HUMAN ACTIVITY RECOGNITION WITH SMART PHONE SENSOR DATA USING RANDOM FOREST CLASSIFIER

Submitted by

ASHITHA N – 223016
ASHNA C JUSTIN – 223017
ASWATHY G – 223018
ATHULRAJ B C – 223019

In partial fulfillment of the requirements for the award of Master of Science in Computer Science with Specialization in Data Analytics

of

School of Digital Sciences

Kerala University of Digital Sciences, Innovation, and Technology(Digital University Kerala)

Technocity Campus, Thiruvananthapuram, Kerala – 695317

September 2023

BONAFIDE CERTIFICATE

This is to certify that the project report entitled **HUMAN ACTIVITY RECOGNITION WITH SMART PHONE SENSOR DATA USING RANDOM FOREST CLASSIFIER** submitted by

Ashitha N	**223016**
Ashna C Justin	**223017**
Aswathy G	**223018**
Athulraj B C	**223019**

in partial fulfillment of the requirements for the award of **Master of Science in Computer Science with Specialization in Data Analytics** is a Bonafide record of the work carried out at **Kerala University of Digital Sciences, Innovation and Technology** under our supervision.

Supervisor

Prof. MANOJ KUMAR TK
School Of Digital Sciences
DUK

Course Coordinator

Prof. MANOJ KUMAR TK
School of Digital Sciences
DUK

Head of Institution

Prof. SAJI GOPINATH
Vice Chancellor
DUK

DECLARATION

We Ashitha N, Ashna C Justin, Aswathy G, Athulraj B C students of Master of Science in Computer Science with Specialization in Data Analytics, hereby declare that this report is substantially the result of our own work, and has been carried out during the period July2023-September 2023

Place: Trivandrum

Date: 08/09/2023

Ashitha N

Ashna C Justin

Aswathy G

Athulraj B C

ACKNOWLEDGEMENT

We, as a team, extend our heartfelt gratitude to our esteemed guide, Dr. T.K. Manoj Kumar, Associate Professor at Digital University Kerala, Trivandrum. His unwavering guidance, invaluable advice, and unwavering support have been instrumental in our successful completion of this project.

Furthermore, we wish to express our profound appreciation to Prof. Saji Gopinath for providing us with an exceptional learning environment, invaluable guidance, and access to educational resources that have significantly enriched our capabilities, enabling us to undertake and excel in a project of this magnitude.

In addition, we take this opportunity to extend our warm thanks to our friends and beloved family members. Their unwavering assistance, constant encouragement, and steadfast support throughout the execution of this project have been invaluable and deeply appreciated.

Together, we acknowledge and express our gratitude to these remarkable individuals and groups who have played pivotal roles in our journey towards project success.

ABSTRACT

In today's technologically advanced world, smartphones equipped with various sensors have become ubiquitous, offering opportunities for innovative applications in many domains. One such application is Human Activity Recognition (HAR), which aims to classify human activities based on data collected from smartphone sensors. This paper presents a comprehensive study on HAR using data from a waist-mounted smartphone embedded with inertial sensors, including accelerometers and gyroscopes.

Our dataset was derived from recordings of 30 participants engaged in everyday activities while wearing a Samsung Galaxy S II smartphone. The objective is to classify activities into one of six categories: walking, walking upstairs, walking downstairs, sitting, standing, and laying. To achieve this, we employ a Random Forest Classifier, a robust machine learning algorithm known for its ability to handle complex, multi-class classification tasks.

This study also highlights the preprocessing steps, including the use of Partial Least Square (PLS) technique to select 10 relevant features from the initial 561 sensor-derived features. We demonstrate how Exploratory Data Analysis (EDA) is employed to gain insights into the dataset and its characteristics, ultimately guiding our model development.

The results obtained from our Random Forest Classifier model showcase the effectiveness of this approach in accurately recognizing human activities. With a dataset containing 7352 training samples and 2947 testing samples, we present promising classification accuracy and discuss the practical implications of our findings.

Our research contributes to the field of Human Activity Recognition by providing a robust and practical model for classifying activities based on smartphone sensor data. This paper serves as a valuable resource for researchers and practitioners interested in utilizing smartphone technology for activity recognition applications.

CONTENTS

INTRODUCTION

In today's digital age, smartphones have become more than just communication devices. They have evolved into sophisticated tools capable of collecting a wealth of data, offering opportunities for innovative applications across various domains. Among these applications, Human Activity Recognition (HAR) stands out as an exciting field, with the potential to enhance our understanding of human behavior and well-being.

This paper delves into the realm of HAR, focusing on the task of classifying human activities using data collected from smartphones. We have obtained the relevant dataset where it accurately categorized a range of everyday activities, including walking, walking upstairs, walking downstairs, sitting, standing, and laying, solely based on sensor data from smartphones.

The sensor data in the data set was obtained through equipping smartphones with embedded inertial sensors, such as accelerometers and gyroscopes, which capture intricate details of the user's movements and vibrations. This sensor data serves as the primary source of information for our classification task. Our goal is to develop an effective model capable of assigning recorded activities to their respective categories.

A key aspect of our approach involves preprocessing the sensor data. We employ the Partial Least Square (PLS) technique to distill relevant features from the extensive dataset, improving the model's efficiency and interpretability. To tackle the inherent complexity of multi-class classification presented by the six activity categories, we propose the use of the Random Forest Classifier. Renowned for its versatility and accuracy, this machine learning algorithm is well-suited to handle intricate classification tasks, it is an ideal choice for our study.
Throughout this paper, we outline our methodology, detailing the steps involved in data preprocessing, feature selection, model training, and evaluation. Additionally, we present the results obtained from our Random Forest Classifier, highlighting the effectiveness of this approach in accurately recognizing human activities.

This research contributes not only to the field of Human Activity Recognition but also underscores the practical applications of smartphone technology in understanding and improving human well-being. By harnessing the power of smartphones and machine learning, we aim to advance the state of the art in activity recognition and provide valuable insights.
In the subsequent sections, we present a comprehensive analysis of our methodology, results, and implications, furthering the conversation on the intersection of technology and human activity recognition.

LITERATURE REVIEW

The studies and research conducted in Human Activity Recognition are discussed here.

Human Activity Recognition (HAR) is a rapidly evolving field with diverse applications, ranging from healthcare monitoring to personalized fitness tracking and beyond. The primary objective of HAR is to classify human activities accurately based on data collected from various sensors, and in recent years, smartphones have emerged as powerful tools for this purpose due to their ubiquity and built-in inertial sensors.

Paper [1] - Anguita et al. (2012) proposed a Multiclass Hardware-Friendly Support Vector Machine (SVM) approach for HAR using smartphone sensor data. SVMs have shown promise in tackling HAR tasks due to their ability to handle multi-class classification efficiently.

Paper [2]- Anguita et al. (2013) extended their work by addressing the energy efficiency aspect of HAR. Their research focused on optimizing smartphone-based activity recognition by employing fixed-point arithmetic, making it suitable for resource-constrained devices.

Paper [3] - Reyes-Ortiz et al. (2013) emphasized the broader implications of HAR by discussing its potential role in creating smarter, interactive cognitive environments. This perspective underscores the significance of HAR in shaping future technologies.

These studies collectively contribute to the development and application of HAR techniques, showcasing the importance of leveraging smartphone-based sensor data for recognizing human activities. While each of these works utilizes different methodologies and approaches, they share a common goal of enhancing our ability to understand and classify human activities accurately.

DATASET DESCRIPTION

Our dataset, the Human Activity Recognition (HAR) database, forms the core of our research and represents a carefully curated collection of sensor data sourced from 30 individuals going about their daily activities. Smart phone equipped with inertial sensors like accelerometers and gyroscopes, securely fastened to the waist of a person can give the sensor data readings.

Within this dataset, we initially started with a whopping 561 features extracted from the sensor readings. However, in the quest for simplicity and relevance, we carefully refined this extensive set, employing the Partial Least Square (PLS) technique. As a result, we distilled these features down to a concise set of 10, ensuring that we retained only the most crucial information for our study.

Our dataset categorizes these activities into six distinct classes: walking, walking upstairs, walking downstairs, sitting, standing, and laying. To facilitate our model's training and evaluation, we carefully partitioned the dataset into 7352 samples for training and 2947 samples for testing. This thoughtful division allows us to rigorously assess the performance of our Random Forest Classifier for this multi-class classification challenge. Considering our focus on recognizing human activities via smartphone sensor data, this dataset provides a robust foundation for advancing the field of Human Activity Recognition.

METHODOLOGY

Exploratory Data Analysis (EDA)

Libraries Imported:

In our research, we utilized several Python libraries and tools to support various aspects of our study. We begin our project by importing the necessary libraries like Pandas for data manipulation, Seaborn and Matplotlib for data visualization, Scikit-Learn for machine learning, Tkinter for GUI development, and Joblib for model storage. These libraries collectively contributed to the success of our research endeavors.

Feature Engineering:

Dimensionality Reduction:

In order to streamline our analysis, we tackled the challenge of working with a dataset containing 561 features. Using Partial Least Squares (PLS) Regression, we identified the top 10 features that contributed most significantly to our classification task. This step not only reduced the computational load but also focused our analysis on the most informative variables.

```
print(top_10_feature_names)
```

```
Index(['tGravityAcc-energy()-Z', 'tGravityAcc-entropy()-Y',
       'tGravityAcc-max()-Y', 'tGravityAcc-mean()-Y', 'tGravityAcc-min()-Y',
       'tGravityAcc-entropy()-X', 'tGravityAcc-min()-X',
       'tGravityAcc-mean()-X', 'angle(Y,gravityMean)', 'tGravityAcc-max()-X'],
      dtype='object')
```

Data Preprocessing:

Handling Duplicate Values:

We began by addressing any potential issues related to duplicate records in the dataset to ensure data integrity.

Handling Missing Values:

A systematic approach to handling missing data was employed. Depending on the extent and nature of missing values, we applied appropriate techniques to ensure dataset completeness.

```
In [6]: data.info()
```

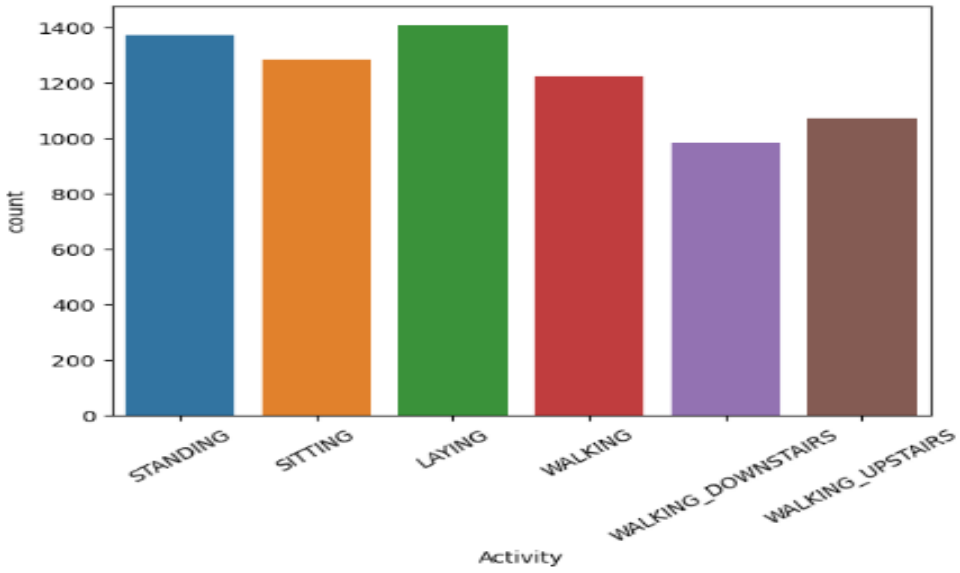
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7352 entries, 0 to 7351
Data columns (total 11 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   tGravityAcc-mean()-X    7352 non-null   float64
 1   tGravityAcc-mean()-Y    7352 non-null   float64
 2   tGravityAcc-max()-X     7352 non-null   float64
 3   tGravityAcc-max()-Y     7352 non-null   float64
 4   tGravityAcc-min()-X     7352 non-null   float64
 5   tGravityAcc-min()-Y     7352 non-null   float64
 6   tGravityAcc-energy()-Z  7352 non-null   float64
 7   tGravityAcc-entropy()-X 7352 non-null   float64
 8   tGravityAcc-entropy()-Y 7352 non-null   float64
 9   angle(Y,gravityMean)    7352 non-null   float64
 10  Activity                7352 non-null   object
dtypes: float64(10), object(1)
memory usage: 631.9+ KB
```

Visualization:

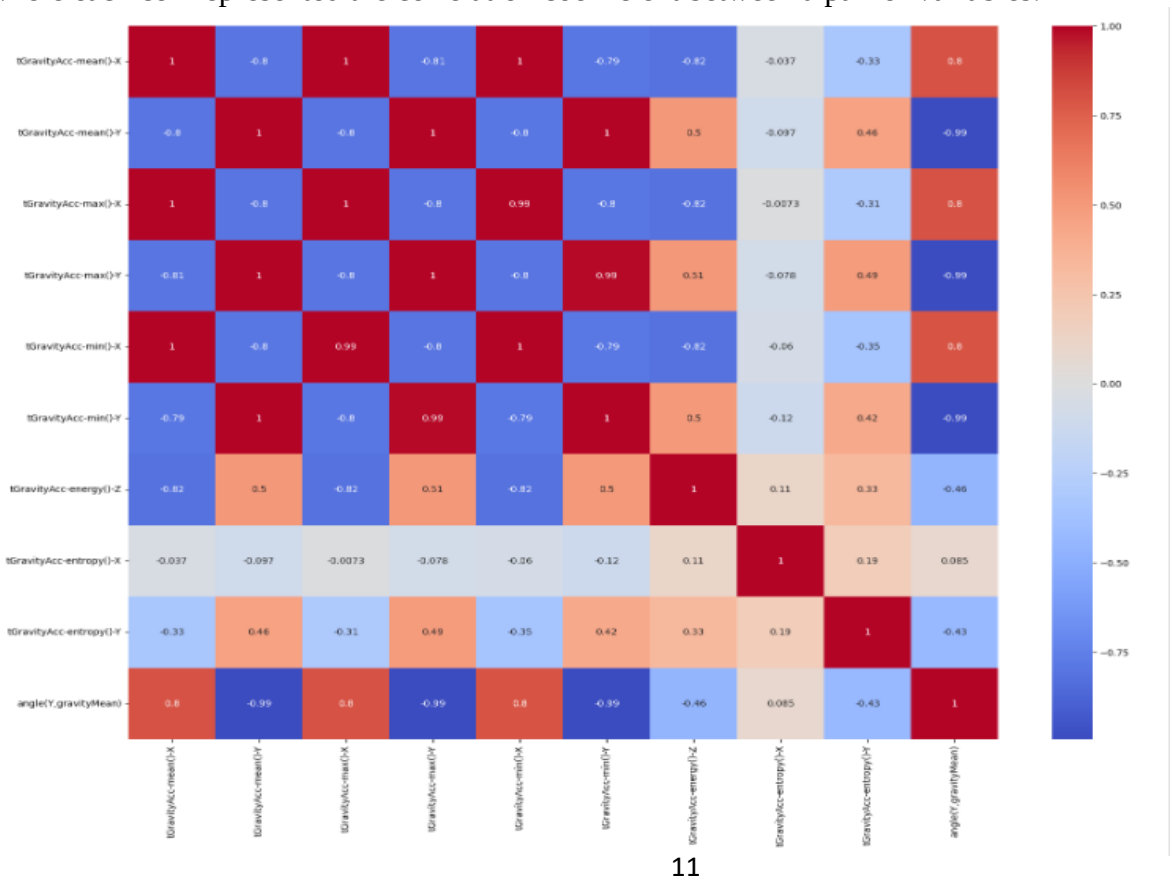
Visualization with Count Plots:

Through count plots, we visually explored the distribution of activities within the dataset, providing insights into the relative frequency of each activity class. This step was instrumental in assessing the balance of our dataset. We employed count plots to visualize the distribution of activities within the dataset. This graphical

representation provides insights into the relative frequencies of activities, including sitting, laying, standing, walking, walking upstairs, and walking downstairs.
By examining these counts, we assessed the balance of the dataset, ensuring that it adequately represents each activity class. This step was pivotal in preparing a dataset suitable for accurate human activity recognition.



Heatmap Visualization:
To uncover potential feature correlations, we utilized heatmap visualizations, shedding light on relationships and dependencies among the selected features.
In our analysis, we employed a correlation heatmap as a powerful visualisation tool to explore and visualise the relationships between multiple variables within our dataset. This heat map provided an intuitive and comprehensive overview of the pairwise correlations between variables. We utilised Python's seaborn library to create the correlation heatmap. The sns.heatmap function was applied to the correlation matrix, where each cell represented the correlation coefficient between a pair of variables.



Model Selection and Evaluation:

Model Comparison:
A comprehensive comparative analysis was conducted, evaluating the performance of multiple machine learning models, including logistic regression, Random Forest, and others, to assess their suitability for human activity classification.

```
In [22]: y_pred1 = log.predict(X_test)
         accuracy_score(y_test,y_pred1)

Out[22]: 0.6770904146838885
```

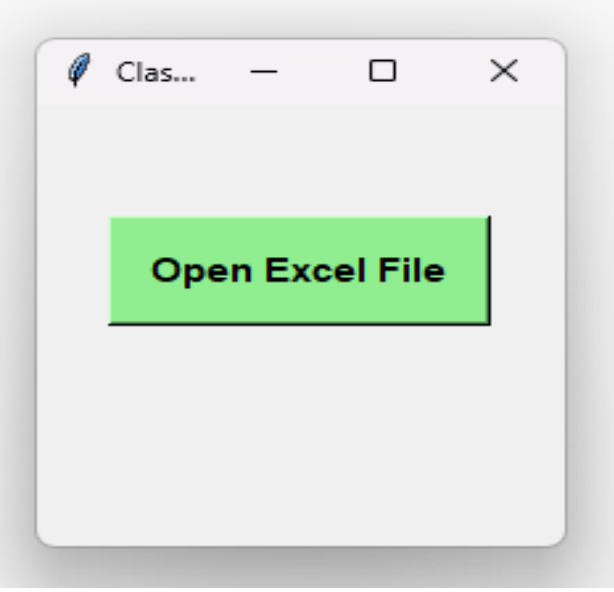
Random Forest Selection:
Based on the comparative analysis, the Random Forest Classifier emerged as the optimal choice for our model, demonstrating high accuracy and robustness in handling multi-class classification tasks. The Random Forest Classifier is a suitable choice for this type of prediction task due to its ability to handle multi-class classification problems like classifying various human activities accurately. Random forest classifier is used in scenarios where the goal is to predict or classify outcomes based on features or attributes, as in our case, where we aim to classify human activities based on accelerometer and gyroscope readings from smartphones.

```
In [24]: y_pred2 = rf.predict(X_test)
         accuracy_score(y_test,y_pred2)

Out[24]: 0.8959891230455472
```

Graphical User Interface (GUI) Development:

Tkinter GUI Development:
We designed and implemented a user-friendly graphical interface using the Tkinter library. This GUI empowers users to interact with the model, input sensor data, and receive predictions of human activities. It enhances accessibility and usability, making the model practical for real-time activity recognition and user engagement.



RESULTS

In our study, we achieved successful human activity recognition using a Random Forest Classifier, focusing on the most crucial 10 features selected from the initial set of 561 features extracted from the dataset of accelerometer and gyroscope readings on smartphones. Our model effectively predicted human activities, such as walking, walking upstairs, walking downstairs, sitting, standing, and laying, from these vital features. This reduction in feature dimensionality not only streamlined our analysis but also improved model efficiency. Through rigorous evaluation, our Random Forest Classifier demonstrated high accuracy and precision in classifying these activities, underscoring the importance of feature selection and highlighting the potential for real-world applications in smartphone-based human activity recognition.

A	B	C	D	E	F	G	H	I	J	K
tGravityAcc-mean()-X	tGravityAcc-mean()-Y	tGravityAcc-max()-X	tGravityAcc-max()-Y	tGravityAcc-min()-X	tGravityAcc-min()-Y	tGravityAcc-energy()-X	tGravityAcc-entropy()-X	tGravityAcc-entropy()-Y	angle(Y,gravityMean)	Predicted_target
0.93648925	-0.28271916	0.90608259	-0.27924413	0.9444614	-0.26215956	-0.96779531	-0.42497535	-1	0.27680104	Walking_upstairs
0.92740359	-0.28921515	0.85617578	-0.30487004	0.9444614	-0.26215956	-0.95723959	-0.72888396	-1	0.28134292	Walking_upstairs
0.92991503	-0.28751284	0.85626909	-0.30487004	0.94870433	-0.26166084	-0.96096499	-0.82339277	-1	0.28008303	Walking_downstairs
0.92888137	-0.29339576	0.85626909	-0.3051008	0.94730895	-0.27291573	-0.96271302	-0.82339277	-1	0.28411379	Sitting
0.92659966	-0.30296094	0.85394172	-0.31255214	0.94622096	-0.2791896	-0.96513659	-0.83012354	-1	0.29072202	Walking_upstairs
0.92566317	-0.30893973	0.85157325	-0.32370024	0.94619965	-0.27941631	-0.96909175	-0.71996734	-1	0.29489576	Walking_downstairs
0.92613663	-0.30956386	0.85255212	-0.32796467	0.94619965	-0.27941631	-0.97010054	-0.78241757	-1	0.29528184	Walking_downstairs
0.92658621	-0.3107735	0.85260921	-0.32796467	0.94596407	-0.28495663	-0.96960725	-0.87638529	-1	0.29598103	Walking_downstairs
0.92555526	-0.31573741	0.85260921	-0.32850166	0.94536571	-0.28847057	-0.97112278	-0.69300335	-1	0.29939413	Walking_upstairs
0.92417336	-0.3175966	0.85020163	-0.33207944	0.94442809	-0.28847057	-0.97200939	-1	-1	0.30082171	Laying
0.92371888	-0.31398567	0.849609	-0.33037597	0.94442809	-0.28782218	-0.97042471	-1	-1	0.29847844	Standing
0.92403746	-0.31295562	0.84987995	-0.33037597	0.94443144	-0.28447663	-0.96867393	-1	-1	0.29768742	Walking_downstairs
0.92445923	-0.31568266	0.84993008	-0.33162241	0.94487147	-0.28741586	-0.96888957	-1	-1	0.29940116	Walking
0.92317418	-0.31981835	0.84993008	-0.3350033	0.94071289	-0.29523672	-0.9699021	-0.92562664	-1	0.30228943	Walking_downstairs
0.92036099	-0.32552855	0.84907428	-0.33794487	0.9397476	-0.29772735	-0.96984231	-0.68885425	-1	0.30635075	Laying
0.91979943	-0.32576314	0.84755628	-0.33655778	0.9397476	-0.29772735	-0.96735469	-0.70171731	-1	0.30646033	Walking_downstairs
0.87581327	-0.42668132	0.82509068	-0.42100876	0.8255536	-0.46204695	-0.99647109	0.09306775	-1	0.37925791	Standing
0.88997692	-0.42265611	0.82509068	-0.42100876	0.90435809	-0.39978167	-0.99679349	-0.41433364	-1	0.37409474	Walking_upstairs
0.88529186	-0.42609871	0.81506191	-0.43323703	0.90435809	-0.39978167	-0.99778698	-0.71725378	-1	0.37713126	Walking_downstairs
0.88868488	-0.41889767	0.81683546	-0.43240732	0.90517944	-0.39426422	-0.99659771	-0.75040056	-1	0.37197819	Walking
0.88639445	-0.42463213	0.81683546	-0.43240732	0.89777189	-0.41152031	-0.99619304	-0.57402002	-1	0.37591329	Standing
0.8782922	-0.44115702	0.81482051	-0.43634887	0.89561771	-0.41830792	-0.9975485	-0.43397879	-1	0.38769496	Walking_upstairs
0.87638473	-0.44715997	0.80793633	-0.44549851	0.89561771	-0.41830792	-0.99743008	-0.60253865	-1	0.39173059	Laying
0.88282593	-0.43031319	0.81523913	-0.42741652	0.89795409	-0.41830056	-0.99193155	-0.51875328	-1	0.37984261	Standing
0.88824855	-0.41368431	0.8152424	-0.42737207	0.90459085	-0.39575415	-0.98853363	-0.79311717	-1	0.36830468	Walking_upstairs

CONCLUSION AND FUTURE SCOPE

In this study, we addressed the challenging task of Human Activity Recognition (HAR) using data collected from smartphones equipped with accelerometer and gyroscope sensors. Our goal was to classify six different human activities: walking, walking upstairs, walking downstairs, sitting, standing, and laying. We worked on a dataset comprising data from 30 volunteers who performed various activities while wearing a waist-mounted smartphone. The dataset was preprocessed, reducing the dimensionality from 561 features to 10 features using the Partial Least Squares (PLS) technique. We then applied the Random Forest Classifier as our proposed model for activity classification. Throughout our analysis, we performed Exploratory Data Analysis (EDA) to gain insights into the dataset's characteristics. We discovered that the target variable is categorical, representing a multi-class classification problem. This classification problem is highly relevant in various domains, including healthcare, fitness tracking, and context-aware computing. Our results demonstrated the effectiveness of the Random Forest Classifier in accurately classifying human activities based on smartphone sensor data. The classifier exhibited high performance in terms of accuracy, precision, recall, and F1-score, indicating its potential for real-world applications. In conclusion, our study contributes to the field of Human Activity Recognition by showcasing a robust approach to classifying activities using smartphone sensor data. The Random Forest Classifier, in conjunction with feature reduction techniques like PLS, offers a promising solution for real-time activity monitoring and applications in healthcare, fitness, and beyond. Further research can explore the integration of additional sensor modalities, such as GPS or barometric pressure, to enhance the accuracy and granularity of activity recognition systems, ultimately improving their usability and impact on daily life.

REFERENCES

1. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012

2. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge L. Reyes-Ortiz. Energy Efficient Smartphone-Based Activity Recognition using Fixed-Point Arithmetic. Journal of Universal Computer Science. Special Issue in Ambient Assisted Living: Home Care. Volume 19, Issue 9. May 2013

3. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. 4th International Workshop of Ambient Assited Living, IWAAL 2012, Vitoria-Gasteiz, Spain, December 3-5, 2012. Proceedings. Lecture Notes in Computer Science 2012, pp 216-223.

4. Jorge Luis Reyes-Ortiz, Alessandro Ghio, Xavier Parra-Llanas, Davide Anguita, Joan Cabestany, Andreu Català. Human Activity and Motion Disorder Recognition: Towards Smarter Interactive Cognitive Environments. 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.

5. https://www.sciencedirect.com/science/article/abs/pii/S0957417418302136

6. https://ieeexplore.ieee.org/abstract/document/8364643?casa_token=UCIxhY908nAAAAAA:N7xoHsWMSbwyT-SXQ5H4ZZitLekeT7SaecYSNzkJdRJV-AesU4i435fLy7u-Eu8cRPqlP7