

Human Activity Recognition Through Wearable Sensor Data: An Analytical Study

Poonam Gupta¹, Dr. Aradhana Kumari Singh²

¹ Department of Computer Science & Engineering, MATS University, Raipur

² Department of Computer Science & Engineering, MATS University, Raipur

Abstract - Human Activity Recognition (HAR) using smartphones has become an emerging area of research with applications in healthcare, fitness, and assisted living. Unlike traditional HAR systems that require multiple wearable sensors, this study proposes a lightweight framework that leverages only built-in smartphone sensors such as accelerometers, gyroscopes, gravity, and linear acceleration. A dataset comprising twelve activities—including six standard motions (walking, sitting, standing, etc.) and six extended physical activities (cycling, swimming, tennis, jump rope, push-ups, and football)—was collected using the Andro Sensor application. Preprocessing involved noise removal, statistical feature extraction, and visualization through t-SNE clustering. Four supervised machine learning classifiers—Logistic Regression, Support Vector Machines (SVM), Decision Trees, and K-Nearest Neighbors—were evaluated using accuracy, precision, recall, and F1-score. Experimental results demonstrate that the proposed system achieves robust recognition, with SVM (RBF kernel) attaining the highest accuracy of 86.48% and an average ROC-AUC score of 0.93. The findings highlight the feasibility of smartphone-based HAR systems that balance computational efficiency with high accuracy, making them suitable for real-world applications in health monitoring and activity tracking.

Key Words: Smartphone sensors, accelerometer, Gyroscope, machine learning, Support Vector Machine (SVM), classification, t-SNE visualization

1. INTRODUCTION

Human activity recognition (HAR) on smartphones is a new field with important uses in assisted living, fitness, and healthcare. Smartphones can continuously track and analyze user movements by utilizing built-in sensors like gyroscopes and accelerometers. This allows systems to recognize activities like sitting, running, walking, and even detecting falls. This technology has the potential to support fitness tracking applications, enhance personal health monitoring, and offer prompt interventions in healthcare settings. HAR has made significant strides, but a number of significant obstacles remain in the way of its broad use, particularly when it comes to real-time applications on devices with limited resources like smartphones.

The nature of the sensor data itself is one of the main obstacles in HAR. Due to varying user behaviours, environmental factors, and device placements, the data

gathered from smartphone sensors is frequently noisy, high-dimensional, and subject to variability. It is challenging to create generalized models that can reliably identify activities across a range of populations and contexts because of this variability. Furthermore, the limited processing power and battery life of smartphones limit real-time implementation, requiring algorithms that are computationally efficient and lightweight.

The majority of HAR systems in use today are based on conventional machine learning models, which either prioritize efficiency at the expense of performance or aim for high accuracy at the expense of computational efficiency. Furthermore, filter-based or wrapper-based techniques are frequently used to approach feature selection, a crucial part of HAR. Although wrapper methods are more accurate, they are computationally prohibitive for real-time applications, and filter methods are computationally cheap but often miss crucial feature dependencies. Therefore, sophisticated feature selection methods that can strike a balance between computational efficiency and accuracy are required.

HAR's classification models also have optimization issues. These models are usually hyperparameter tuned manually or by brute force techniques, which take a lot of time and frequently produce less-than-ideal results. This necessitates investigating optimization strategies like metaheuristic algorithms, which can more effectively look for near-optimal solutions. Metaheuristic techniques have not been widely used in HAR systems, despite their success in other fields, especially when it comes to situations involving smartphone real-time decision-making.

Furthermore, transfer learning has become a potent method for improving model adaptability, especially in situations where labelled data is limited. Although transfer learning has demonstrated promise in fields like natural language processing and image recognition, its potential in HAR is still untapped. Additionally, by allowing the model to selectively focus on the most significant sensor signals, attention mechanisms—which enable models to focus on the most relevant parts of the input data—could greatly enhance the performance of transfer learning in HAR. A General Structure of HAR system is shown in fig. 1.

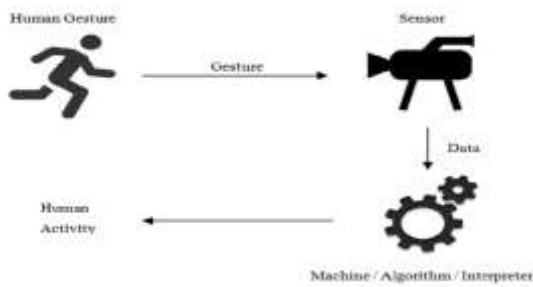


Fig. 1. General Structure of HAR System (*Human Activity Recognition*)

The development of smartphone-based human activity recognition (HAR), which has great promise for a number of fields like human-computer interaction, fitness, and healthcare, provided motivation. Since smartphones with sophisticated sensors are so common, HAR systems can provide real-time information about the health and behavior of users. Despite this promise, there are still a number of obstacles to overcome in order to create accurate and efficient HAR systems that function well on devices with limited resources, such as smartphones. Some Key Motivations as follows:

A. Need for Real-World Applicability

HAR systems are essential for applications like tracking exercise regimens, keeping an eye on patient health, and identifying anomalous behavior like falls. But existing systems frequently have trouble striking a balance between the computational limitations of smartphones—such as their short battery life and processing power—and the requirement for high accuracy. The goal of this research is to create more effective HAR algorithms that can function in practical situations without putting undue strain on the device.

B. Challenges of High-Dimensional and Noisy Data

The information gathered from smartphone sensors is frequently noisy, high-dimensional, and variable depending on the user and the environment. Reduced accuracy results from existing HAR methods' frequent inability to filter out irrelevant information. Thus, the goal of this research is to create sophisticated feature selection methods that can simplify data while maintaining important details for activity recognition.

C. Enhancing Model Generalization Across Users and Environments

Making sure that models generalize well across various users and activity contexts is one of the main challenges in HAR systems. Because of domain shifts, models that were trained on particular datasets typically perform poorly when used with different users or environments. By investigating transfer learning strategies, this study aims to address this problem. It places special emphasis on integrating attention mechanisms to enhance adaptability and model generalization.

D. Optimization of Classifier Performance

Current HAR classifiers frequently have fixed parameters, which restricts their capacity to adjust to novel datasets or activity patterns. This may lead to poor generalization to new data and decreased recognition accuracy. A viable way to dynamically modify classifier parameters and boost performance without incurring undue computational overhead is through metaheuristic optimization techniques.

The goal of the study is to increase the robustness and flexibility of HAR classifiers by incorporating metaheuristic optimization.

E. Addressing the Research Gap in Efficient HAR Algorithms

Even though HAR has received a lot of attention lately, there is still work to be done in order to create algorithms that are both computationally efficient and extremely accurate for use on smartphones. The majority of existing solutions either focus on efficiency at the expense of performance or prioritize accuracy, which results in high computational costs. By creating methods that strike a balance between the need for high accuracy and low computational requirements, this research aims to close this gap and make HAR systems more practical for daily use.

2. RELATED WORK

We divided the previous work related to our contributions in two sections: a brief review of already done work (Section II-A) and a Literature review table (Section II-B).

A. Brief Review Of The Work Already Done

In the field of human activity recognition (HAR), improving classification performance is vital. One study proposed a novel feature selection method, modified Guided Regularized Random Forest (MGRRF), which successfully enhanced recognition accuracy to 98% on a public HAR dataset using five classifiers: RF, k-NN, logistic regression, SVM, and XGBoost. This method focuses on selecting more interpretable features, reducing model bias, and avoiding overfitting.

Enhancing smartphone-based HAR has been another area of interest. A parameter-free ensemble feature selection method was introduced, combining traditional feature selection with association rule mining. This approach improved classification efficiency and significantly reduced computation time on UCI-HAR and UCI-HAPT datasets.

Smartphone sensor data has also been used to classify physical activities and detect falls. In one study, data from participants performing various activities and fall events were collected. Machine learning algorithms, including SVM and RF, achieved high accuracy rates, while k-NN excelled in fall classification.

Feature vector dimensionality reduction and accurate activity classification have been addressed using a hybrid approach combining LDA and MLP methods. Evaluated on the UCI-HAR dataset, this method achieved high metrics in accuracy, precision, recall, and F1-score, outperforming state-of-the-art models.

Real-life environments pose unique challenges for HAR. One investigation utilized a dataset collected from fixed sensor placement and single sensor usage in existing datasets.

B. Literature Review:

Human Activity Recognition (HAR) using sensor data has gained significant attention due to its applications in smart environments, healthcare, and wearable technologies. Islam and Talukder [1] developed a smartphone-based HAR system utilizing exploratory data analysis, visualization,

dimensionality reduction, and ensemble learning, achieving 97.12% accuracy; however, the ensemble model posed high computational overhead, limiting real-time deployment. To enhance deep-learning-based HAR, Dahou et al. [2] proposed the MLCNNwav model integrating residual CNNs with trainable wavelet transforms, which improved feature representation and generalization, though its reliance on high-performance hardware restricts deployment in low-power wearables. A multi-structural approach was introduced by Saha et al. [3] through FusionActNet, combining static and dynamic feature extraction via dual residual networks, resulting in 97.35% accuracy on UCI-HAR, but its computational complexity remains a bottleneck. Wang et al. [4] explored spectrogram-based sensor representation with transfer-learning CNNs and wrapper-based feature selection, achieving up to 97.97% accuracy across datasets while reducing feature dimensionality; however, the framework was limited to accelerometer and gyroscope data, potentially affecting generalizability. To address wearable resource constraints, Yi et al. [5] designed a lightweight HAR model using statistical features along with RF and PQ-CNN classifiers, obtaining high F1 scores above 0.94, though its dependency on single-accelerometer inputs limits activity from 22 participants were recorded using accelerometers and a chest-mounted camera, with SVM yielding the best results among seven machine learning models. This work highlighted issues like non-diversity. Imanzadeh et al. [6] focused on real-world HAR challenges by collecting a noisy dataset and employing a weighted hybrid deep ensemble, achieving high accuracy despite the dataset's limited size and sensor variability. For healthcare-oriented HAR, Yi-Chun Lai et al. [7] combined ML and DL models including ANFIS, SVM, RF, and 1D-CNN across UCI-HAR, MHEALTH, and WISDM datasets, demonstrating strong performance, though limitations were not reported. Garcia-Gonzalez et al. [8] emphasized real-life settings by applying ML models to smartphone sensor data, where tree-based classifiers, especially Random Forest, significantly improved recognition accuracy from 74.39% to 92.97%. Logacjov et al. [9] introduced the HARTH dataset and evaluated several ML and DL algorithms, identifying SVM as the top performer with an F1-score of 0.81, though sensor placement inconsistency affected accuracy. A broader perspective was provided by Ferrari et al. [10], who surveyed methods across acquisition, preprocessing, segmentation, feature extraction, and classification, highlighting strengths and weaknesses but without specifying limitations. Finally, Mekruksavanich et al. [11] proposed a hybrid 4-layer CNN-LSTM model optimized via Bayesian tuning, outperforming baseline LSTMs on the UCI-HAR dataset, although limitations were not explicitly stated. Overall, the literature shows consistent progress toward higher accuracy and robustness through deep learning, feature fusion, and hybrid modeling, while common challenges include high computational cost, limited sensor modalities, and constrained real-world applicability.

3. METHDOLOGY

A. Overview:

The main aim of this paper is to implement a Human Activity Recognition System (HAR System). The purpose of this system is to identify the action which the user would be doing, solely based off of the changes in motion of the user's body during the performance of specific actions[12].

Most HAR systems implemented would use specialised motion sensors that would be secured to the users' body, including, but not limited to, the waist, chest, arms and legs. However, the main problem with this type of system is the complex setup the user would be required to wear during the activity, in addition to the added expenses when purchasing these sensors. Considering the simplicity of the application, many users are more likely to get discouraged in using such a complex, albeit excessive, set up. As a result of the rapid advancements in the technological field as well as the efforts of many researchers, this setup has been reduced to needing only a smartphone. This initial set-up made use of a bulkier mounting system through the use of a belt, an aspect of the set-up which can be improved upon, with the users' comfort being the main priority[12].

Therefore, for this paper, we were aiming to develop a simple Human Activity Recognition prototype which only uses the built-in sensors found in an average smartphone and eliminating the use of a belt mount, allowing the user to carry their phone in their pockets. While this may result in less accurate predictions, it allows users to retain their usual habits; keeping their phone in their pockets[12].

B. DataSets Overview:

The Kaggle-HAR dataset was collected using the initial set-up made use of a bulkier mounting system through the use of a belt, an aspect of the set-up which can be improved upon, with the users' comfort being the main priority. Here, two separate datasets were gathered by the three members working on this APT. The first dataset was done to mimic that made in the paper one with six total actions: Walking, Walking Downstairs, Walking Upstairs, Sitting, Standing, and Laying. This dataset was created in order to compare the difference in results gathered and processed. However, we also collected a second dataset in which we chose physical activities which were not included in the existing data. These also required body movement from the user and were recorded through the accelerometer and gyroscope sensors found in the smartphone. The final new activities are Cycling, Football, Swimming, Tennis, Jump Rope and Push-ups. In summary, the main aim of this paper was not only to interpret the original six activities that most Human Activity Recognition papers tend to focus on, but also to recognise another six unique physical activities[12].

The process of classifying the data with a high accuracy can be divided into two steps: data collection and modelling. In order to collect a sufficient amount of data, the free app 'AndroSensor' was used. Using this app allowed for the collection of data using the four main inertia sensors: gyroscope, gravity, accelerometer and linear acceleration. The data collected consists of roughly one hour worth of data for each of the 12 activities mentioned above, allowing for the model to be developed with an even distribution of data across all the categories[12].

C. Data Collection and Preprocessing:

After the data is collected, it is pre-processed. The pre-processing entails the removal of "NaN" and duplicated values while also generating statistical readings from the 'csv' file produced by 'Andro Sensor'. After being processed, the data is analysed via a t-SNE algorithm which aids the visualisation of data clusters. Finally, the data is modelled and

classified using four different supervised machine learning algorithms: Logistic Regression, Support Vector Machines, Decision Trees and K-Nearest Neighbours [12]. The workflow depicted in Figure 2 illustrates the sequential and interconnected nature of the proposed framework.

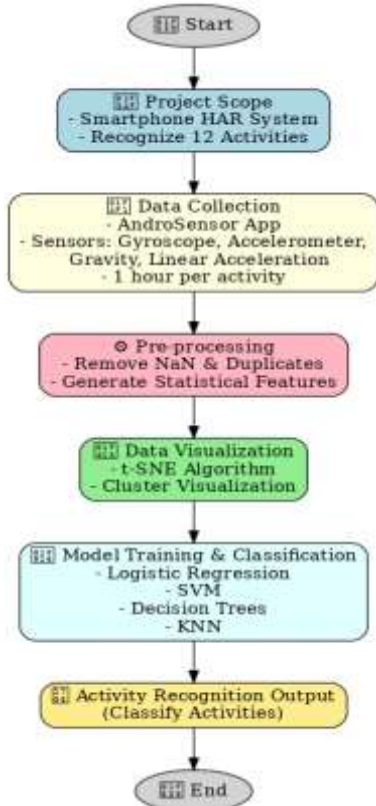


Fig 2: Workflow Diagram of Proposed Method

D. Data Visualization:

The t-SNE visualization used all the features from HAR dataset, excluding the 'Activity' column, which was the target variable. These features are essentially a comprehensive set of extracted characteristics from various sensor data, likely collected from accelerometers, gyroscopes, and possibly other motion sensors. Specifically, the features can be broadly categorized as statistical measures derived from:

- **ACCELEROMETER** data (X, Y, Z axes and combined XYZ): This includes mean, mad (mean absolute deviation), max, min, std (standard deviation), energy, iqr (interquartile range), and entropy for each axis, as well as combined measures like
ACCELEROMETER_XYZ~sma (signal magnitude area).
- **GYROSCOPE** data (X, Y, Z axes and combined XYZ): Similar statistical measures (mean, mad, max, min, std, energy, iqr, entropy) are extracted for gyroscope readings, along with GYROSCOPE_XY~correlation, GYROSCOPE_YZ~correlation, GYROSCOPE_ZX~correlation, and GYROSCOPE_XYZ~sma.
- **GRAVITY** data (X, Y, Z axes and combined XYZ): Again, various statistical measures and GRAVITY_XYZ~sma.

- **LINEAR_ACCELERATION** data (X, Y, Z axes and combined XYZ): Includes statistical measures and

correlations like :

LINEAR_ACCELERATION_XY~correlation,
LINEAR_ACCELERATION_YZ~correlation,
LINEAR_ACCELERATION_ZX~correlation, and
LINEAR_ACCELERATION_XYZ~sma.

These diverse features capture different aspects of motion and movement patterns, which t-SNE then uses to map similar activity patterns closer together in the 2D space. The t-SNE plots provide a valuable cluster visualization of your Human Activity Recognition (HAR) data.

The t-SNE plots three visualizations, each representing the high-dimensional activity data projected into a two-dimensional space using different perplexity values (5, 10, and 20) is shown in figure 3(a), figure 3(b), figure 3(c).

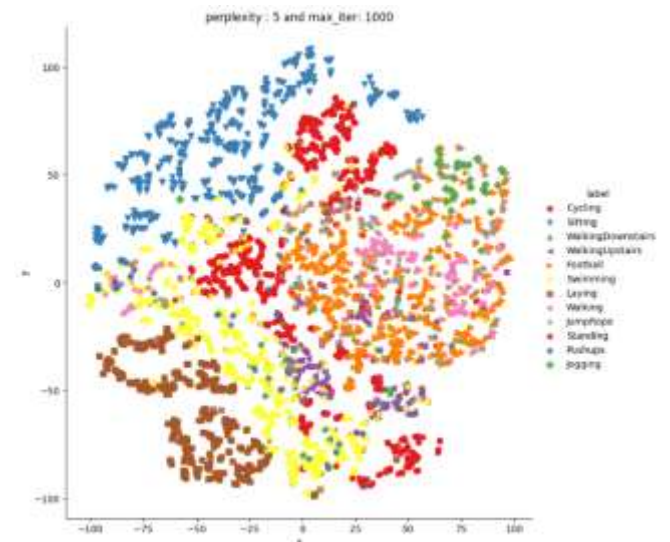


Fig 3(a): t- SNE Visualization with Data Cluster of Perplexity 5

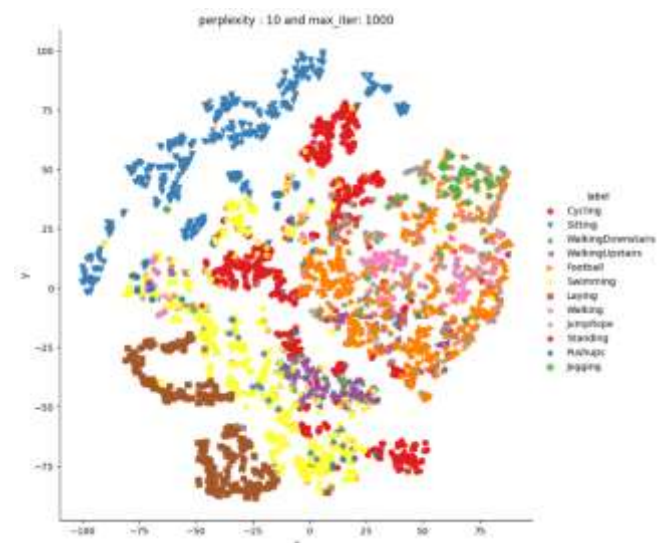


Fig 3(b): t- SNE Visualization with Data Cluster of Perplexity 10

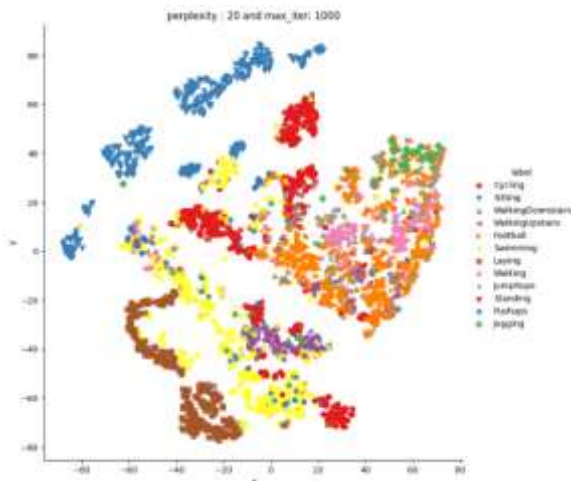


Fig 3(c): t- SNE Visualization with Data Cluster of Perplexity 20

E. Classification Algorithms:

Classification is a method of machine learning in which the number of classes is predetermined, and each instance is labeled with one of these classes to serve as an algorithm guide (Labels that exist provide guidance for the algorithm; hence classification is a supervised learning approach). A Mapping between a predefined class and the training data can be established by learning training data [13]. In this research, machine algorithms such as LR, SVM, DS, KNN, Classifier, were applied to the datasets [23].

Logistic Regression: Although it is called are regression model, logistic regression is regarded as a classification model [14][23].

Support Vector Machine: SVM divides the sample space into two categories based on a hyper-plane, which is discovered through training and learning [15]. Maximizing the margin between the hyperplane and the support vectors is the objective of the SVM [16][23].

Decision Trees: As a classifier and are regression model, decision trees can be used to solve problems in both classification and regression by dividing the data into smaller segments and filling in the leaf nodes with results [17]. Decision trees are popular because they are able to provide a simple representation of complex problems that are easier to interpret, as well as their ability to support inference tasks through the production of logic rules for classification [18][23].

K-nearest neighbours: KNN is built upon the most basic assumption that underlies all predictions: As a result of the same characteristics, most likely, they will produce similar results [19]. When new data are provided to the algorithm, the algorithm compares the new data to the closest data to determine a data class [20]. Though simple, the method is grounded in non-parametric density estimation [21] and outperforms more sophisticated methods in many cases [22][23].

F. Evaluation Metrics:

To evaluate the effectiveness of the method, the study used performance metrics such as accuracy, precision, recall, and F1 score. These metrics were calculated using the generated confusion matrix values. The confusion matrix is a

representation that illustrates how well a model performs in classifying samples for each category in classification. It provides information on true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN), which indicate correct and incorrect predictions for positive and negative classes. TP and TN signify correct predictions, while FP and FN signify incorrect predictions. Equations show performance metrics of accuracy, precision, recall, and F1-score, respectively [23]:

Accuracy: Accuracy measures the overall correctness of a predictive model by calculating the ratio of correctly classified instances to the total number of instances as shown in Eq. (1).[24]

$$\text{Accuracy} = \frac{\sum_{c=1}^K TP_c}{N} \quad (1)$$

Precision: Precision focuses on the proportion of true positive predictions out of all positive predictions made by a model. It measures the ability of the model to accurately identify positive instances as shown in Eq. (2)[24].

$$\text{Precision} = \frac{1}{K} \sum_{c=1}^K \frac{TP_c}{TP_c + FP_c}. (2)$$

Recall: Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive instances that are correctly identified by the model as shown in (3)[24].

$$\text{Recall} = \frac{1}{K} \sum_{c=1}^K \frac{TP_c}{TP_c + FN_c}. \quad (3)$$

F1 score: The F1 score combines precision and recall into a single value, balancing the two metrics as shown in (4)[24].

$$F1 = \frac{1}{K} \sum_{c=1}^K \frac{2 \cdot TP_c}{2 \cdot TP_c + FP_c + FN_c}. \quad (4)$$

Accuracy represents the proportion of correct predictions in relation to the overall number of input samples. By measuring the precision of the model, we can identify how accurate it is at classifying positive samples. In other words, using precision, we aim to identify all positive samples correctly as positive, and not misidentify a negative sample as positive. On the other hand, as far as the call is concerned, it doesn't matter what happens to negative samples, just how positive samples are classified. As a result, if all positive samples are labelled Positive by the model, then the recall will be 100%, regardless of whether all negative samples are incorrectly labelled Positive. It is always possible to maximize one or the other but not both because of some trade-off. This is because high recall requires the generation of incorrect results, which reduces precision [25]. Therefore, the F1score becomes necessary for them to maintain a trade-off [23].

4. EXPERIMENTAL RESULTS & DISCUSSION

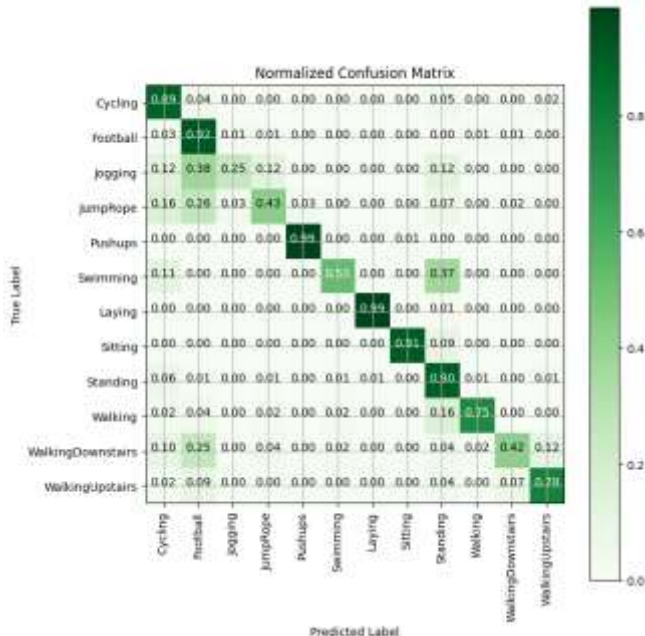
In this section, we delve into the specifics of the experimental configuration and showcase the experimental results using metrics such as accuracy, precision, recall, and F-score.

A. Performance results using Support Vector Machine(SVM)
SVM is popularly used in HAR literature to recognise human physical activities. In this paper SVM uses Radial Basis Function and Polynomial Function for dealing with non-linearly separable data.

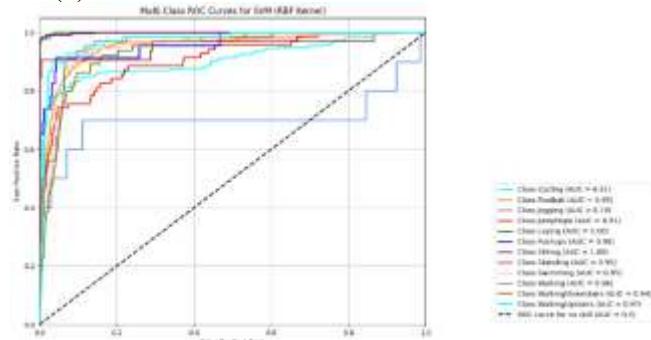
B. *Radial Basis Kernel(RBF)*: The performance results are shown in Fig. 4. Fig. 4(a) shows the classification report using the RBF SVM classifier. Fig. 4(a) shows the precision, recall, f1- score, and accuracy for all 12 different activities, it's average accuracy is 86.48%. Fig. 4(b) shows the confusion matrix of the using RBF SVM classifier with test data to represent the type I (false positive) and II (false negative) classification error. Grid search with 05-fold cross-validation is used in SVM with $C = 90$, gamma and scale chosen by the grid search as the best parameters. The average ROC-AUC score is 0.93. Fig. 4(c) shows the ROC-AUC score for each activity. Fig.4(d) shows the Classification Accuracy of all activities using RBF Kernel SVM Classifier.

	precision	recall	f1-score	support
Cycling	0.84	0.89	0.87	233
Football	0.83	0.92	0.87	250
Jogging	0.29	0.25	0.27	8
JumpRope	0.76	0.43	0.55	58
Laying	0.99	0.99	0.99	142
Pushups	0.77	0.53	0.62	19
Sitting	0.99	0.99	0.99	207
Standing	0.91	0.91	0.91	22
Swimming	0.81	0.90	0.86	199
Walking	0.90	0.75	0.82	51
WalkingDownstairs	0.71	0.42	0.53	52
WalkingUpstairs	0.76	0.78	0.77	54
accuracy			0.86	1295
macro avg	0.80	0.73	0.75	1295
weighted avg	0.86	0.86	0.86	1295

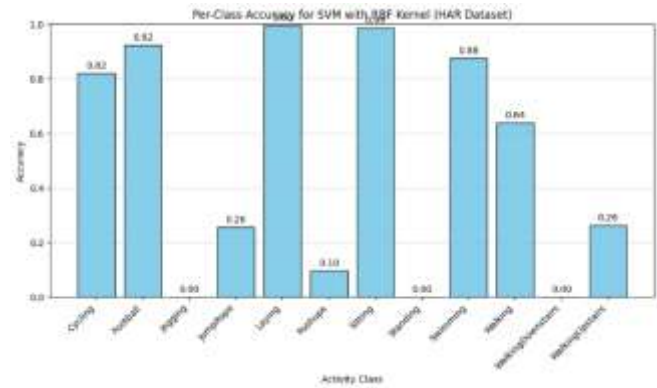
(a) Classification Report of RBF Kernel SVM Classifier



(b) Confusion matrix of RBF Kernel SVM Classifier



(c) ROC Curve of RBF Kernel SVM Classifier



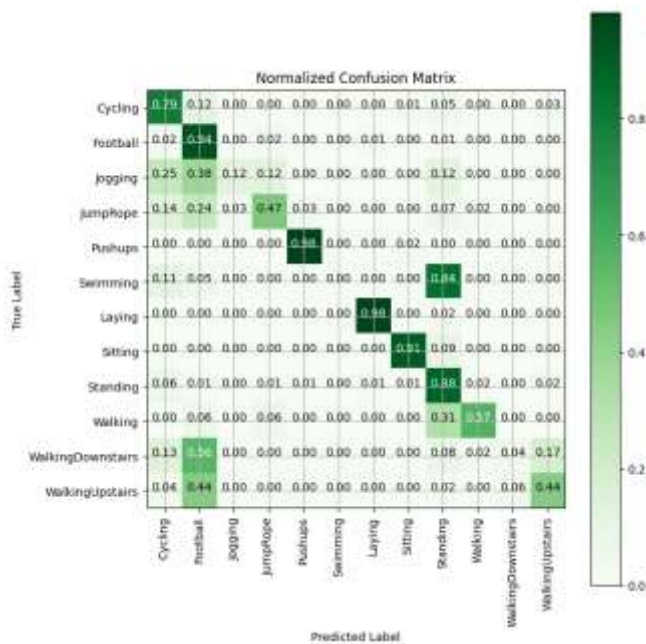
(d) Classification Accuracy of all activities using RBF Kernel SVM Classifier

Fig.4: Performance using RBF Kernel SVM classifier.

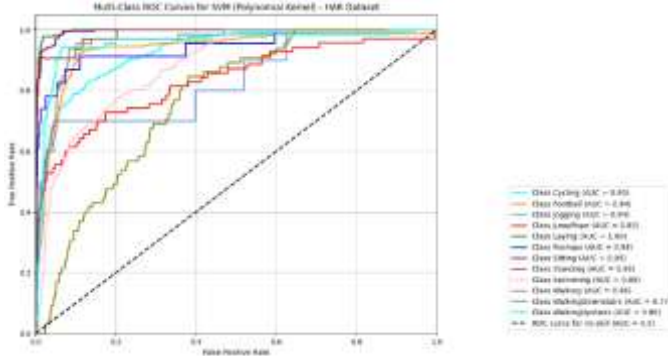
B. *Polynomial Kernel*: The performance results are shown in Fig. 5. Fig. 5(a) shows the classification report using the Polynomial Kernel SVM classifier. Fig. 5(a) shows the precision, recall, f1- score, and accuracy for all 12 different activities, it's average accuracy is 80.15%. Fig. 5(b) shows the confusion matrix of the using Polynomial Kernel SVM classifier with test data to represent the type I (false positive) and II (false negative) classification error. Grid search with 05-fold cross-validation is used in SVM with $C = 16$, degree=1, Kernel= 'Poly' chosen by the grid search as the best parameters. The average ROC-AUC score is 0.919. Fig. 5(c) shows the ROC-AUC score for each activity. Fig.5(d) shows the Classification Accuracy of all activities using Polynomial Kernel SVM Classifier.

	precision	recall	f1-score	support
Cycling	0.83	0.79	0.81	233
Football	0.69	0.94	0.80	250
Jogging	0.25	0.12	0.17	8
JumpRope	0.71	0.47	0.56	58
Laying	0.98	0.98	0.98	142
Pushups	0.80	0.00	0.00	19
Sitting	0.98	0.98	0.98	207
Standing	0.77	0.91	0.83	22
Swimming	0.74	0.88	0.80	199
Walking	0.83	0.57	0.67	51
WalkingDownstairs	0.40	0.04	0.07	52
WalkingUpstairs	0.56	0.44	0.49	54
accuracy			0.80	1295
macro avg	0.65	0.59	0.60	1295
weighted avg	0.78	0.80	0.78	1295

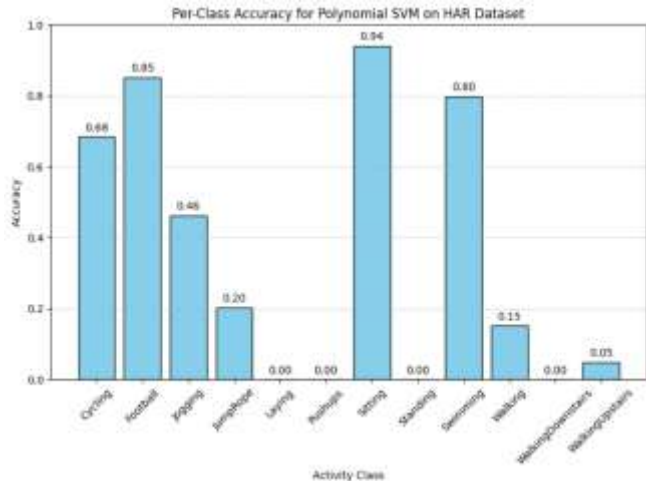
(a) Classification Report of RBF Kernel SVM Classifier



(b) Confusion matrix of Polynomial Kernel SVM Classifier



(c) ROC Curve of Polynomial Kernel SVM Classifier



(d) Classification Accuracy of all activities using Polynomial Kernel SVM Classifier

Fig.5: Performance using Polynomial Kernel SVM classifier

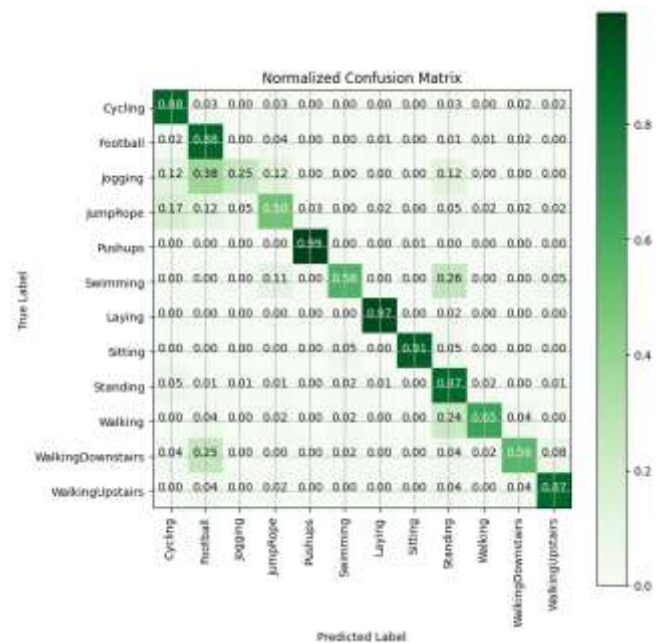
Performance results using Logistic Regression (LR)

LR is popularly used in HAR literature to recognise human physical activities. The performance results are shown in Fig. 6. Fig. 6(a) shows the classification report using LR classifier. Fig. 6(a) shows the precision, recall, f1- score, and accuracy for all 12 different activities, it's average accuracy is 85.86%. Fig. 6(b) shows the confusion matrix of the using LR classifier with test data to represent the type I (false positive) and II (false negative) classification error. Grid search with

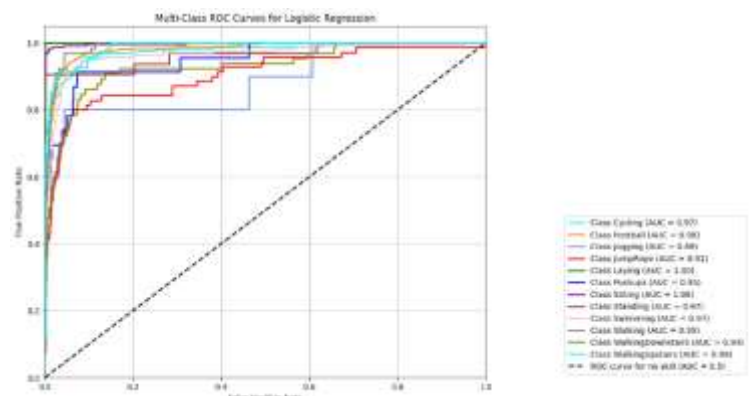
05-fold cross-validation is used in LR with $C = 0.125$, multiclass= 'multinomial', solver= 'newton-cg' chosen by the grid search as the best parameters. The average ROC-AUC score is 0.961. Fig. 6(c) shows the ROC-AUC score for each activity. Fig. 6(d) shows the Classification Accuracy of all activities using LR Classifier.

	precision	recall	f1-score	support
Cycling	0.88	0.88	0.88	233
Football	0.86	0.88	0.87	250
Jogging	0.22	0.25	0.24	8
JumpRope	0.55	0.50	0.52	58
Laying	0.99	0.99	0.99	142
Pushups	0.65	0.58	0.61	19
Sitting	0.97	0.97	0.97	207
Standing	0.95	0.91	0.93	22
Swimming	0.81	0.87	0.84	199
Walking	0.82	0.65	0.73	51
WalkingDownstairs	0.67	0.56	0.61	52
WalkingUpstairs	0.80	0.87	0.83	54
accuracy			0.86	1295
macro avg	0.76	0.74	0.75	1295
weighted avg	0.86	0.86	0.86	1295

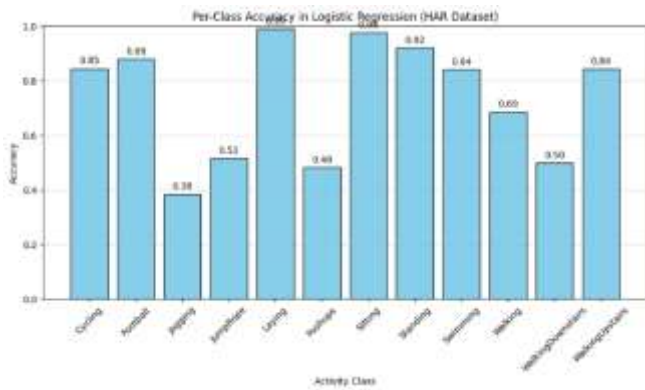
(a) Classification Report of LR Classifier



(b) Confusion matrix of LR Classifier



(c) ROC Curve of LR Classifier



(d) Classification Accuracy of all activities using LR Classifier

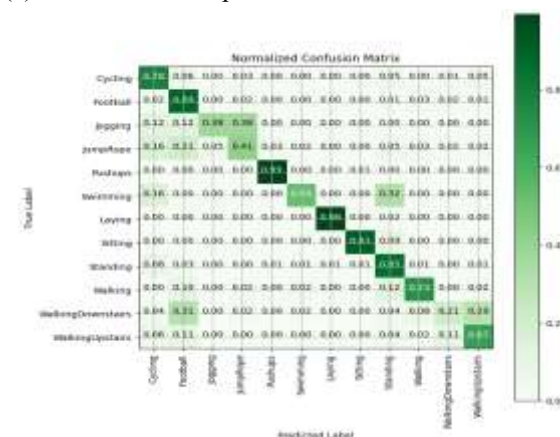
Fig.6: Performance using LR classifier

Performance results using k-nearest neighbor(KNN)

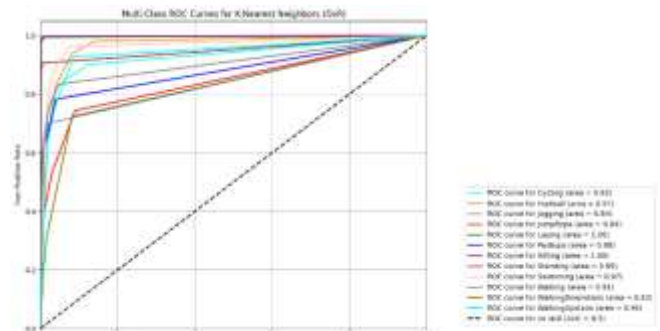
KNN is popularly used in HAR literature to recognise human physical activities. The performance results are shown in Fig. 7. Fig. 7(a) shows the classification report using KNN classifier. Fig. 7(a) shows the precision, recall, f1- score, and accuracy for all 12 different activities, it's average accuracy is 81.77%. Fig. 7(b) shows the confusion matrix of the using KNN classifier with test data to represent the type I (false positive) and II (false negative) classification error. Grid search with 05-fold cross-validation is used in KNN with C = 0.125, multiclass= 'multinomial', solver= 'newton-cg' chosen by the grid search as the best parameters. The average ROC-AUC score is 0.922. Fig. 7(c) shows the ROC-AUC score for each activity. Fig. 7(d) shows the Classification Accuracy of all activities using KNN Classifier.

	precision	recall	f1-score	support
Cycling	0.83	0.78	0.80	233
Football	0.79	0.89	0.83	250
Jogging	0.50	0.38	0.43	8
JumpRope	0.59	0.41	0.48	58
Laying	0.97	0.99	0.98	142
Pushups	0.62	0.53	0.57	19
Sitting	0.99	0.98	0.99	207
Standing	0.83	0.91	0.87	22
Swimming	0.81	0.85	0.83	199
Walking	0.67	0.73	0.70	51
WalkingDownstairs	0.46	0.21	0.29	52
WalkingUpstairs	0.53	0.67	0.59	54
accuracy			0.82	1295
macro avg	0.72	0.69	0.70	1295
weighted avg	0.81	0.82	0.81	1295

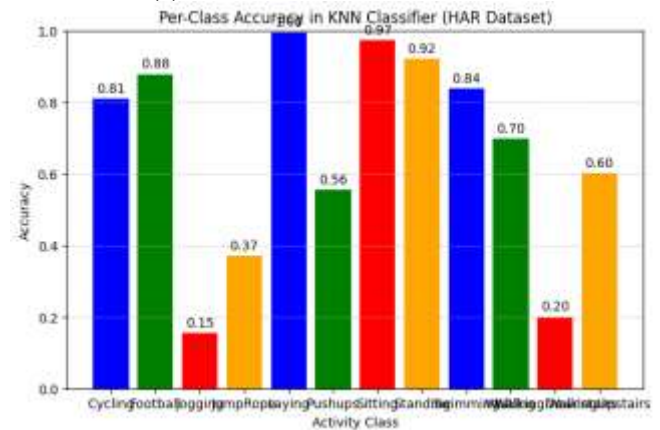
(a) Classification Report of KNN Classifier



(b)Confusion matrix of KNN Classifier



(c) ROC Curve of KNN Classifier



(d) Classification Accuracy of all activities using KNN Classifier

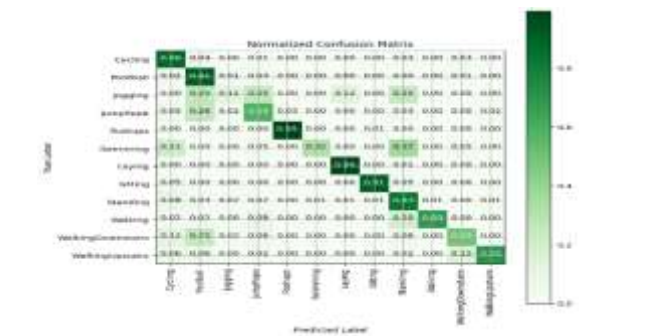
Fig.7: Performance using KNN classifier

Performance results using Decision Tree

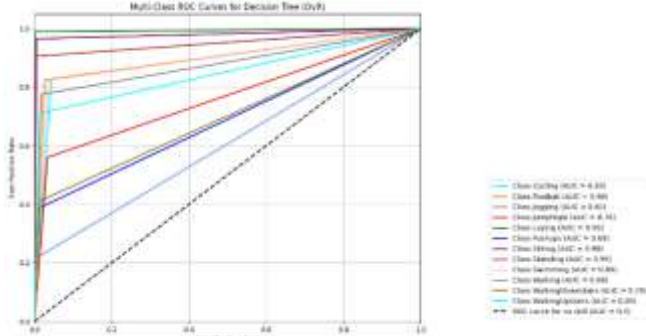
Decision Tree is popularly used in HAR literature to recognise human physical activities. The performance results are shown in Fig. 8. Fig. 8(a) shows the classification report using Decision Tree classifier. Fig. 8(a) shows the precision, recall, f1- score, and accuracy for all 12 different activities, it's average accuracy is 85.17%. Fig. 8(b) shows the confusion matrix of the using Decision Tree classifier with test data to represent the type I (false positive) and II (false negative) classification error. Grid search with 05-fold cross-validation is used in Decision Tree with criterion='entropy', max_depth='9', splitter='best' chosen by the grid search as the best parameters. The average ROC-AUC score is 0.841. Fig. 8(c) shows the ROC-AUC score for each activity. Fig. 8(d) shows the Classification Accuracy of all activities using Decision Tree Classifier.

	precision	recall	f1-score	support
Cycling	0.84	0.88	0.86	233
Football	0.82	0.91	0.87	250
Jogging	0.11	0.12	0.12	8
JumpRope	0.57	0.59	0.58	58
Laying	0.99	0.99	0.99	142
Pushups	0.75	0.32	0.44	19
Sitting	0.99	0.98	0.98	207
Standing	0.87	0.91	0.89	22
Swimming	0.82	0.83	0.82	199
Walking	0.97	0.69	0.80	51
WalkingDownstairs	0.58	0.50	0.54	52
WalkingUpstairs	0.90	0.70	0.79	54
accuracy			0.85	1295
macro avg	0.77	0.70	0.72	1295
weighted avg	0.85	0.85	0.85	1295

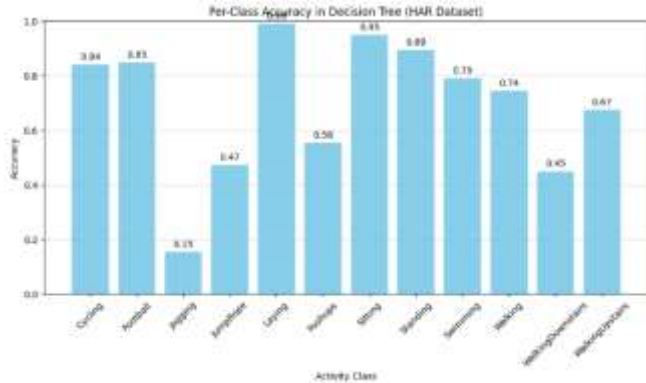
(a) Classification Report of Decision Tree Classifier



(b) Confusion matrix of Decision Tree Classifier



(c) ROC Curve of Decision Tree Classifier



(d) Classification Accuracy of all activities using Decision Tree Classifier

Fig.8: Performance using Decision Tree classifier

Comparison Between the Models

Across all five models, certain activities consistently stand out as strong performers, with **Laying, Sitting, and Standing** achieving very high precision, recall, and F1-scores above 91%. Activities like **Cycling, Football, and Swimming** also show stable results, with F1-scores typically ranging from the low 80s to mid-80s. However, some activities remain challenging: **Jogging** performs poorly across all models, especially for RBF SVM 2 (17% F1) and Decision Tree (12% F1), while **Pushups** show high variability, with RBF SVM 2 failing completely (0%) and others ranging between 44–62%. Similarly, **Walking Downstairs** is weak across models, particularly for the Unknown model (29% F1). At the model level, **RBF SVM** and **Logistic Regression** are the most

balanced, both achieving 86% accuracy and weighted F1-scores, though they struggle with Jogging and Jump Rope. **Decision Tree** performs strongly on stationary activities but falters on Jogging and Pushups, while the **KNN model** shows moderate performance with strengths in Football and Sitting but weaknesses in Walking Downstairs and Jogging. **Polynomial SVM** is the weakest overall, with 80% accuracy and a macro F1 of 60%, performing well only in Football recall and Standing but failing on Pushups and Jogging. All the classifier having various activities shown in comparison barchart in fig 9,10 and 11.

In terms of overall accuracy, the comparison shows that **RBF SVM** and **Logistic Regression** are the strongest performers, each achieving **86.00% accuracy** with balanced precision, recall, and F1-scores across most activities. The **Decision Tree** model follows closely with **85.00% accuracy**, excelling in stationary activities like Laying, Sitting, and Standing but struggling with Jogging and Pushups. The **KNN model** delivers moderate results at **82.00% accuracy**, showing strengths in Football and Sitting but weaknesses in Walking Downstairs and Jogging. Finally, **Polynomial SVM** is the weakest overall, with only **80.00% accuracy** and particularly poor performance on Pushups and Jogging, despite strong recall in Football and Standing. Thus, if accuracy is the primary metric, **RBF SVM and Logistic Regression emerge as the best models**, offering the most reliable and consistent performance across the dataset is shown in Table I.

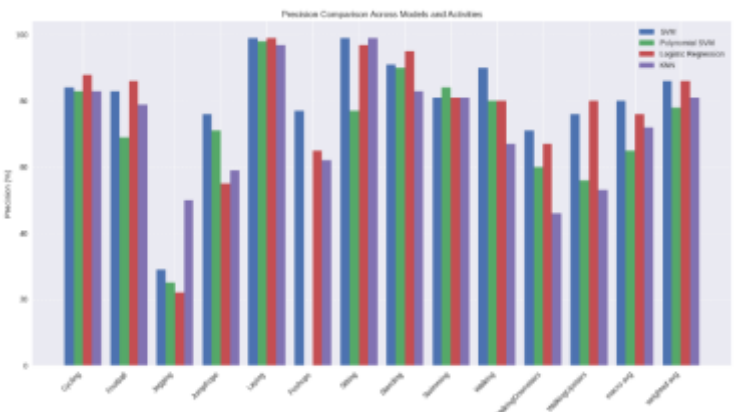


Fig. 9: Comparison chart for Precision

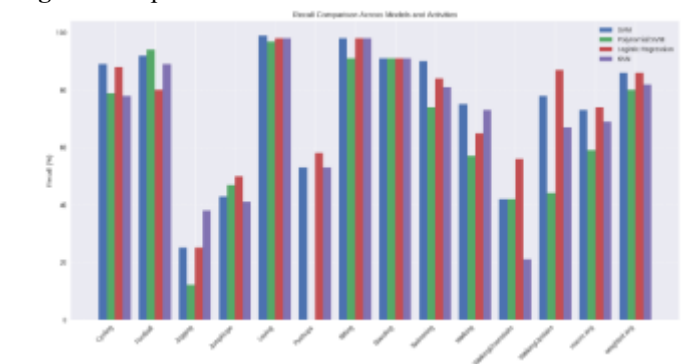


Fig. 10: Comparison chart for Recall

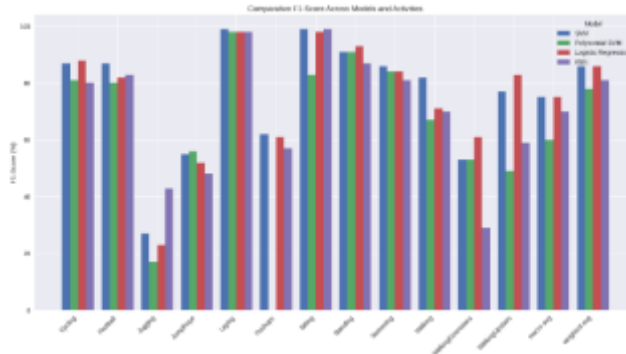


Fig. 11: Comparison chart for F1 Score

Table I. Classification performance of different machine learning methods on the Kaggle-HAR dataset of various activities.

Activity	SVM(RBF)			SVM(Polynomial Kernel)			Logistic Regression			KNN			Decision Tree		
	Precision (%)	Recall (%)	F1 Score (%)	Precision (%)	Recall (%)	F1 Score (%)	Precision (%)	Recall (%)	F1 Score (%)	Precision (%)	Recall (%)	F1 Score (%)	Precision (%)	Recall (%)	F1 Score (%)
Cycling	84.00 %	89.00 %	87.00 %	83.00 %	79.00 %	81.00 %	88.00 %	88.00 %	88.00 %	83.00 %	78.00 %	80.00 %	84.00 %	88.00 %	86.00 %
Football	83.00 %	92.00 %	87.00 %	69.00 %	94.00 %	80.00 %	86.00 %	80.00 %	82.00 %	79.00 %	89.00 %	83.00 %	82.00 %	91.00 %	87.00 %
Jogging	29.00 %	25.00 %	27.00 %	25.00 %	12.00 %	17.00 %	22.00 %	25.00 %	23.00 %	50.00 %	38.00 %	43.00 %	11.00 %	12.00 %	12.00 %
Jump Rope	76.00 %	43.00 %	55.00 %	71.00 %	47.00 %	56.00 %	55.00 %	50.00 %	52.00 %	59.00 %	41.00 %	48.00 %	57.00 %	59.00 %	58.00 %
Laying	99.00 %	99.00 %	99.00 %	98.00 %	97.00 %	98.00 %	99.00 %	98.00 %	98.00 %	97.00 %	98.00 %	98.00 %	99.00 %	98.00 %	99.00 %
Pushups	77.00 %	53.00 %	62.00 %	0.00 %	0.00 %	0.00 %	65.00 %	58.00 %	61.00 %	62.00 %	53.00 %	57.00 %	75.00 %	32.00 %	44.00 %
Sitting	99.00 %	98.00 %	99.00 %	77.00 %	91.00 %	83.00 %	97.00 %	98.00 %	98.00 %	99.00 %	98.00 %	99.00 %	99.00 %	98.00 %	98.00 %
Standing	91.00 %	91.00 %	91.00 %	90.00 %	91.00 %	91.00 %	95.00 %	91.00 %	93.00 %	83.00 %	91.00 %	87.00 %	87.00 %	91.00 %	89.00 %
Swimming	81.00 %	90.00 %	86.00 %	84.00 %	74.00 %	84.00 %	81.00 %	84.00 %	84.00 %	81.00 %	81.00 %	81.00 %	82.00 %	83.00 %	82.00 %
Walking	90.00 %	75.00 %	82.00 %	80.00 %	57.00 %	67.00 %	80.00 %	65.00 %	71.00 %	67.00 %	73.00 %	70.00 %	97.00 %	60.00 %	74.00 %
Walking Downstairs	71.00 %	42.00 %	53.00 %	60.00 %	42.00 %	53.00 %	67.00 %	56.00 %	61.00 %	46.00 %	21.00 %	29.00 %	58.00 %	50.00 %	54.00 %
Walking Upstairs	76.00 %	78.00 %	77.00 %	56.00 %	44.00 %	49.00 %	80.00 %	87.00 %	83.00 %	53.00 %	67.00 %	59.00 %	90.00 %	70.00 %	79.00 %

Accuracy	86.48%			80.15%			85.86%			81.77%			85.17%		
Macro Avg	80.00 %	73.00 %	75.00 %	65.00 %	59.00 %	60.00 %	76.00 %	74.00 %	75.00 %	72.00 %	69.00 %	70.00 %	77.00 %	70.00 %	72.00 %
Weighted Avg	86.00 %	86.00 %	86.00 %	78.00 %	80.00 %	78.00 %	86.00 %	86.00 %	86.00 %	81.00 %	82.00 %	81.00 %	85.00 %	85.00 %	85.00 %

5. CONCLUSION

In terms of overall accuracy, the comparison shows that **RBF SVM** and **Logistic Regression** are the strongest performers, each achieving **86.00% accuracy** with balanced precision, recall, and F1-scores across most activities. The **Decision Tree** model follows closely with **85.00% accuracy**, excelling in stationary activities like Lying, Sitting, and Standing but struggling with Jogging and Pushups. The **KNN model** delivers moderate results at **82.00% accuracy**, showing strengths in Football and Sitting but weaknesses in Walking Downstairs and Jogging. Finally, **Polynomial SVM** is the weakest overall, with only **80.00% accuracy** and particularly poor performance on Pushups and Jogging, despite strong recall in Football and Standing. We conclude that, if accuracy is the primary metric, **RBF SVM and Logistic Regression emerge as the best models**, offering the most reliable and consistent performance across the dataset.

We can explore the proposed method using various datasets and for more activities in the future. Finally, some other machine learning algorithms can be used to prove the effectiveness and we can use feature selection method.

ACKNOWLEDGEMENT

On behalf of all authors, the corresponding author states that there is no conflict of interest.

REFERENCES

- [1] S. M. Islam and K. H. Talukder, "Exploratory analysis of smartphone sensor data for human activity recognition," *IEEE Access*, vol. 11, pp. 1–10, 2023.
- [2] A. Dahou, M. A. Al-Qaness, M. Abd Elaziz, and A. M. Helmi, "MLCNNwav: Multilevel convolutional neural network with wavelet transformations for sensor-based human activity recognition," *IEEE Internet Things J.*, vol. 11, no. 1, pp. 820–828, 2023.
- [3] U. Saha, S. Saha, M. T. Kabir, S. A. Fattah, and M. Saquib, "Decoding human activities: Analyzing wearable accelerometer and gyroscope data for activity recognition," *IEEE Sens. Lett.*, vol. 8, no. 1, pp. 1–4, 2024.Y.
- [4] Y. Wang et al., "A novel deep multifeature extraction framework based on attention mechanism using wearable sensor data for human activity recognition," *IEEE Sens. J.*, vol. 23, no. 7, pp. 7188–7198, 2023.
- [5] M. K. Yi, W. K. Lee, and S. O. Hwang, "A human activity recognition method based on lightweight feature extraction combined with pruned and quantized CNN for wearable

device," *IEEE Trans. Consum. Electron.*, vol. 69, no. 3, pp. 657–670, 2023.

[6] S. Imanzadeh, J. Tanha, and M. Jalili, "Ensemble of deep learning techniques to human activity recognition using smart phone signals," *Multimed. Tools Appl.*, pp. 1–30, 2024.

[7] Y.-C. Lai, S.-Y. Chiang, Y.-C. Kan, and H.-C. Lin, "Coupling analysis of multiple machine learning models for human activity recognition," *Computers, Materials & Continua*, vol. 79, no. 3, pp. 3783–3803, 2024, doi: 10.32604/cmc.2024.050376.

[8] D. Garcia-Gonzalez, D. Rivero, E. Fernandez-Blanco, and M. R. Luaces, "New machine learning approaches for real-life human activity recognition using smartphone sensor-based data," *Knowledge-Based Systems*, vol. 279, p. 110260, 2023, doi: 10.1016/j.knosys.2023.110260.

[9] A. Logacjov, K. Bach, A. Kongsvold, H. B. Bårdstu, and P. J. Mork, "HARTH: A human activity recognition dataset for machine learning," *Sensors*, vol. 21, no. 23, p. 7853, 2021, doi: 10.3390/s21237853.

[10] A. Ferrari, D. Micucci, M. Mobilio, and P. Napolitano, "Trends in human activity recognition using smartphones," *Journal of Reliable Intelligent Environments*, 2021, doi: 10.1007/s40860-021-00147-0.

[11] S. Mekruksavanich and A. Jitpattanukul, "LSTM networks using smartphone data for sensor-based human activity recognition in smart homes," *Sensors*, vol. 21, no. 5, p. 1636, 2021, doi: 10.3390/s21051636.

[12] <https://www.kaggle.com/datasets/owenagius/inertia-sensors-for-human-activity-recognition/data>.

[13] Z. Bi, Y. Han, C. Huang, and M. Wang, "Gaussian naive Bayesian data classification model based on clustering algorithm," in *Proc. 2019 Int. Conf. Modeling, Analysis, Simulation Technologies and Applications (MASTA 2019)*, Atlantis Press, 2019, pp. 396–400.

[14] Subasi, A. *Practical machine learning for data analysis using python*; AcademicPress,2020.

[15] S. Li et al., "Text classification based on machine learning methods," 2019.

[16] R. Gove and J. Faytong, "Machine learning and event-based software testing: Classifiers for identifying infeasible GUI event sequences," in *Advances in Computers*, vol. 86, Elsevier, 2012, pp. 109–135.

[17] S. M. Basha and D. S. Rajput, "Survey on evaluating the performance of machine learning algorithms: Past contributions and future roadmap," in *Deep Learning and Parallel Computing Environment for Bioengineering Systems*, Elsevier, 2019, pp. 153–164.

[18] N. B. Amor, S. Benferhat, and Z. Elouedi, "Qualitative classification with possibilistic decision trees," in *Modern Information Processing*, Elsevier, 2006, pp. 159–169.

- [19] J. S. Richman, "Multivariate neighborhood sample entropy: A method for data reduction and prediction of complex data," in *Methods in Enzymology*, vol. 487, Elsevier, 2011, pp. 397–408.
- [20] D. Chanal, N. Y. Steiner, R. Petrone, D. Chamagne, and M. C. Péra, "Online diagnosis of PEM fuel cell by fuzzy C-means clustering,"
- [21] P. Nadkarni, "Core technologies: Data mining," in *Chapter 10*, 2016.
- [22] T. M. Ebbels, "Non-linear methods for the analysis of metabolic profiles," in *The Handbook of Metabonomics and Metabolomics*, Elsevier, 2007, pp. 201–226.
- [23] M. Vazan, E. Sharifi, H. Farahani, and S. Madadi, "Leveraging LDA feature extraction to augment human activity recognition accuracy," *Preprints*, 2024, doi: 10.20944/preprints202406.1528.v1.
- [24] S. Reshmi and E. Ramanujam, "An ensemble maximal feature subset selection for smartphone-based human activity recognition," *Journal of Network and Computer Applications*, vol. 236, 103875, 2024, doi: 10.1016/j.jnca.2024.103875.
- [25] S. Chakraborty, S. Sambhavi, and A. Nandy, "Deep learning in gait abnormality detection: Principles and illustrations," in *Bioinformatics and Medical Applications: Big Data Using Deep Learning Algorithms*, 2022, pp. 63–72.