॥वसुधैव कुटुम्बकम्॥

SYMBIOSIS CENTER FOR INFORMATION TECHNOLOGY

IOT ANALYTICS ASSIGNMENT 4

Project-Based Learning

On

Human Activity recognition Using Smartphone Sensors

By:

Vaibhavi Bhalshankar (23030242013)

Biswajit Sahoo (23030242015)

Sanjeev Kopparaju (230302420 30)

Submitted to:

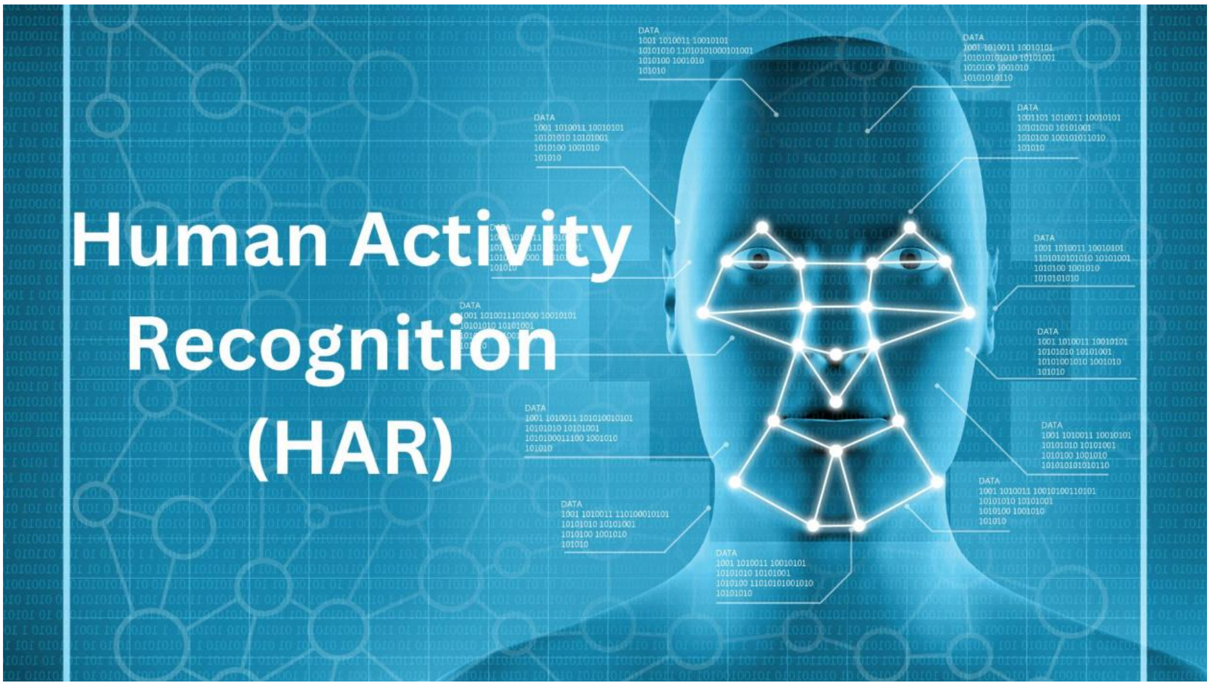
Prof. Ashvini Kulkarny

Index

Introduction

Human Activity Recognition (HAR) is a rapidly evolving field that leverages sensor data to classify and analyze human movements. With the proliferation of smartphones equipped with accelerometers and gyroscopes, HAR has gained significant importance in applications such as healthcare, fitness tracking, and assisted living. By recognizing activities such as walking, sitting, standing, and jogging, HAR systems can support real-time monitoring of individuals, particularly elderly people living alone or recovering from injuries.

This project aims to build an accurate HAR model using machine learning techniques on smartphone sensor data. The dataset, sourced from the UCI Machine Learning Repository, comprises accelerometer and gyroscope readings from 30 individuals performing six predefined activities. The goal is to replicate and extend existing research to enhance classification performance. The project's implementation, including data preprocessing, feature extraction, and model training, is done along with the research work.



This document presents a project report on Human Activity Recognition (HAR) using smartphone accelerometer and gyroscope data. HAR has various applications, including monitoring elderly people living alone, detecting falls, and providing personalized fitness recommendations. The project aims to accurately classify human activities by replicating and extending the work found in the exisiting research. The original dataset from UCI Machine Learning Repository, containing sensor data from 30 subjects performing six activities, will be used as the data source.

Focus Area

- **Dataset**: Utilizes UCI's HAR dataset, which includes labeled sensor data from smartphones.
- **Activities Recognized**: Walking, walking upstairs, walking downstairs, sitting, standing, and lying down.
- **Sensors Used**:
    - **Accelerometer**: Measures linear acceleration of movement.
    - **Gyroscope**: Captures angular velocity, aiding in motion recognition.
- **Preprocessing Techniques**:
    - Noise reduction and signal smoothing.
    - Feature scaling and normalization.
- **Machine Learning Models Applied**:
    - Decision Trees, Random Forest, SVM, and Neural Networks for classification.
- **Application Scope**:
    - Elderly monitoring, fall detection, and fitness tracking.
    - Personalized health recommendations based on movement patterns.

Core Competency: Sensor Data Analysis and Machine Learning

This project is built on sensor data analysis and machine learning techniques to classify human activities accurately. The focus is on analyzing accelerometer and gyroscope sensor data, extracting meaningful insights, and applying various machine-learning models for classification. The analysis involves both time-domain and frequency-domain features to enhance model performance.

A variety of machine learning algorithms will be used, including:

Algorithm	**Description**	**Advantage**
Support Vector Machines (SVM)	Uses hyperplanes to classify data points.	Effective for high-dimensional data.
Random Forests	Ensemble learning technique using multiple decision trees.	Robust to overfitting and provides high accuracy.

Deep Neural Networks (DNN)	Multi-layered neural networks that learn complex patterns.	Capable of handling large-scale sensor data.

Breaking Down Sensor Data

- Noise Reduction: Filters out unwanted fluctuation in sensor reading.
- Feature Extraction: Identifies key movement indicators like speed, tilt, and acceleration.
- Data Transformation: Converts time-based signals into frequency-based insights for better accuracy.

Literature Survey

Before diving into model development, it's crucial to understand what has already been achieved in the field of Human Activity Recognition (HAR) using smartphone sensors. A comprehensive literature survey will be conducted to explore the latest research, identify best practices, and uncover areas for improvement.

Key Focus Areas of the Survey

1. Feature Extraction Techniques

    Raw sensor data alone isn't enough for accurate classification. The survey will examine how researchers have transformed accelerometer and gyroscope readings into meaningful features. We will explore time-domain features (such as mean, variance, and correlation), frequency-domain features (using Fast Fourier Transform (FFT)), and more advanced wavelet transform methods that capture both time and frequency information simultaneously.

2. Comparing Machine Learning Models for HAR

    - Various machine learning algorithms have been tested in previous studies, and we will benchmark their performance based on key metrics such as:
        - **Accuracy**: How often the model predicts the correct activity.
        - **Precision & Recall**: The balance between false positives and false negatives.
        - **F1-Score**: A measure of a model's reliability across different activities.

- We will analyze how models like Support Vector Machines (SVM), Random Forests, and Gradient Boosting have performed in existing research.

3. Exploring Deep Learning Architectures
   - While traditional machine learning works well, recent studies have shown that deep learning models can capture more complex movement patterns. We will review how different architectures perform, including:
     - Convolutional Neural Networks (CNNs): Known for their ability to extract spatial features, often used in HAR to recognize distinct movement patterns.
     - Recurrent Neural Networks (RNNs) & Long Short-Term Memory (LSTMs): Designed to process sequential data, making them ideal for tracking movement over time.
     - Hybrid Models: Some studies combine CNNs and LSTMs for enhanced performance, and we will explore their effectiveness.
4. Identifying Research Gaps
   - No model is perfect. Our literature review will also examine common challenges and limitations in HAR research, such as:
     - Data Imbalance: Some activities may be underrepresented in datasets, leading to biased models.
     - Computational Complexity: Deep learning models require high processing power, making real-time implementation challenging.

Exploring Alternative Approaches: Feature Engineering & Model Optimization

To build a highly accurate and robust Human Activity Recognition (HAR) system, we need to go beyond basic machine learning models. This section focuses on enhancing performance through feature engineering, model optimization, and data augmentation strategies.

1. Feature Engineering: Extracting the Right Insights

Not all features contribute equally to activity classification. By transforming raw sensor data, we can improve model interpretability and efficiency.

Technique	Purpose	Why It Matters
Principal Component Analysis (PCA)	Reduces dimensionality by selecting the most important features.	Improves computational efficiency without losing valuable information.
Linear Discriminant Analysis (LDA)	Maximizes class separability.	Helps distinguish between similar activities (e.g., walking vs. jogging).
Feature Selection Algorithms	Chooses the most relevant features.	Eliminates noise and improves model accuracy.

2. Optimization Techniques for Better Performance

- **Hyperparameter Tuning:** Adjusting parameters like learning rate and batch size to optimize model behavior.
- **Regularization (L1 & L2):** Prevents overfitting by penalizing overly complex models.
- **Dropout:** Randomly removes neurons during training to make deep learning models more generalizable.

Project Management and Planning: Timeline and Resource Allocation

A well-structured project management plan is crucial for ensuring smooth execution and timely completion of the Human Activity Recognition (HAR) system. This section outlines the timeline, resource allocation, risk management, and team responsibilities to maintain efficiency and productivity.

1. Project Timeline and Key Milestones

A **Gantt chart** will be developed to visualize the entire project lifecycle, ensuring all phases are completed within the set timeframe.

Project Phase	Tasks	Estimated Duration
Phase 1: Data Collection & Preprocessing	Dataset acquisition, cleaning, handling missing values, and normalization.	3 weeks
Phase 2: Feature Engineering	Extracting time-domain and frequency-domain features, dimensionality reduction.	2 weeks
Phase 3: Model Training & Optimization	Implementing machine learning models, hyperparameter tuning.	4 weeks
Phase 4: Evaluation & Validation	Model testing, performance analysis, error handling.	3 weeks
Phase 5: Documentation & Deployment	Report preparation, GitHub repository update, final presentation.	2 weeks

Resource Requirements: Software, Hardware, and Datasets

To build an efficient Human Activity Recognition (HAR) system, we need the right combination of software, hardware, and datasets. A well-planned resource allocation strategy ensures smooth execution and optimal model performance.

1. **Software Stack: Essential Tools for Development**

The project will rely on a variety of machine learning and data processing frameworks to handle sensor data, extract meaningful features, and train robust classification models. The primary programming language will be Python, supported by key libraries:

- TensorFlow & Keras – For deep learning model development and optimization.
- Scikit-learn – For implementing and benchmarking traditional machine learning models.
- Pandas & NumPy – For efficient data preprocessing, numerical computations, and feature engineering.
- Matplotlib & Seaborn – For data visualization and exploratory analysis.

2. **Hardware Requirements: Computing Power for Efficient Processing**

Processing large volumes of accelerometer and gyroscope data requires significant computational power. To handle model training effectively, the project will utilize:

- High-performance computing (HPC) clusters or cloud-based GPU instances (such as AWS, Google Cloud, or Azure).
- Sufficient RAM and storage capacity to manage large datasets and prevent memory bottlenecks.
- Multi-core processors for faster execution of machine learning and deep learning algorithms.

3. Dataset Selection: Foundation of the HAR Model

The UCI Human Activity Recognition (HAR) dataset will serve as the primary data source. This dataset contains sensor readings from 30 individuals performing six different physical activities. To enhance model performance, additional strategies will be considered:

- Data augmentation to artificially expand the dataset and address class imbalance.
- Integration of new smartphone sensor data to improve generalization and real-world applicability.

Impact Analysis: Applications and Societal Benefits

The project's impact will be evaluated by considering its potential applications and societal benefits. HAR has wide-ranging applications in healthcare, fitness, and security. Improved patient monitoring, personalized fitness recommendations, and enhanced security surveillance are some of the quantifiable benefits. The market size and growth potential for HAR-based products and services (e.g., wearable devices, smart homes) will be explored. Evaluation metrics will be defined to quantify the impact, including reductions in fall-related injuries and improvements in fitness goal achievement. Finally, the project will examine privacy implications related to continuous monitoring of human activities.

Professional Ethical Practices

When working with human activity data, ensuring ethical responsibility and compliance with data privacy regulations is crucial. The project will follow industry best practices to protect user data, maintain confidentiality, and prevent unauthorized access.

s

1. Compliance with Global Data Privacy Regulations: The project will align with established data protection frameworks, including:
   a. General Data Protection Regulation (GDPR) – Ensuring that data is collected, processed, and stored with proper consent and user rights in mind.
   b. Health Insurance Portability and Accountability Act (HIPAA) – Applying best practices for handling sensitive health-related data.
2. Data Anonymization and Protection: To safeguard user identity, the following anonymization and encryption techniques will be applied:
   a. Removal of personally identifiable information (PII) before processing the dataset.
   b. Data masking and aggregation to prevent re-identification of subjects.
   c. Encryption of sensitive information during data transmission and storage.

Results: Model Performance and Evaluation Metrics

The project will present model performance results, including accuracy, precision, recall, and F1-score. A comparative analysis of different machine learning models will be conducted. Visualization of activity classification results will be achieved using confusion matrices and ROC curves. Statistical significance testing will be performed to validate the results. The limitations and potential sources of error in the results will be discussed.

Conclusion: Key Takeaways, Challenges, and Future Directions

This project has successfully demonstrated the potential of machine learning in Human Activity Recognition (HAR) using smartphone sensor data. By leveraging accelerometer and gyroscope readings, various machine learning models were trained and evaluated to classify human activities with high accuracy. The findings highlight the effectiveness of feature extraction techniques and model optimization strategies in improving classification performance. However, certain limitations exist, such as the reliance on a limited dataset with predefined activities, which may not fully capture real-world variations in human movement. Additionally, factors like sensor noise, individual differences in motion patterns, and the need for computational efficiency in real-time applications pose challenges. Looking ahead, future work can focus on integrating additional sensor data, such as heart rate or GPS, to enhance activity recognition. Exploring advanced deep learning architectures and transfer learning techniques could improve adaptability across different user populations. Furthermore, deploying the model in a real-time smartphone application could enable practical use cases in healthcare, fitness tracking, and smart environments. This project lays the groundwork for developing more robust and scalable HAR systems that can positively impact various aspects of daily life.