ENHANCING HUMAN ACTIVITY RECOGNITION THROUGH MULTIMODAL ENSEMBLE: A FUSION OF TRADITIONAL AND DEEP LEARNING MODELS

Eureka D[1], Dr. Sridevi Ponmalar[2]

[1]*Department of Data Science and Business Systems, M.Tech - Data Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai - 603 203, Tamil Nadu, India*
[2]*Department of Computational Intelligence, Assistant Professor, SRM Institute of Science and Technology, Kattankulathur, Chennai - 603 203, Tamil Nadu, India*

---***---

Abstract - *Human activity recognition (HAR) is a challenging task that has many applications in domains such as patient-care, sports, health-care, security, elderly-care, and a variety of other applications. HAR aims to identify the activities of daily living (ADL) performed by a person using sensor data collected from wearable devices or smartphones. Our proposal for this project is an improved HAR system that integrates sensor data with the fusion of Ensembled Traditional ML models and DeepLearning techniques to achieve high accuracy and resilience. We compare each model's individual performance to the ensembled model. We used the HAR dataset from the UCI repository for our model. The data is collected using a waist-mounted smartphone with embedded accelerometer and gyroscope sensors to capture the 3-axial linear acceleration and angular velocity of the person(subject). Butterworth low-pass filter is applied to separate the body acceleration and gravity components of the acceleration signal. We use five classical machine learning classifiers: RandomForest, GradientBoosting, ExtraTreeClassifier, K-NearestNeighbors, SupportVectorMachine, as well as DeepLearning models: MultiLayer Perceptron, DNN. We evaluate our system on the publicly available HAR dataset constructed using recordings of thirty participants engaging in six activities of daily living: sitting, standing, lying, walking, walking upwards and downstairs. We compare the performance of our Ensembled system with the individual models' performance and show that our system achieves good accuracy. This system can be used for various HAR applications that require high accuracy and reliability. The next parts provide a brief overview of how our model works, the dataset, and potential enhancements.*

***Key Wors*: HAR, UCI repository, Ensemble, Traditional ML, Deep Learning, ADL**

1.INTRODUCTION

1.1 Importance of HAR in Modern Applications

The evolution of wearable devices has opened new frontiers for understanding human behavior. Recognizing daily activities from sensor data is crucial for applications ranging from healthcare to personalized user interfaces[1][12]. Our project addresses the challenge of creating a comprehensive HAR system that surpasses the limitations of existing methods. The 563 features collected in this experiment are derived from the sensor signals (accelerometer and gyroscope) of the smartphone worn by 30 subjects. Time domain characteristics and frequency domain features are the two groups into which they are separated.

Time domain features: These are obtained by applying various statistical functions to the sensor signals in each window of 2.56 seconds. Some of the functions are mean, standard deviation, maximum, minimum, correlation, etc. The time domain features also include the activity label and the subject identifier for each window.

Frequency domain features: These are obtained through the sensor signals' use of a Fast Fourier Transform (FFT) and then computing various spectral measures such as energy, entropy, frequency bands, etc. The frequency domain features also include the angle between the gravity vector and each of the sensor axes.

The total number of features is 561 (from the time and frequency domains) plus 2 (activity label and subject identifier), which makes 563 features in total.

1.2 Challenges in Existing HAR Systems

In the ever-evolving landscape of technological advancements, Human Activity Recognition (HAR) stands as a crucial domain with applications ranging from healthcare monitoring to smart environments[1][12]. However, the existing HAR systems encounter challenges in accurately identifying a diverse array of human activities, often due to limitations in model adaptability and collaboration between traditional and deep learning approaches.

1.3 A Novel Approach: Integrating Traditional ML and Deep Learning

This project report unveils a novel initiative dedicated to addressing these challenges head-on. Our primary objective is to enhance the adaptability and performance of HAR systems by seamlessly integrating the strengths of traditional machine learning (ML) models with the depth of neural networks. Unlike conventional approaches that typically focus on developing only Traditional ML models or Neural Network models, our innovative strategy involves ensembling the best-performing models from both paradigms.

Through meticulous experimentation and iterative refinement, our project introduces a hybrid ensemble model that brings together the predictive power of traditional ML models, renowned for their interpretability and generalization, with the nuanced understanding of complex patterns offered by state-of-the-art Neural Network models.

1.4 Contribution to Advancing HAR Technology

The previous research mostly explores building Traditional ML Models or Deep Learning models alone, or Ensemble of Traditional ML models or Deep Learning models alone. This project delves into the nuances of HAR, exploring the intricacies of human movement recognition and aiming to overcome the limitations observed in the current systems. This article will provide you with a comprehensive overview of the project's objectives, methodology, experimental setup, results, and conclusions and future enhancements. In our pursuit, we aim to contribute not only to the progression of the HAR system but also to broader discourse on harmonizing diverse Machine Learning methodologies for enhanced predictive capabilities, pushing the boundaries of HAR through a fusion of wisdom and innovation.

2. RELATED WORK

In the domain of Human Activity Recognition (HAR), various machine learning (ML) and deep learning (DL) models have been explored to accurately detect and classify human activities from sensor data. This section provides an overview of the related work conducted in our project area, focusing on the following models: Support Vector Machines (SVM), Random Forest, Decision Tree, Gradient Boosting, Naive Bayes, Multilayer Perceptron (MLP), and a hybrid designs that incorporate both long short-term memory and convolutional neural networks and so.

Support Vector Machines (SVM) [10][11]- Support Vector Machines have been extensively studied for HAR due to their ability to handle high-dimensional data and nonlinear decision boundaries. Previous works [10][11] have utilized SVM classifiers with handcrafted features extracted from accelerometer and gyroscope data to recognize activities with promising results of 96.33% accuracy [10] by gridsearch to get the best hyperparameters and compared with other ML models.

Gradient Boosting and Extreme Gradient Boosting (XGBoost) - known for their ensemble learning approach, have been applied to HAR with the aim of improving classification accuracy by combining multiple weak learners. XGBoost combines the strengths of gradient boosting with enhanced regularization techniques to achieve high performance in classification tasks. Previous research[13] has demonstrated the effectiveness of XGBoost in accurately recognizing human activities from sensor data, making it a popular choice in HAR applications.

Multilayer Perceptron (MLP) - Multilayer Perceptron, a fundamental architecture in artificial neural networks, has been utilized in HAR to learn hierarchical representations of sensor data. Previous studies[14] have demonstrated the effectiveness of MLP models in recognizing diverse human activities.The models developed achieved over 98% accuracy in recognizing various activities using data from wrist and ankle-worn devices.

Convolutional Neural Networks (CNN) + Long Short-Term Memory (LSTM) [7][9]- Hybrid designs that incorporate both long short-term memory and convolutional neural networks have emerged as effective HAR models capable of capturing spatial and temporal relationships in sensor data[7]. These models use CNNs to extract spatial characteristics from sensor readings and LSTM networks to simulate sequential patterns over time, yielding forefront performance in activity identification tests. [7] is LSTM-Conv model that uses UCI, WISDM and OPPORTUNITY dataset to evaluate its performance and achieved accuracy of 95.78%, 95.85% and 92.63%. [9] uses Conv-LSTM model and achieved accuracy of 91.6%

Deep LSTM Models - Deep LSTM architectures, consisting of multiple LSTM layers, have been investigated in HAR[6], demonstrating state-of-the-art performance in recognizing human activities with long-term dependencies and intricate temporal dynamics.

Ensemble Models[2][3][4][5][15] - Ensemble learning techniques, such as Stacking, Bagging, and Boosting, have also been employed in HAR to improve classification performance by combining predictions from multiple base models. The reference work [2] uses Voting Ensemble of

Logistic Regression, MLP, SVM, KNN, Gaussian Naive Bayes and Random Forest Models performed on the MHEALTH and USC-HAD datasets and achieved accuracy of 94.72% and 86.90 respectively. The reference work [3] uses the J48 decision tree, Multi-Layer Perceptrons (MLP) and Logistic Regression Ensemble model, where the model building is carried out on a WISDM dataset. Ensemble of Multiple CNN models have also been carried out in the HAR system and achieved accuracy of about 94% [4]. [5] uses Voting Ensemble of K-Means Clustering and PCA.

The related work in our project area encompasses a range of ML and DL models, each offering unique strengths and capabilities for HAR. While traditional ML algorithms like SVM, Random Forest, and Naive Bayes provide interpretability and robustness, deep learning architectures such as MLPs and CNN+LSTM hybrids offer superior performance in capturing complex activity patterns from raw sensor data.

3. DATASET DESCRIPTION

We obtained the HAR sensor data for our model from the UCI repository (open source repository).

3.1 Experiment Details

The Dataset was created by recording the activities of 30 subjects around the age of 19-48 performing six daily life activities with the smartphone sensors mounted on their waist to capture their movements. The smartphone captured the triaxial acceleration and angular velocity at 50Hz.

3.2 Sensor signals

Accelerometer and Gyroscope are the sensors used to capture the activities. The raw sensor signals obtained from the sensors are preprocessed using noise filters, and samples were taken in sliding windows with a 50% overlap and a duration of 2.56 seconds (128 readings per window).

Accelerometer - Measures linear acceleration forces. Collect data about the movement and orientation of the body.

Acceleration is the rate of change of velocity,

$$\alpha = \delta v/\delta t = \delta 2x/\delta t2$$

Gyroscope - Measures rotational motion. Position and change of orientation of an object.

3.3 Feature Vector

The Dataset provides 561 features with time and frequency domain variables for each window. The sensor signals' properties include mean(μ), standard deviation(σ), correlation, entropy and so on.

Time domain characteristics are produced by applying several statistical algorithms to sensor inputs within a 2.56-second frame. The functions include mean, standard deviation, maximum, minimum, correlation, and so forth. The time domain elements additionally contain each window's activity label and topic identification.

Frequency domain characteristics are derived by performing a Fast Fourier Transform (FFT) on sensor inputs and then computing spectral measurements like energy, entropy, frequency bands, and so on. The frequency domain characteristics also include the angle between the gravity vector and each sensor axis.

To calculate the number of output records (or feature vectors) obtained from this process, you can use the following formula:

Number of Output Records = Total Duration of Data / (Window Duration×Overlap)

Total Duration of Data = Total number of sensor readings×Sampling Period

Total number of sensor readings = Total Duration of Data / Sampling Period

3.4 Activity label and subject identifier

The dataset categorizes each observation window with one of six activities: walking, walking upstairs, walking downstairs, sitting, standing, or lying down. Additionally, the dataset includes information about the individual subject who performed each activity.

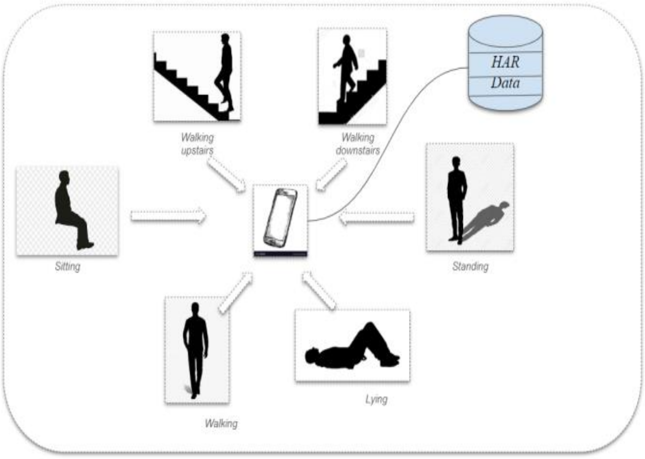

Fig -1: HAR activity data collection

The above Fig -1 shows the HAR activity data collected through smartphone sensors and stored in HAR Dataset.

4. METHODOLOGY OF RESEARCH

The proposed methodology for the Human Activity Recognition (HAR) system is as mentioned below.

Data Collection, Feature Analysis, Train-Test Split, Feature Transformation, Model Building (Ensemble Learning, Adaptive Fusion Learning), and Model Evaluation.

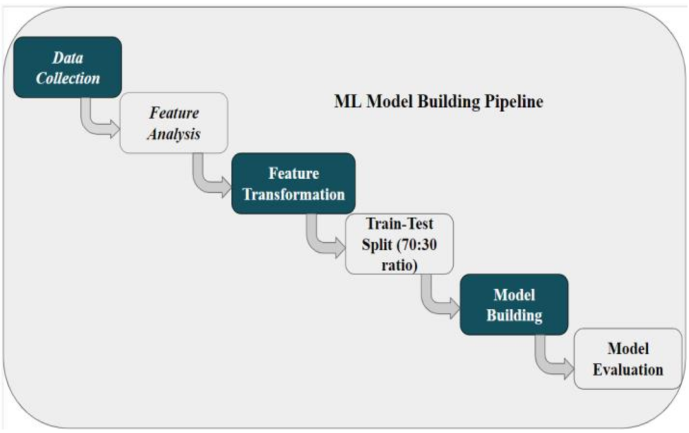

Fig -2: Model Building Pipeline

4.1 Data Collection

The HAR sensor data is collected from UCI repository, capturing a diverse set of human activities in various contexts. The dataset includes 561 input features, subject of activity, and target feature of six activities. The data is randomly split in a 70:30 ratio of train and test set ensuring a representative sample for training and evaluation.

4.2 Feature Analysis

Conducting feature analysis involves a meticulous examination of sensor data to extract relevant features. This phase aims to create a comprehensive feature set for subsequent modeling. The HAR Dataset does not contain any outlier or any missing values. On analyzing the dataset, it has been observed that the dataset exhibits a near-even distribution among the various classes, indicating a balanced representation. Fig.3. shows the visual representation of the activity distribution in our dataset.

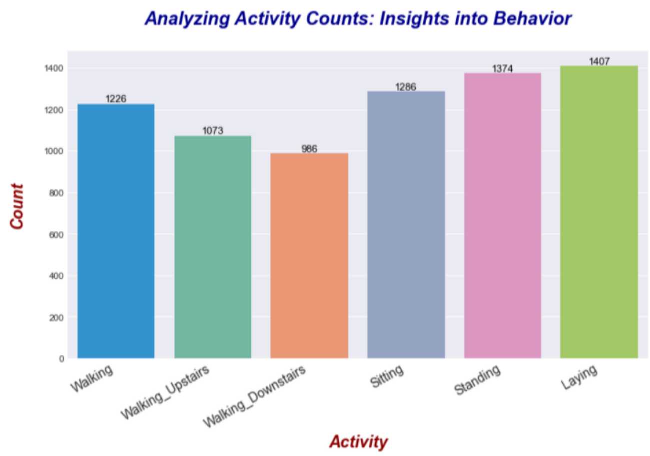

Chart -1: Activity Distribution Data

4.3 Train-Test Split

The dataset is partitioned and some amount of data is taken separately from the original set to evaluate models' performance. A stratified split ensures a balanced representation of activities in both sets, minimizing biases and enhancing generalization. The HAR dataset from UCI repository is split randomly in a train to test data ratio of 70:30, which we further separate the input and output features for the upcoming process.

4.4 Feature Transformation

Feature transformation is the process of changing or converting a dataset's current properties in order to increase its modeling applicability. We used the StandardScaler Transformation to scale all of the inputs from -3 to +3. The numerical features will be scaled by the Standard Scaler with a mean of 0 and a standard deviation of 1. This prevents features with larger scales from dominating the model training process and ensures that all characteristics contribute evenly to the model. Consider putting all numerical values on the same scale to guarantee that they are handled equally during model training.

The Activity has been labeled as follows,

Table -1: Activity Label

Label	Activities
1	Walking
2	Walking_upstairs
3	Walking_downstairs
4	Sitting
5	Standing
6	Laying

4.5 Model Building

Ensemble models mitigate overfitting and enhance generalization by leveraging the diversity of individual classifiers, thereby achieving higher accuracy and robustness in activity recognition tasks[2][3][4][5][15].

According to the findings, in order to enhance the application and the performance of HAR, Traditional Machine Learning (ML) models[10][11][13] and Deep Learning (DL) models[6][7][9][14] were trained individually, as well as an ensemble of Traditional Machine Learning models[5] or an ensemble of Deep Learning models[4] alone.

In our project, we propose ensembling the traditional machine learning approach with the cutting-edge deep learning model to extract characteristics more precisely and accurately. Combining traditional machine learning models with cutting-edge deep learning techniques through ensemble learning offers a potent strategy for enhancing performance and interpretability across various domains. By leveraging the feature extraction capabilities of deep learning alongside the interpretability and robustness of traditional models, ensembles can yield more accurate and nuanced predictions. This fusion not only improves model performance but also fosters better understanding of complex data relationships, making it a valuable approach for tackling real-world problems effectively. This study goes into the nuances of HAR, investigating the complexities of human movement identification and attempting to overcome the limits seen in existing systems.

A diverse set of models is developed, including Random Forest Classifier, Gradient Boosting Classifier, ExtraTrees Classifier, Support Vector Classifier(SVC), K-Neighbors Classifier, and MultiLayer Perceptron(MLP - a Neural Network model) and a Dense Neural Network Model(DNN).

We have built two Ensembling models, one with Stacking method and other with Voting Classifier method.

We Stacked the Random Forest Machine Learning Model with Dense Neural Network Model using Logistic Regression as the meta learner (with 1 input layer, 1 hidden layer and 1 output layer), and achieved an accuracy of 95%.

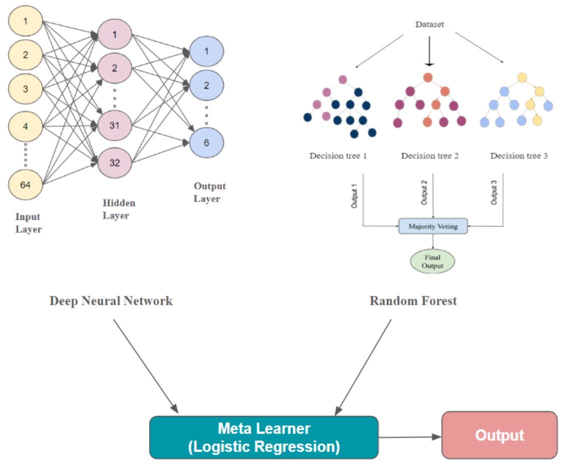

Fig -3: Stacking Random Forest and DNN with Logistic Regression Meta Learner

We applied the Voting Classifier method on RandomForest classifier, GradientBoosting classifier, ExtraTrees Classifier, SVC, KNeighbours (Traditional ML Models) and MultiLayer Perceptron with two hidden layers (Neural Network model).

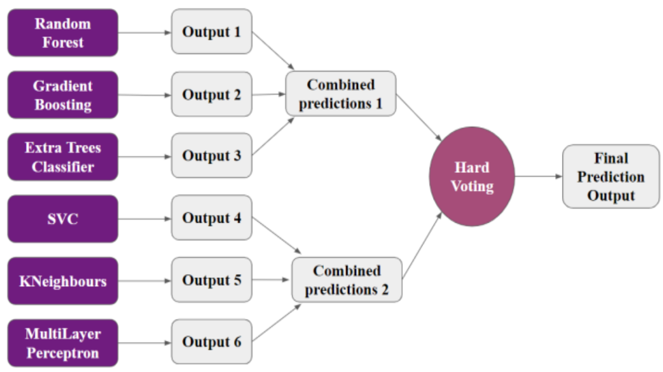

Fig -4: Ensembling Traditional ML with Neural Network

To Compare the Individual models' performance with the Ensembled models' performance, each model is individually trained on the HAR dataset with the training data and then evaluated the performance of each model with the testing data.

4.6 Ensemble Learning

Ensemble methods, such as Voting and Stacking, are implemented to combine the predictions of the best-performing traditional ML models. This ensemble is then compared against individual models to evaluate the effectiveness of combining different machine learning paradigms.

4.7 Model Evaluation

Models' performance is appraised using normative measurements such as accuracy, precision, recall, F1-score, and confusion matrices. Cross-validation is employed to ensure the robustness and reliability of results. The comparative analysis between traditional ML and deep learning models guides further refinement. Comparing the accuracy of other models,

	Precision_Score	Recall_Score	F1_Score	accuracy
RandomForest Classifier	0.925023	0.921078	0.922228	0.923990
DeepLearning Model	0.933311	0.927866	0.928971	0.928741
Ensemble Stack (ML + DL)	0.949775	0.949081	0.949308	0.950119
GradientBoosting Classifier	0.939851	0.937125	0.937977	0.938582
ExtraTrees Classifier	0.942229	0.936958	0.938284	0.940618
Support Vector Classifier	0.957464	0.955386	0.956106	0.956566
KNeighbours Classifier	0.897957	0.884119	0.886254	0.888700
MultiLayer Perceptron	0.948162	0.944992	0.945808	0.946047
Ensemble Voting (ML5 + DL1)	0.974708	0.974223	0.974400	0.974550

Fig -5: Performance Evaluation

The above Fig -6 shows the performance of ML models in an increasing order. Comparing the individual models' performance, we can see that the Voting Ensemble model gives better accuracy of about 97.45% and we can see that the Stacking Ensemble method gives 95.01% of accuracy.

5. RESULTS, CONCLUSIONS AND FUTURE WORKS

As per Fig-6 We can see that the Ensemble Model gives better accuracy results compared to the other individual models.

Fig -6: Performance improvement in plot

This research underscores the potential of hybrid ensemble methods in HAR, suggesting avenues for future exploration including alternative ensemble techniques, diverse datasets, and real-world deployment scenarios. Through a structured and rigorous approach, this study contributes to advancing the state-of-the-art in HAR methodologies, offering insights into the fusion of machine learning and deep learning paradigms. [8] references the challenges and approaches with vision based indoor activity recognition which can be taken into account for future works.

REFERENCES

[1] Yuemeng Cheng, Kan Wang, Hao Xu, Tangan Li, & Qinghui Jin, Daxiang Cui. "Recent developments in sensors for wearable device applications", Analytical and Bioanalytical Chemistry 13 Aug 2021.

[2] Nguyen, H.D. , Tran, K. P. 2, Zeng, X. , Koehl, L. and Tartare, G. "Wearable Sensor Data Based Human Activity Recognition using Machine Learning: A new approach", arXiv:1905.03809v1 [cs.LG] 9 May 2019.

[3] Cagatay Catal, Selin Tufekci, Elif Pirmit, and Guner Kocabag. "On the use of ensemble of classifiers for accelerometer-based activity recognition", Applied Soft Computing, 37:1018–1022, 2015.

[4] Narjis Zehra; Syed Hamza Azeem; Muhammad Farhan. "Human Activity Recognition Through Ensemble Learning of Multiple Convolutional Neural Networks", 2021 55th Annual Conference on Information Sciences and Systems (CISS) 10.1109/CISS50987.2021.9400290 IEEE 2021.

[5] Dinesh Kumar, Vishal Bharti, "Ensemble Classification for Human Activity Recognition with Feature Reduction", 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), pp.602-607, 2022.

[6] Tarun Kumar, Binod Prasad, "Prediction of Human Physical Activities With Consumer Smartphone Sensor Data", 2021 IEEE Bombay Section Signature Conference (IBSSC), pp.1-6, 2021.

[7] K. Xia, J. Huang and H. Wang, "LSTM-CNN Architecture for Human Activity Recognition", IEEE Access, vol. 8, pp. 56 855-56 866, 2020.

[8] Geetanjali Bhola, Dinesh Kumar Vishwakarma, "A review of vision-based indoor HAR: state-of-the-art, challenges, and future prospects", Multimedia Tools and Applications, 2023.

[9] Mohd Halim Mohd Noor, Sen Yan Tan, Mohd Nadhir Ab Wahab, "Deep Temporal Conv-LSTM for Activity Recognition", Neural Processing Letters, vol.54, no.5, pp.4027, 2022.

[10] Jakaria Rabbi, Md. Tahmid Hasan Fuad, Md. Abdul Awal, "Human Activity Analysis and Recognition from Smartphones using Machine Learning Techniques", arXiv:2103.16490v1 [cs.LG] 30 Mar 2021.

[11] Davide Anuita, 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 on Ambient Assisted Living IWAAL 2012: Ambient Assisted Living and Home care pp 216-223, 2012.

[12] Ong Chin Ann and Lau Bee Theng, “Human activity recognition: a review. In Control System, Computing and Engineering (ICCSCE)”, 2014 IEEE International Conference on, pages 389–393. IEEE, 2014.

[13] Celal Alagoz, “Comparative Analysis of XGBoost and Minirocket Algorithms for Human Activity Recognition”, 1st Bilsel International Sur Scientific Researches Congress, 10-11 February, 2024.

[14] Majidzadeh Gorjani, O.; Byrtus, R.; Dohnal, J.; Bilik, P.; Koziorek, J.; Martinek, R. “Human Activity Classification Using Multilayer Perceptron”. Sensors 2021, 21, 6207. https://doi.org/ 10.3390/s21186207.

[15] Abdulrahman Alruban, Hind Alobaidi, Nathan Clarke' Fudong Li, “Physical Activity Recognition by Utilising Smartphone Sensor Signals”, arXiv:2201.08688 January, 2022.