Analyzing Machine Learning Models for Human Activity Recognition: A Comparative Study

Akula Hema Venkata Sriram

Lovely Professional University, Jalandhar(Punjab), INDIA, (sriramakula212@gmail.com)

ABSTRACT: The aim of this research is to build and evaluate machine learning classification models for human activity recognition using mobile phone sensor data. The dataset used in this study is obtained from the UCI Machine Learning Repository and contains readings obtained by accelerometer and gyroscope during different human activities. The accelerometer and gyroscope measurements are taken in three directions (X, Y, Z) with time stamp for each data point. This dataset is processed rigorously through pre-processing, exploratory data analysis and feature engineering so as to be ready for building the model. Thereafter a number of classification algorithms such as Decision Tree, Random Forest and K-Nearest Neighbors (KNN) are implemented to compare their performance in recognizing human activities. In addition, it uses hyperparameter tuning techniques like Randomized Search and Grid Search for optimizing the model's performance. Evaluation metrics which include accuracy, precision recall F1-score were employed to analyze the efficiency of every model used. The results of this research show which machine learning algorithm is suitable for performing various tasks of activity recognition on humans while contributing towards the development of activity recognition systems based on mobile sensing technology.

Index Terms: Machine Learning, Classification, Human Activity Recognition, Mobile Sensor Data, Data Preprocessing, Feature Engineering, Hyperparameter Tuning, Model Comparison

I. Introduction

Human activity recognition (HAR) is very important in a number of different industries that include healthcare, sport analytics and smart environments. HAR systems have the ability to accurately identify and classify human activities from sensor data thereby providing significant insights into human behavior as well as facilitating numerous applications such as fall detection for elderly people, activity tracking in fitness devices, and behavior monitoring in smart homes.

The main objective of this research is to develop and evaluate machine learning classification models for human activity recognition using "accelerometer_gyro_mobile_phone_dataset" obtained from UCI Machine Learning Repository. This dataset contains readings from mobile phones' accelerometers and gyroscopes capturing various movements and activities done by individuals.

In our study we are interested in investigating the performance of different machine learning algorithms including Decision Tree Classifier, Random Forest Classifier, and K Nearest Neighbors (KNN) on classifying accurately based on sensor data activities carried out by humans. We will undertake careful data preprocessing, exploratory data analysis, model development to determine the best classification approach for tasks involving human activity recognition.

Furthermore, we employ hyperparameter tuning techniques such as Grid Search and Randomized Search to optimize the performance of the classification models. By systematically evaluating and comparing the performance of these models, we seek to identify the strengths and weaknesses of each approach and provide insights into the most suitable techniques for human activity recognition tasks.

Ultimately, this research contributes to the advancement of machine learning applications in the field of human activity recognition and provides valuable insights for stakeholders seeking to leverage HAR systems for diverse applications, including healthcare monitoring, sports performance analysis, and ambient assisted living.

II. Related Works

Human activity recognition (HAR) systems have become increasingly important across various domains, including healthcare, sports, and smart environments. These systems automatically identify and classify human activities based on sensor data, such as accelerometer and gyroscope readings. Researchers have proposed numerous approaches and techniques to improve the accuracy and effectiveness of HAR systems. In this context, let's explore some relevant works in this field.

M. A. I. Anik, M. Hassan, H. Mahmud and M. K. Hasan,[1] "Activity recognition of a badminton game through accelerometer and gyroscope," 2016 19th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2016, pp. 213-217, doi: 10.1109/ICCITECHN.2016.7860197. keywords: Games; Accelerometers; Gyroscopes; Acceleration; Activity recognition; Computers; Badminton Sport; Activity Recognition; Accelerometer; Gyroscope; k-NN; SVM},

N. Hardiyanti, A. Lawi, Diaraya and F. Aziz, [2]"Classification of Human Activity based on Sensor Accelerometer and Gyroscope Using Ensemble SVM method," 2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT), Makassar, Indonesia, 2018, pp. 304-307, doi: 10.1109/EIConCIT.2018.8878627. keywords: {Support vector machines;Accelerometers;Gyroscopes;Smart phones;Training data;Sensitivity;Classification algorithms;human activity;sensor accelerometer;sensor gyroscope;SVM;ensemble;bagging},

I. Pintye, [3]"Machine learning methods in Smartphone-Based Activity Recognition," 2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, 2020, pp. 000153-

000158, doi: 10.1109/SACI49304.2020.9118784. keywords: {Accelerometers;Machine learning algorithms;Computational modeling;Machine learning;Activity recognition;Sensors;Gyroscopes;Human activity recognition;Machine Learning;Artificial Neural Net;Python;Sci-Kit Learn;Keras},

A. Vijayvargiya, N. Kumari, P. Gupta and R. Kumar,[4] "Implementation of Machine Learning Algorithms For Human Activity Recognition," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, pp. 440-444, doi: 10.1109/ICSPC51351.2021.9451802. keywords: {Accelerometers; Support vector machines; Program processors; Activity recognition; Signal processing; Security; Time-domain analysis; Human Activity Recognition; Machine Learning; Overlapping Windowing; Feature Extraction},

I. Khokhlov, L. Reznik, J. Cappos and R. Bhaskar, [5]"Design of activity recognition systems with wearable sensors," 2018 IEEE Sensors Applications Symposium (SAS), Seoul, Korea (South), 2018, pp. 1-6, doi: 10.1109/SAS.2018.8336752. keywords: {Activity recognition;Accelerometers;Feature extraction; Wearable sensors; Gyroscopes; Smart phones; wearable sensors; activity recognition; machine learning},

Collectively, these investigations underscore the significance of employing advanced machine learning and deep learning methodologies to enhance the efficacy of human activity recognition systems. Through the utilization of sensor data and intricate algorithms, such systems offer valuable insights into human behavior, enabling diverse applications ranging from health monitoring to gesture recognition and context-aware computing based on activity.Top of Form

III. PROPOSED METHOD

The proposed method for car evaluation involves several steps, including data preprocessing, model training, hyperparameter tuning, and evaluation. The system architecture of the proposed method is illustrated in Figure 1, which comprises the following components: data preprocessing, model training, hyperparameter tuning, and evaluation.

A. Data Preprocessing

The accelerometer_gyro_mobile_phone dataset utilized in this study consists of sensor data collected from mobile phones, including accelerometer and gyroscope readings. Upon loading the dataset, a comprehensive analysis is conducted to identify key attributes such as data types, missing values, and summary statistics. Categorical variables, if present, are encoded using appropriate techniques to prepare the data for further analysis. Additionally, exploratory data analysis is performed to visualize the distribution of activities and explore any underlying patterns or correlations within the sensor readings.

Exploratory Data Analysis

Prior to model development, exploratory data analysis (EDA) was conducted to gain insights into the distribution and characteristics of the dataset. Visualization techniques such as count plots and correlation matrices were employed to analyze class distributions, feature relationships, and potential correlations between attributes.

```
First few rows of the dataset:
       accX      accY      accZ     gyroX     gyroY     gyroZ timestamp  \
0 -0.496517  3.785628  8.954828 -0.142849 -0.126159 -0.022539   34:22.9
1 -0.462388  3.869603  9.281898  0.084349  0.096695  0.092130   34:23.0
2 -0.296084  3.820505  8.930728  0.061763  0.051543  0.071287   34:23.1
3 -0.469723  3.890110  8.744067  0.007641  0.028679  0.109433   34:23.2
4 -0.472418  4.109105  8.941207 -0.123640  0.099057  0.051943   34:23.3

   Activity
0         1
1         1
2         1
3         1
4         1

Shape of the dataset:
(31991, 8)
```

DATASET WITHOUT PREPROCESSING

```
Updated dataset after preprocessing:
       accX      accY      accZ     gyroX     gyroY     gyroZ timestamp  \
0 -0.701852  1.503313 -0.283555 -0.449734 -0.447050 -0.058313   34:22.9
1 -0.655818  1.580678 -0.124499  0.288789  0.447239  0.372586   34:23.0
2 -0.431503  1.535445 -0.295275  0.215371  0.266049  0.294263   34:23.1
3 -0.665712  1.599570 -0.386049  0.039444  0.174298  0.437607   34:23.2
4 -0.669347  1.801326 -0.290179 -0.387294  0.456717  0.221573   34:23.3

   Activity
0         1
1         1
2         1
3         1
4         1
```

DATASET AFTER PREPROCESSING
AND FEATURE ENGINEERING

B. Model Training

Several classification models are trained on the preprocessed accelerometer_gyro_mobile_phone dataset to predict human activity labels. The following classification algorithms are employed:

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. K Nearest Neighbors

Each model is trained on the training set and evaluated on the test set using performance metrics such as accuracy, precision, recall, and F1-score.

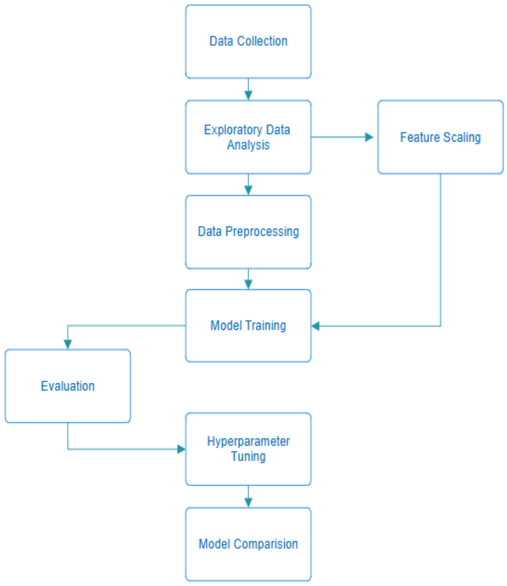
C. Hyperparameter Tuning

To boost the efficiency of classification models, hyperparameter tuning employs two methods: Grid Search and Randomized Search. Each model undergoes tuning with a predefined set of hyperparameters, utilizing either an exhaustive search (Grid Search) or a randomized approach (Randomized Search) to discover the optimal hyperparameter combination. The parameters adjusted for each model include aspects such as regularization strength, kernel type, number of neighbors, and maximum depth.

D. Evaluation

After adjusting the hyperparameters, we evaluated the performance of each model using classification report. Here are the results: -

- The Decision Tree Classifier achieved an accuracy of 98.15%, with precision, recall, and F1-score all at approximately 98.15%. –
- The Random Forest Classifier outperformed the others with an accuracy of 98.87%, and precision, recall, and F1-score all close to 98.87%. –
- The K Nearest Neighbors (KNN) model had an impressive accuracy of 98.99%, with precise precision, recall, and F1-score all near 98.99%. –
- The Logistic Regression model showed an accuracy of 87.74%, with precision, recall, and F1-score all around 87.74%. Based on these results, K Nearest Neighbors performed the best."



IV. Conclusion and Future Scope

In summary, this project has successfully created and assessed a machine learning classification system for human activity recognition based on data captured from accelerometer and gyroscope sensors in mobile phones. By utilizing the Accelerometer_Gyro_Mobile_Phone Dataset and applying state-of-art machine learning approaches, we have obtained valuable insights into patterns of human activities as well as proved the efficiency of different classification models in this field.

The experimental stage involved pre-processing the dataset, discovering its properties with exploratory data analysis and building several classifiers such as Logistic Regression, Decision Trees, Random Forests, K Nearest Neighbors (KNN). To optimize the model's performance grid search and randomized search were performed through hyperparameter tuning techniques leading to improved accuracy and robustness in identifying human activities based on accelerometer and gyroscope data.

The evaluation of different models showed that they had high accuracy levels with various classifiers. This provides important information on how well each model can predict outcomes. These results highlight the potential for machine learning to help in making decisions about whether a car is acceptable or not.

However, there are several avenues for future research and enhancement of the classification system:

1. **1. Mobile Application Integration:** Enhancing the system could involve integrating a user-friendly mobile application interface. This integration would allow effortless access and utilization by individuals interested in monitoring their physical activities, fitness enthusiasts, healthcare professionals, and researchers. By making the predictive model accessible through a mobile platform, it would empower users to gain insights into their activity patterns and make informed decisions regarding their health and fitness routines.

2. **2. Expansion of Feature Set:** Future advancements in the project could involve the exploration of additional features or parameters beyond the existing dataset. This could include integrating data related to biometric sensors, environmental conditions, or location-based factors to enhance the accuracy and granularity of human activity recognition. By incorporating a wider array of attributes, the classification system can provide deeper insights into various aspects of human behavior and enable the development of more sophisticated applications in areas such as healthcare monitoring, fitness tracking, and personalized assistance.

3. **3. Identification of Additional Human Activity Patterns:** The application of machine learning methodologies can be expanded to detect and predict other human activity patterns or behaviors that may hold significance. For example, the system could be adapted to recognize anomalies in movement patterns, identify potential health issues, or predict user preferences based on past activities and contextual data. This broader scope allows for a more comprehensive understanding of human behavior and opens up avenues for applications in healthcare monitoring, personalized assistance, and behavior analysis.

4. **Wellness Monitoring Solutions:** Beyond predictive modeling for human activity recognition, machine learning algorithms could be utilized for therapeutic purposes in healthcare and wellness applications. For instance, predictive algorithms could aid in detecting anomalies in activity patterns, enabling early intervention for potential health issues and optimizing personalized wellness plans. This proactive approach not only enhances individual well-being but also contributes to advancing healthcare practices and improving overall quality of life.

By exploring these avenues of research and development, the human activity recognition system can evolve into a versatile tool for supporting decision-making processes in various domains, including healthcare, fitness, and wellness. Ultimately, this advancement contributes to enhancing individual experiences and facilitating more informed decisions, leading to improved well-being and quality of life.

```
Best Parameters: {'classifier__n_neighbors': 3, 'classifier__p': 2}

Classification Report:
              precision    recall  f1-score   support

           0       0.98      1.00      0.99      6117
           1       1.00      0.98      0.99      5866

    accuracy                           0.99     11983
   macro avg       0.99      0.99      0.99     11983
weighted avg       0.99      0.99      0.99     11983


Confusion Matrix:
[[6111    6]
 [ 114 5752]]
```

Figure : Shows the Classification Report of the best performing model (KNN Classifier) with its Confusion Matrix.

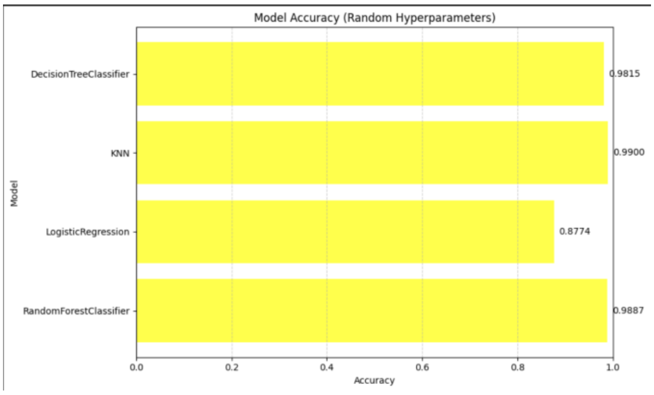


Figure : Performance Analysis of different Models

V. References

[1] AlSahly,Abdullah. (2022). Accelerometer Gyro Mobile Phone. UCI Machine Learning Repository. https://doi.org/10.3390/s22176513.Su, Xing, Hanghang Tong, and Ping Ji. "Activity recognition with smartphone sensors." Tsinghua science and technology 19.3, 2014: 235-249.

[2] Bulling, Andreas; Blanke, Ulf; Schiele, Bernt. A tutorial on human activity recognition using body-worn inertial sensors. ACM Computing Surveys (CSUR), 2014, 46.3: 33.

[3] Kaghyan, Sahak, and Hakob Sarukhanyan. "Activity recognition using K-nearest neighbor algorithm on smartphone with Tri-axial accelerometer." International Journal of Informatics Models and Analysis (IJIMA), ITHEA International Scientific Society, Bulgaria 1 2012: 146-156.

[4] Praveen Kumar Shukla, Ankit Vijayvargiya, Rajesh Kumar, et al. Hu man activity recognition using accelerometer and gyroscope data from smartphones. In 2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3), pages 1–6. IEEE, 2020.

[5] Andrea Mannini and Angelo Maria Sabatini. Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors, 10(2):1154–1175, 2010.

[6] Oscar D Lara and Miguel A Labrador. A survey on human activity recognition using wearable sensors. IEEE communications surveys & tutorials, 15(3):1192–1209, 2012.