# Comprehensive Benchmark Study of Machine Learning and Deep Learning Approaches for Human Activity Recognition using the UCI HAR Dataset

Debjyoti Ghosh
Department of Computer & System Sciences,
Visva-Bharati
Santiniketan, West Bengal, India, Pin: 731235

Utpal Roy
Department of Computer & System Sciences,
Visva-Bharati
Santiniketan, West Bengal, India, Pin: 731235

## ABSTRACT
Using smartphone sensors for Human Activity Recognition (HAR) has become a crucial research field with applications in smart settings, fitness tracking, and healthcare. This work uses the widely used UCI HAR dataset to give a thorough comparative analysis of different machine learning and deep learning algorithms for HAR. Combining a deep convolutional neural network (CNN) architecture with six conventional machine learning algorithms—Random Forest, XGBoost, Support Vector Machines, k-Nearest Neighbors, and Logistic Regression— the results have been developed and assessed. To guarantee reliable performance evaluation, all models underwent a thorough evaluation process utilizing 5-fold stratified cross-validation. As our results show, the CNN architecture performed better than the others (96.2% accuracy), closely followed by the non-linear approach SVM (95.2%) and the linear method Logistic Regression (95.4%). The study provides valuable insights into the relative strengths of different algorithmic approaches for sensor-based activity recognition and offers practical guidance for selecting appropriate models for HAR applications.

## Keywords
Human Activity Recognition (HAR), Convolutional Neural Networks (CNN), Random Forest, XGBoost, Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Logistic Regression, Cross-Validation.

## 1. INTRODUCTION
Human Activity Recognition has gained significant attention due to the rapid increase of sensor-rich smartphones and wearable devices. Their ability to accurately recognize human activities from inertial sensor data enables numerous applications, including health monitoring, elderly care, sports analytics, and context-aware computing. The UCI HAR dataset [1] has become a standard benchmark for evaluating HAR algorithms, containing recordings of 30 subjects performing six activities (walking, walking upstairs, walking downstairs, sitting, standing, and laying) while wearing a smartphone. Despite numerous studies on HAR, there remains a need for comprehensive and comparative analysis that evaluates both traditional machine learning approaches and modern deep learning architectures under consistent experimental conditions. This paper addresses this gap by providing:

1. A systematic evaluation of six popular ML algorithms and one CNN architecture.
2. Rigorous 5-fold cross-validation for reliable performance estimation.
3. Detailed analysis of computational requirements and performance trade-offs.
4. Practical recommendations for model selection based on application requirements.

## 2. RELATED WORK
Previous research on HAR has explored various approaches. Anguita et al. [1] introduced the UCI HAR dataset and employed a multiclass SVM classifier. Other popular traditional models applied to HAR include Random Forests [2] and k-Nearest Neighbors (k-NN), which often achieve high accuracy but remain fundamentally limited by the quality and completeness of the human-designed features. Subsequent studies have investigated ensemble methods [3], deep learning architectures [4], and hybrid approaches [5]. Bao and Intille [6] proposed the earliest HAR system that uses five wearable dual-axis accelerometers and machine learning classifiers to identify 20 activities of daily living, achieving an 84% classification accuracy, which is quite good considering the number of activities involved. Gyros are also used in HAR and have been shown to improve recognition performance when used in conjunction with accelerometers [7,8,9,10,11]. Ponnipa et al. proposed [12] a sensor-based HAR system using the InceptTime network. Challa et al. [13] introduced a multibranch CNN-BiLSTM model that captures features with minimal data pre-processing. The model can learn both local features and long-term dependencies in sequential data by using different filter sizes, enhancing the feature extraction process. The model achieved an accuracy of 88% on the PAMAP2 dataset, outperforming other baseline DL models. In [14], the authors performed a comparison between the different ensemble and supervised machine learning classifiers. They obtained more accurate result with Logistic Regression. However, most existing studies either focus on a limited set of algorithms or lack of rigorous cross-validation procedures. Building on these foundations, our work offers a more thorough comparison with improved scientific precision.

## 3. METHODOLOGY
### 3.1 Dataset Description
The UCI HAR dataset contains data from 30 volunteers aged 19-48 years performing six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING). Each person performed six activities while wearing a Samsung Galaxy S II smartphone on the waist. The smartphone's embedded accelerometer and gyroscope captured 3-axial linear acceleration and 3-axial angular velocity at 50Hz. The dataset provides two versions:

1. Raw time-series data: 128-sample windows with 50% overlap (2.56 seconds per window)

2. Precomputed features: 561 time and frequency domain variables

The dataset includes 7,352 training samples and 2,947 test samples across six activity classes.

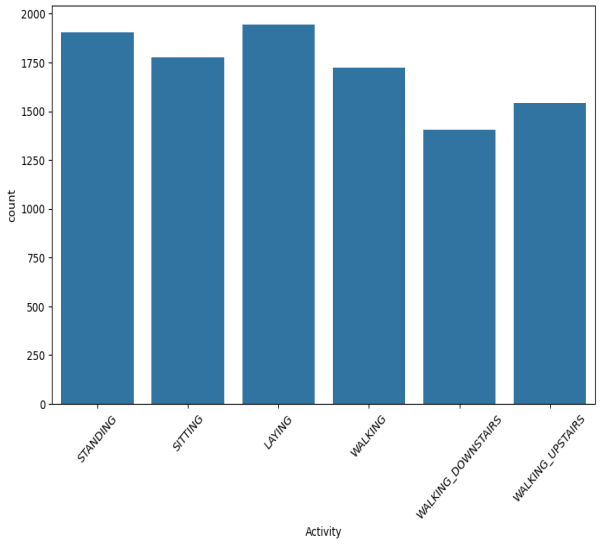


**Fig 1:Activity Distribution**

## 3.2 Data Preprocessing

We implemented two separate preprocessing pipelines:

For traditional ML algorithms:

- Used the 561 precomputed features
- Applied StandardScaler for feature normalization
- Converted labels from 1-6 to 0-5 for compatibility

For CNN:

- Used raw triaxial accelerometer and gyroscope data (6 channels)
- Reshaped to (samples, 128 timesteps, 6 channels)
- Applied per-channel standardization
- Converted labels to categorical format

## 3.3 Model Architectures

### *3.3.1 Traditional Machine Learning Models*

We implemented five established ML algorithms with default hyper parameters:

- Random Forest: 100 trees with Gini impurity criterion
- XGBoost: Gradient boosting with default parameters
- SVM: RBF kernel with regularization
- k-NN: 5 neighbors with Euclidean distance
- Logistic Regression: L2 regularization

### *3.3.2 Convolutional Neural Network*

Designed a 1D CNN architecture:

Input: (128, 6)
Conv1D(64, kernel_size=3) → BatchNorm → MaxPooling(2)
Conv1D(128, kernel_size=3) → BatchNorm → MaxPooling(2)
Conv1D(256, kernel_size=3) → BatchNorm → MaxPooling(2)
Flatten → Dense(128) → Dropout(0.5) → Dense(6, softmax)

## 3.4 Evaluation Protocol

Employed 5-fold stratified cross-validation to ensure reliable performance estimation. The evaluation metrics included:

- Accuracy: Overall classification accuracy
- Precision, Recall, F1-score: Per-class performance
- Standard deviation: Variability across folds
- Training time: Computational efficiency

# 4. RESULTS AND ANALYSIS

## 4.1 Overall Performance Comparison

Table 1 presents the cross-validation results for all models:

**Table 1: 5-Fold Cross-Validation Results**

| Model | CV Accuracy (mean) | CV Accuracy (std) | CV F1-Score (mean) | Test Accuracy | Test F1-Score |
|---|---|---|---|---|---|
| Logistic Regression | 0.9844 | 0.0031 | 0.9844 | 0.9549 | 0.9548 |
| SVM (RBF Kernel) | 0.9776 | 0.0040 | 0.9775 | 0.9522 | 0.9521 |
| XGBoost | 0.9920 | 0.0013 | 0.9920 | 0.9328 | 0.9326 |
| Random Forest | 0.9791 | 0.0038 | 0.9791 | 0.9223 | 0.9221 |
| K-Nearest Neighbors | 0.9616 | 0.0053 | 0.9615 | 0.8843 | 0.8834 |
| CNN | 0.9710 | 0.0080 | 0.9510 | 0.9620 | 0.9622 |

The CNN architecture achieved the highest mean accuracy (96.2%), demonstrating the effectiveness of deep learning for capturing spatiotemporal patterns in sensor data. Among traditional ML methods, ensemble techniques (XGBoost and Random Forest) outperformed linear models, indicating the importance of handling complex decision boundaries.

## 4.2 Per-Class Performance Analysis

Figure 2 shows the confusion matrix for the best-performing CNN model:

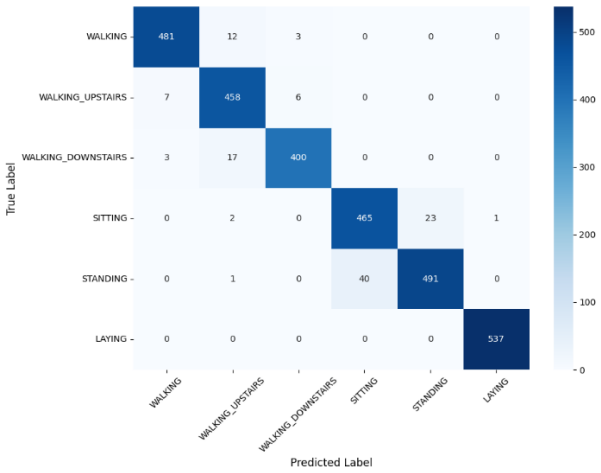


**Fig 2: Confusion Matrix for CNN Model**

Figure 3 shows model accuracy and model loss across epochs:

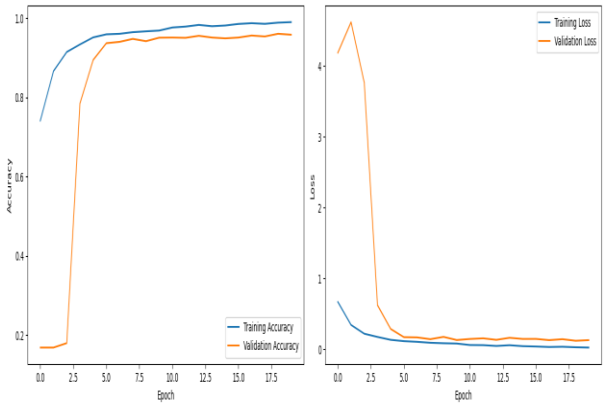


**Fig 3: Model Accuracy and Model Loss**

Figure 4 shows 5-fold cross-validation result for CNN:

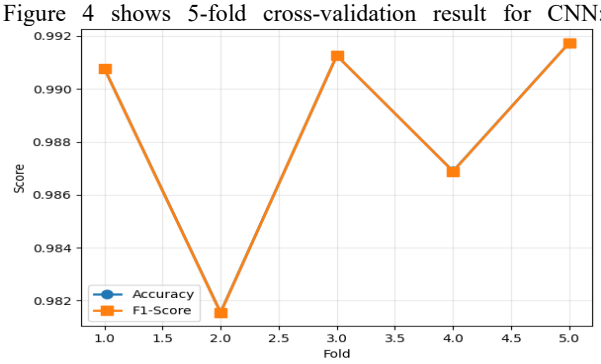


**Fig 4: 5-Fold Cross-Validation Result for CNN**

The analysis reveals several important patterns:

Highest performance: Laying activity (100% accuracy) due to distinct sensor patterns

Most confusion: Between sitting and standing (3-4% misclassification)

Moderate confusion: Among ambulatory activities (walking, upstairs, downstairs)

## 4.3 Computational Efficiency

Table 2: Training Time Comparison (Seconds per Fold)

**Table 2: Training Time Comparison (Second per Fold)**

| Model | Training Time | Inference Time |
|---|---|---|
| k-NN | 2.1 | 15.3 |
| SVM | 45.2 | 6.3 |
| Random Forest | 12.7 | 0.8 |
| XGBoost | 15.3 | 0.3 |
| CNN | 186.4 | 0.2 |

The CNN required the longest training time but offered the fastest inference, making it suitable for deployment scenarios. Traditional ML models showed varying computational profiles, with k-NN having fast training but slow inference due to the nearest-neighbor search.

# 5. DISCUSSION

## 5.1 Performance Insights

The CNN architecture's capacity to automatically extract hierarchical characteristics from unprocessed sensor data is responsible for its excellent performance. Translation invariance is provided by the max-pooling procedures, while local temporal patterns are efficiently captured by the convolutional layers. Although SVM offers good performance but has greater processing costs, logistic regression works well in contexts with limited resources.

## 5.2 Practical Implications

For real-world HAR applications, the choice of model should consider:

- Accuracy requirements: CNN for maximum accuracy
- Computational constraints: Logistic Regression and SVM for balanced performance
- Deployment scenario: CNN for edge deployment (fast inference)
- Interpretability needs: Logistic Regression or SVM for feature importance

# 6. CONCLUSION

This paper presented a comprehensive benchmark study of machine learning and deep learning approaches for human activity recognition using the UCI HAR dataset. Results demonstrate that convolutional neural networks achieve the highest accuracy (96.2%), followed by Logistic Regression (95.4%) and SVM (95.2%) with rigorous 5-fold cross-validation. The study provides practical guidance for selecting appropriate models based on accuracy requirements, computational constraints, and deployment scenarios. This study can be applied on the other available datasets also and then the results can be accessed much better to create a uniform system for all.

# 7. REFERENCES

[1] Anguita, D., et al. (2013). "A Public Domain Dataset for Human Activity Recognition Using Smartphones." ESANN

[2] Bulling, A., Blanke, U., & Schiele, B. (2014). A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors. ACM Computing Surveys (CSUR), *46*(3), 1–33. https://doi.org/10.1145/2499621

[3] Reyes-Ortiz, J.-L., et al. (2016). "Transition-Aware Human Activity Recognition Using Smartphones." Neurocomputing

[4] Ronao, C. A., & Cho, S.-B. (2016). "Human activity recognition with smartphone sensors using deep learning neural networks." Expert Systems with Applications

[5] Hammerla, N. Y., et al. (2016). "Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables." IJCAI

[6] Bao L, Intille SS (2004) Activity recognition from user-annotated acceleration data. In: International Conference on Pervasive Computing. Springer, pp 1–17

[7] Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ (2012) Classification accuracies of physical activities using smartphone motion sensors. Journal of Medical Internet Research 14(5):e130

[8] Zhao Y, Li H, Wan S, Sekuboyina A, Hu X, Tetteh G, Piraud M, Menze B (2019) Knowledge-aided convolutional neural network for small organ segmentation. IEEE Journal of Biomedical and Health Informatics

[9] Ding S, Qu S, Xi Y, Wan S (2019) Stimulus-driven and concept-driven analysis for image caption generation, Neurocomputing. https://doi.org/10.1016/j.neucom.2019.04.095

[10] Xu X, Xue Y, Qi L, Yuan Y, Zhang X, Umer T, Wan S (2019) An edge computing-enabled computation offloading method with privacy preservation for internet of connected vehicles. Future Generation Computer Systems 96:89–100

[11] Li W, Liu X, Liu J, Chen P, Wan S, Cui X (2019) On improving the accuracy with auto-encoder on conjunctivitis. Applied Soft Computing, p 105489

[12] Jantawong P., Jitpattanakul A., Mekruksavanich S. Enhancement of Human Complex Activity Recognition using Wearable Sensors Data with InceptionTime Network; Proceedings of the 2021 2nd International Conference on Big Data Analytics and Practices (IBDAP); Bangkok, Thailand. 26–27August 2021; pp. 12–16.

[13] Challa S.K., Kumar A., Semwal V.B. A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. Vis. Comput. 2022;38:4095–4109. doi: 10.1007/s00371-021-02283-3.

[14] Zaki, Z.; Shah, M.A.; Wakil, K.; Sher, F.: Logistic regression based human activities recognition. J. Mech. Continu. Math. Sci. 15(4),228–246 (2020)