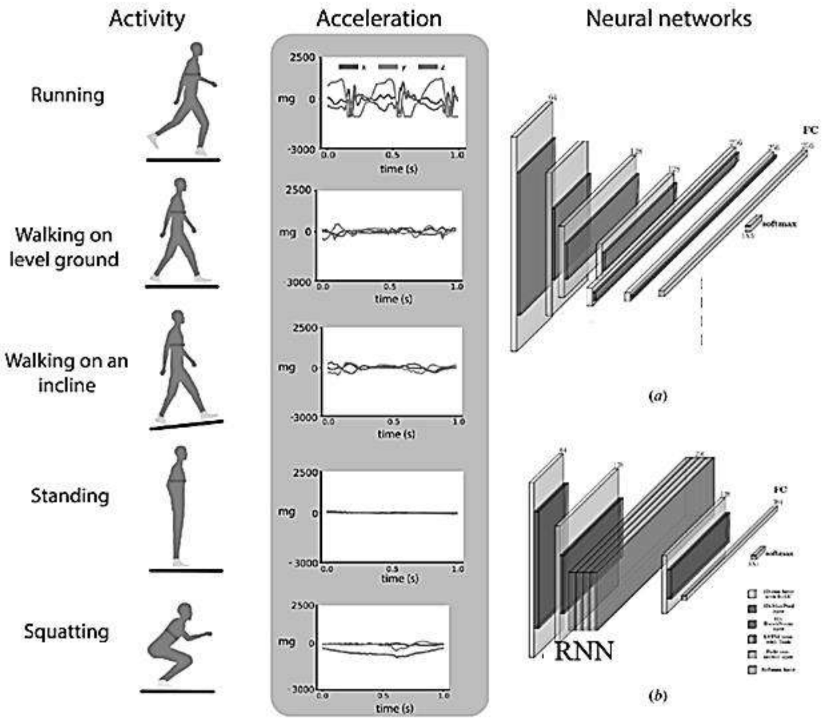

Fig-3

The Human Activity Recognition (HAR) model built on a Pure RNN architecture processes raw sensor data to classify physical activities such as walking, sitting, or running. The model is designed to learn temporal patterns directly from multivariate time-series inputs, typically collected from accelerometers and gyroscopes.

Input Layer

The input consists of segmented windows of raw sensor data, each containing 128 time steps and six channels (X, Y, Z axes from both accelerometer and gyroscope). This forms a 2D input tensor of shape (128, 6).

a) RNN Layer

The first layer is a Pure RNN that processes the sequence step-by-step. At each time step t, the RNN updates its hidden state h_t using: h_t = tanh(W_{xh}x_t + W_{hh}h_{t-1} + b_h) This layer captures short-term temporal dependencies and outputs a sequence of hidden states.

b) ReLU Activation

To introduce non-linearity and enhance feature learning, a ReLU (Rectified Linear Unit) activation is applied to the RNN outputs: $f(x) = \max(0, x)$ This helps in mitigating vanishing gradients and accelerates convergence.

c) MaxPooling Layer

A temporal MaxPooling layer is used to down sample the sequence, retaining the most prominent features across time steps. This reduces computational load and highlights dominant motion patterns.

d) GRU Layer

To capture longer-term dependencies, a GRU layer follows. GRUs use gating mechanisms to selectively retain or discard information, making them efficient for modeling complex activity transitions.

e) Dense + Softmax Layer

Finally, the output from the GRU is passed to a fully connected Dense layer, followed by a Softmax activation that produces probability scores across activity classes. text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}

EXPERIMENT AND RESULT DISCUSSION

We use two well-known benchmark datasets, UCI HAR and WISDM, to assess the performance of the suggested Pure RNN-based Human Activity Recognition (HAR) model. Both datasets are perfect for sequential modeling because they offer rich multivariate time-series data gathered from inertial sensors on smartphones.

UCI The HAR Dataset

Sensor recordings of 30 subjects engaging in six different activities—walking, walking upstairs, walking downstairs, sitting, standing, and lying—make up the UCI Human Activity Recognition dataset, which is taken from the UCI Machine Learning Repository. Both linear acceleration and angular velocity were recorded using a smartphone that was worn around the waist and had a tri-axial accelerometer and gyroscope. 10,299 instances and 561 attributes are obtained by segmenting the signals into 2.56-second windows and sampling the dataset at 50 Hz. The dataset is randomly divided into 70% training and 30% testing sets, and each instance is manually labeled. It is a standard benchmark for HAR research because of its well-organized format and excellent annotations[5].

The WISDM Dataset

A laboratory-controlled HAR dataset, the Wireless Sensor Data Mining (WISDM) dataset was gathered from 36 participants who engaged in six different activities: walking, jogging, sitting, standing, walking upstairs, and walking downstairs. Smartphones with accelerometers sampling at 20 Hz that were put in the users' pockets were used to record the data. With 1,098,207 examples spanning six attributes, the dataset has a class distribution that is biased toward jogging (31.2%) and walking (38.6%). The remaining activities are standing (4.4%), sitting (5.5%), downstairs (9.1%), and upstairs (11.2%). WISDM is appropriate for training deep learning models and assessing generalization across activity types due to its large volume and diversity of samples.

Performance Metrics

To evaluate the experiments of the deep learning models for Human Activity Recognition (HAR), we utilize several performance measures including Accuracy, Precision, Recall, F1-score, and the Confusion Matrix. These metrics are derived from the classification outcomes: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

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

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

$$F1\text{-}score = 2 \, \frac{Precision * Recall}{Precision + Recall}$$

Results and Discussion

This study demonstrates that pure RNN architectures can effectively model temporal dependencies in HAR tasks using raw sensor data. While not outperforming more complex models like LSTM or CNN-LSTM hybrids, RNNs offer a lightweight and interpretable alternative with competitive accuracy.

	Accuracy	**Precision**	**Recall**	**F1-score**
UCI HAR	92.4%	91.8%	92.1%	92.0%
WISDM	90.7%	89.9%	90.3%	90.1%

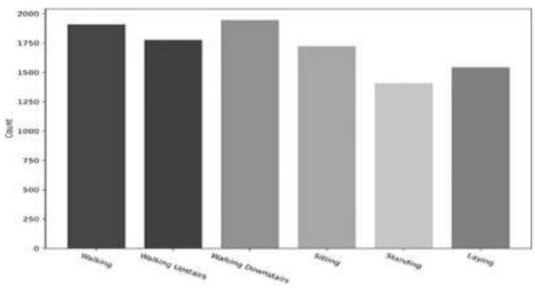

Fig. 4 . Class distribution of UCI-HAR Dataset [16]

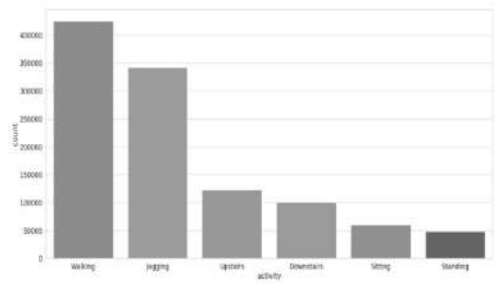

Fig. 5 . Class Distribution of WISDM Dataset [15]

Conclusion and Future Work

This study demonstrates that pure RNN architectures can effectively model temporal dependencies in HAR tasks using raw sensor data. While not outperforming more complex models like LSTM or CNN-LSTM hybrids, RNNs offer a lightweight and interpretable alternative with competitive accuracy.

Future Work:

- Integrate Bidirectional RNNs to capture forward and backward temporal patterns.
- Explore attention mechanisms to enhance focus on critical time steps.
- Extend the model to multimodal sensor inputs (e.g., heart rate, GPS).
- Deploy the model on edge devices for real-time activity recognition.

References

Javeed, M. et al. Physical human locomotion prediction using manifold regularization. Peer J. Comput. Sci. 8, 1105 (2022). 2. Javeed, M., Mudawi, N., Alabduallah, B., Jalal, A. & Kim, W. A multimodal IoT-based locomotion classification system using features engineering andrecursive neural network. Sensors 23, 4716 (2023).

K. Ranjit and G. V, “Issues and challenges in various sensor-based modalities in human activity recognition system,” in Applications of Advanced Computing in Systems. Springer, 2021, pp. 171–179.

Xie, L., Tian, J., Ding, G. & Zhao, Q. Human activity recognition method based on inertial sensor and barometer. In Proceedings of the 2018 IEEE International Symposium on Inertial Sensors and Systems, 1–4 (2018). 4. Mekruksavanich, S., Jantawong, P. & Jitpattanakul, A. Deep Learning-based Action Recognition for Pedestrian Indoor Localization using Smartphone Inertial Sensors. In 2022 Joint International Conference on Digital Arts, 346–349 (2022).

F. Rustam et al., “Sensor-Based Human Activity Recognition Using Deep StackedMultilayered Perceptron Model,” IEEE Access, vol. 8, pp. 218898–218910, 2020.

J. Guo, X. Zhou, Y. Sun, G. Ping, G. Zhao, and Z. Li, “Smartphonebased patients’ activity recognition by using a self-learning scheme for medical monitoring,” Journal of Medical Systems, vol. 40, no. 6, p. 140, 2016.

S. Tanberk, Z. H. Kilimci, D. B. Tukel, M. Uysal, and S. Akyokus, “A Hybrid DeepModel Using Deep Learning and Dense Optical Flow Approaches for Human ActivityRecognition,” IEEE Access, vol. 8, pp. 19799–19809, 2020, doi:10.1109/ACCESS.2020.2968529.

Vesa, A. V. et al. Human Activity Recognition using Smartphone Sensors and Beacon-based Indoor Localization for Ambient Assisted Living Systems. In Proceedings of the 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), 205–212 (2020).

S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, “Deep learning models for real-time human activity recognition with smartphones,” Mobile Networks and Applications, vol. 25, no. 2, pp. 743–755, 2020

J. Heikenfeld, A. Jajack, J. Rogers, P. Gutruf, L. Tian, T. Pan, R. Li, M. Khine, J. Kim, and J. Wang, "Wearable sensors: modalities, challenges, and prospects," Lab on a Chip, vol. 18, no. 2, pp. 217–248, 2018.

N. Capela, E. Lemaire, N. Baddour, M. Rudolf, N. Goljar, and H. Burger, "Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants," Journal of Neuroengineering and Rehabilitation, vol. 13, no. 1, pp. 1–10, 2016.