IMU-based Human Activity Recognition using Machine Learning and Deep Learning models

Saad AlKharji
Computer Engineering Department
Khalifa University
Abu Dhabi, United Arab Emirates
100049549@ku.ac.ae

Aysha Alteneiji
EBTIC (Emirates ICT Innovation Center), Khalifa University
Abu Dhabi, United Arab Emirates
aysha.alteneiji@ku.ac.ae

Kin Poon
EBTIC (Emirates ICT Innovation Center), Khalifa University
Abu Dhabi, United Arab Emirates
kin.poon@ku.ac.ae

2023 6th International Conference on Signal Processing and Information Security (ICSPIS) | 979-8-3503-2959-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSPIS60075.2023.10343567

***Abstract*—Research into Human Activity Recognition (HAR) with wearable sensors is attracting a lot of attention due to its wide range of applications. In this paper, we use a microcontroller with an integrated Inertial Measurement Unit (IMU) to design various Machine Learning (ML) and Deep Learning (DL) models. Five different activities were captured: walking, walking up a staircase, walking down a staircase, jumping, and falling. Both ML and DL models were trained on a variety of IMU data. A comparison was performed among the models based on their accuracy scores and confusion matrices to determine the most effective one. The proposed LSTM model achieved an accuracy score of 99% when trained with accelerometer and gyroscope data from the IMU.**

Keywords—Human Activity Recognition, IMU Sensor, Machine Learning, Deep Learning

I. Introduction

The field of Human Activity Recognition (HAR) has seen leaps in its research due to the increasing accessibility of related equipment including sensors and the wide reach of its applications [1]. The potential applications of HAR extend to several areas, such as healthcare, military, and security. Data for training Machine Learning (ML) models to recognize human activities are obtained either through visual based devices or wearable sensors [2]. Visual information from RGB images and skeletal joint orientations have been used to create 3D representation of activities [3]. In addition, wearable devices equipped with IMU sensors are widely employed for the classification of human activities [4].

However, each approach of HAR presents different challenges. Privacy issues are always a concern when implementing HAR through cameras. Moreover, while wearable sensors minimize the privacy issues and allow for non-intrusive data collection, optimal sensor placement has to be taken into consideration [1].

This paper is focused on comparing the effectiveness of Machine Learning (ML) and Deep Learning (DL) algorithms using IMU sensors for a Human Activity Recognition system that classifies daily movements. The four ML models for HAR are: Decision Tree Classifier, Support Vector Classifier (SVC), K Nearest Neighbors (KNN), and Gaussian Naïve Bayes.

A Decision Tree Classifier is a supervised ML model that is represented by a tree data structure [5]. Each node on the tree represents a choice presented to the algorithm and each edge shows the decision that was carried out. SVC is a type of Support Vector Machine (SMV) which works by using a hyperplane to split the classes into their respective groups. The different classes are separated by a margin which is maximized to reduce the classification error.

The KNN algorithm takes a sample data that needs to be classified and labels it based on the associated labels of its closest neighbors. The number of neighbors is represented by an integer K. The most common label from those K neighbors is computed and then assigned to the sample data. Furthermore, Gaussian Naïve Bayes assumes that every feature is independent in its ability to predict the output [6]. The normal distribution of every feature is used to calculate the likelihood that the output will belong to a certain label.

The framework proposed in this work revolves around:

- Exploring different variations of sensory data and their effects on the performance of the models.
- Conducting comparative analysis between different ML and DL models to discover the best suited model for HAR application. The model performance will be assessed based on the accuracy scores and confusion matrices.
- Evaluating the use of feature engineering and raw sensor data on the models.

II. Related Work

Machine Learning and Deep Learning models have been used extensively for Human Activity Recognition applications. Both approaches are subfields of Artificial Intelligence (AI) that involves algorithm development based on learning a particular pattern from a given set of data. However, ML models rely on handcrafted features and require mathematical techniques to extract meaningful features for different input types. On the other hand, DL models can automatically extract meaningful representations and patterns from raw data. For that reason, DL models typically require large amounts of training data whilst ML models work effectively with smaller handcrafted datasets.

In paper [4], the authors classified different human activities using an IMU sensor. They achieved a maximum accuracy of 96.9% when extracting 15 features using Random Forest feature selection and performing classification with a bagged ensemble classifier. In the result section, they pointed out that reducing the set of features to 9 still yielded a satisfactory accuracy of 96.3%.

The authors in [7] performed classifications of movements using the Handy and PAMAP2 public datasets with three ML models: Random Forest (RF), Bagged Decision Tree (DT), and Support Vector Model (SVM). The samples within the datasets were taken by an IMU sensor placed on the wrist of test subjects. The authors explored each possible combination of the IMU's accelerometer, gyroscope, and magnetometer.

Their results show that utilizing all three sensors lead to higher accuracy.

The study conducted by researchers in [8] utilized multiple wearable IMU sensors attached to their body. They placed the sensors on five different locations: right wrist, left wrist, waist, right ankle, and left ankle. They noticed that the most optimal configuration was placing two sensors: one on the right wrist and one on the right ankle. The raw data from the sensors was used to train a Long-Short Term Memory (LSTM) network. The study also highlighted how different types of activities can be recognized with greater accuracy depending on the type of sensory data used. For example, utilizing only the accelerometer resulted in the best performance for stationary activities which included standing, sitting, and lying. On the other hand, sensor fusion of the accelerometer, gyroscope, and magnetometer was better suited for activities that involved walking.

III. IMU-Based HAR Framework

The framework of the proposed HAR system is illustrated in Fig. 1. The framework consists of four parts: data acquisition, data pre-processing, model training and activity classification. The five different motions considered for this study were: walking on a flat surface, walking up a staircase, walking down a staircase, jumping straight up and falling down.

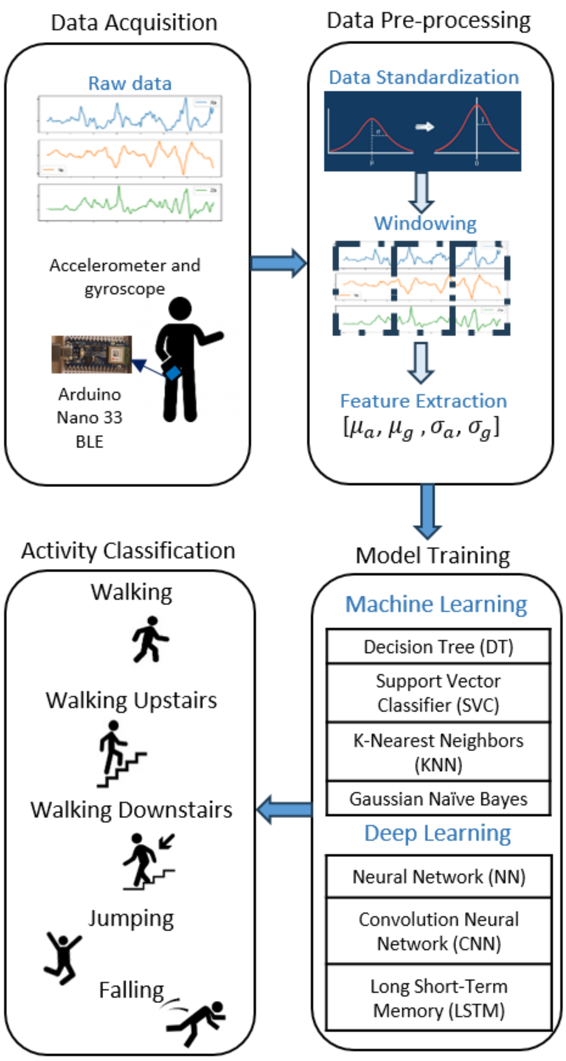


Fig. 1. IMU-Based HAR Framework

A. Data Aquisition

The hardware used to sample the five different motions is an Arduino Nano 33 BLE Sense which is equipped with the LSM9DS1 IMU sensor. HAR systems can accurately classify movements when working exclusively with accelerometer data [9]. However, it is a common practice to incorporate the accelerometer with the gyroscope [4] [10]. The addition of a magnetometer has been shown to be useful when classifying driving, sitting as a passenger in a moving car, and eating [8]. Since the activities of interest for this study do not involve the abovementioned activities, a magnetometer is not utilized in our case.

The Arduino board was placed inside the right pocket during the data capture process. In order to access the IMU sensor, the LSM9DS1 library (the built-in sensor module of the microcontroller) is required. The library has accelerometer and gyroscope sampling rates fixed at 104 Hz. The authors in [11] noted that sampling frequencies can range between 20 Hz and 250 Hz, which confirms that our sampling rate is within the accepted range. The training and testing datasets for each motion were collected separately. The training dataset was sampled for 96 seconds which resulted in 11469 samples. In addition, 24 seconds of sampling time for the testing dataset produced 2856 samples. Thus, the total sampling time for each motion was 120 seconds. The sampled data was stored in csv file format.

B. Data Pre-Processing

Window size for time series data can vary depending on the motion being studied as it can range from 0.08 seconds and up to 30 seconds [12]. Complex activities that involve hand gestures require longer window sizes compared to simpler activities that contain repetitive movements which require window sizes between 2 and 5 seconds [13]. For our application which involves simple activities, a window size of 3 seconds was chosen with an overlap of 50%. Fig. 2 to Fig. 4 illustrate a 3-second capture of raw acceleration and gyroscope data for the walking, falling and walking downstairs activities, respectively. The features considered for training the ML models were extracted from every window frame. In this case, the mean and standard deviation for the tri-axial accelerometer and gyroscope were extracted.

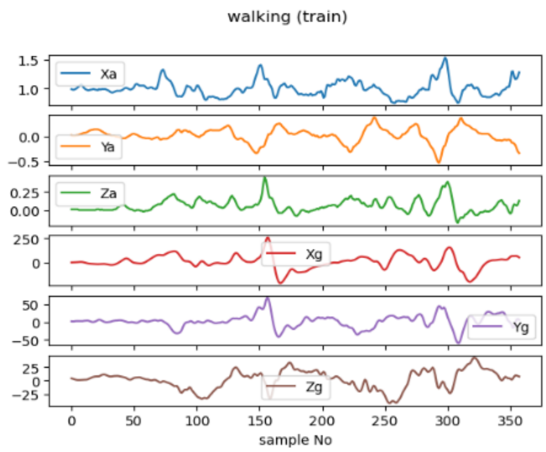


Fig. 2. Raw acc. and gyro. data for walking

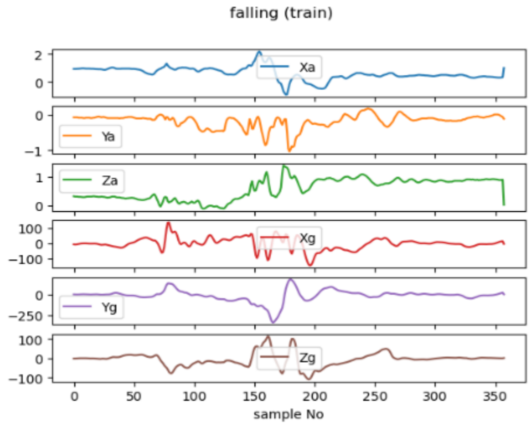


Fig. 3. Raw acc. and gyro. data for falling

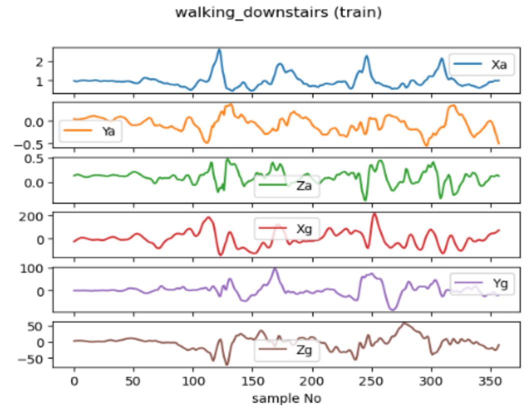


Fig. 4. Raw acc. and gyro. data for walking downstairs

Deep Learning models were trained on two variations of the IMU data. The first variation is the raw data extracted from the accelerometer and gyroscope. The second is the standardized raw data which is achieved by removing the mean and scaling each of the accelerometer and gyroscope data to have a unit variance.

C. *Model Training and Classification*

The four ML models and the three DL models were trained on the pre-processed data and raw data, respectively. Both the ML and DL models were trained using the accelerometer data, the gyroscope data and the combined data. In addition, the training for the DL models was repeated for the standardized raw data. The training accuracy and loss of the NN, CNN, and LSTM models are presented in Fig. 5a to Fig. 5c for both raw and standardized data.

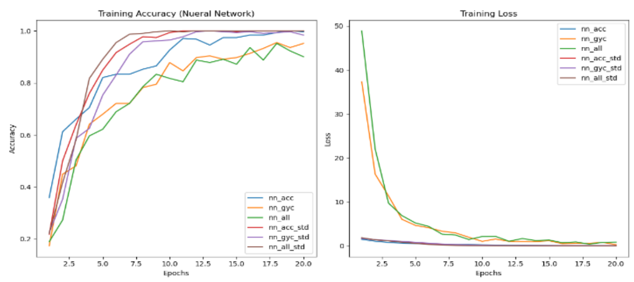


Fig. 5a. Training accuracy and loss of NN model

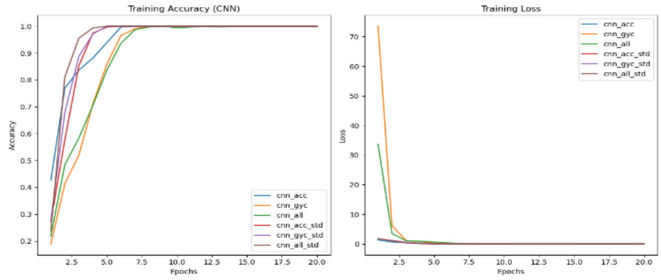


Fig. 5b. Training accuracy and loss of CNN model

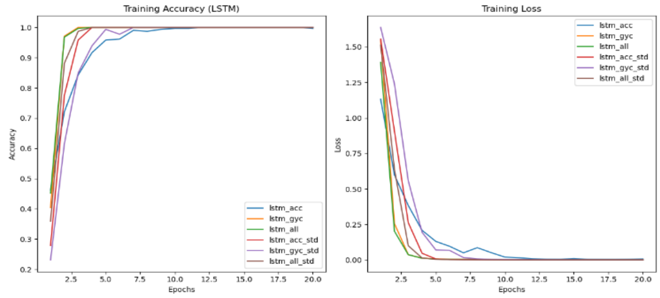


Fig. 5c. Training accuracy and loss of LSTM model

As can be seen from Figs. 5a to 5c, when standardization was applied to the raw data, the convergence rate became faster compared to the non-standardized data. In other words, standardization makes the model reach the optimal loss faster as illustrated in the training loss diagrams. This can be clearly observed in the loss values of the initial epochs for the NN and CNN models. This is due to the fact that standardized features have the same scale (i.e. mean of 0 and standard deviation of 1) which leads to a balanced learning rate.

IV. Result Analysis

A. *Machine Learning (ML)Models*

Based on the results shown in Table 1, it is clear that Gaussian Naïve Bayes outperforms all other models across all categories. It achieves the highest accuracy of 98.6% when using accelerometer and gyroscope data. Moreover, the DT model performs better with both sensors whilst the KNN works better with the accelerometer only. SVC was the worst performing model as it only obtained a maximum accuracy of 80%. When examining the effects of different sensory data, the only noticeable trend is that relying exclusively on the gyroscope leads to the lowest accuracies.

TABLE I. Accuracy Scores of ML Models

ML Models	**IMU Sensors**		
	Accelerometer and Gyroscope	***Accelerometer***	***Gyroscope***
Decision Tree Classifier	**92%**	85.3%	74.6%
SVC	76%	**80%**	**80%**
Gaussian Naïve Bayes	**98.6%**	90.6%	89.3%
KNN (n = 3)	81.3%	**89.3%**	76%

B. *Deep Learning (DL) Models*

In this study, three DL models were proposed with different input data. The aim is to find the model with the best performance. Fig. 6 shows the architectures of the DL models.

The proposed Neural Network (NN) model consists of 2 dense layers, each with 128 and 64 neurons, respectively. The second model is the Convolution Neural Network (CNN) which consists of an input layer, a total of two one-dimensional convolutional layers, followed by a dense layer with 100 neurons and an output layer. The two convolutional layers have 64 filters with a kernel size of 3. In addition, they are followed by a dropout layer to mitigate overfitting and a MaxPooling1D layer to shrink the network size. Convolutional layers extract temporal features by convolving the input with the kernel. The last model is an LSTM model, and it consists of an LSTM layer with 64 units followed by a dense layer with 32 neurons. The output dense layer in the three models consists of 5 neurons representing the five activities to be classified.

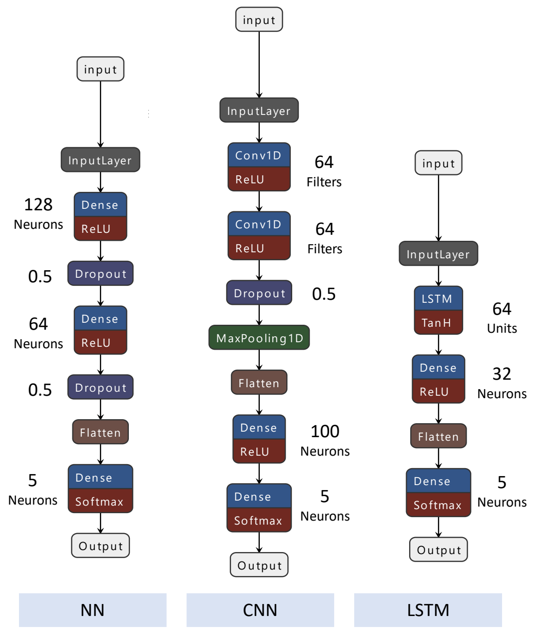


Fig. 6. The proposed DL architectures

The three DL models were trained with different input shapes based on the type of data extracted from the IMU sensor. For example, when the model is trained on accelerometer readings along the X,Y and Z axes, the feature number is set to 3, and when it is trained on both accelerometer and gyroscope data, the feature number is set to 6. The hyper-parameters used in training the Deep Learning models are summarized in Table II.

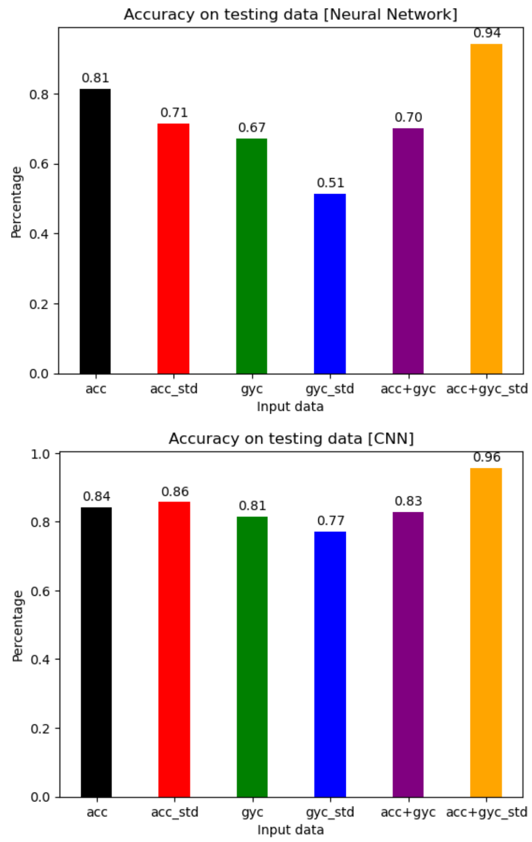
TABLE II. HYPER-PARAMETERS OF THE DL MODELS

Hyper-parameter	Value
Activation Function	ReLU for NN, CNN, LSTM, and Tanh for LSTM
Optimizer	Adam
Loss Function	Categorical Cross entropy
Drop-out Rate	0.5 (only NN and CNN)
Batch Size	32 (NN and CNN), 16 (LSTM)
Epoch	20

The performances of the trained DL models were evaluated using the testing datasets. The models were tested with accelerometer data, gyroscope data, and both accelerometer and gyroscope data. In addition, the standardization of each variant of the IMU data was tested against the non-standardized version. Diagrams in Fig. 7 summarize the accuracy of the aforementioned inputs in a bar chart for each of the DL models. As illustrated in the Fig. 7, The LSTM model outperformed the NN and CNN with almost all input types. In addition, when comparing the performance of NN and CNN, it can be seen that CNN performed better than NN in all input categories. The highest accuracy among all models was (99%) achieved by the LSTM using the standardization of accelerometer and gyroscope data, followed by the CNN (96%) and finally the NN (94%) both tested on the same input.

The effect of using different sensor data is pronounced in the bar charts. The results confirm that using both standardized accelerometer and gyroscope data produced the highest accuracy, and this behavior is consistent in the three models. However, when comparing performances of accelerometer and gyroscope data individually, the results vary among the three models and no conclusion can be drawn from the given charts.

The performance of individual raw accelerometer and raw gyroscope input data are higher in almost all models compared to their standardized versions, this is because each of the individual accelerometer and gyroscope feature data is maintained at a specific scale. Standardizing a single feature does not always result in better performance, but when dealing with multiple features (i.e. accelerometer and gyroscope together), standardization is crucial and can significantly boost the accuracy as demonstrated by the orange bars of the NN, CNN and LSTM models in comparison to the purple bars without standardization.



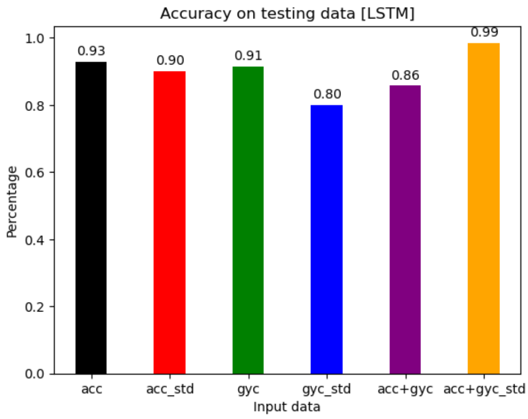


Fig. 7. The accuracy of the proposed DL models on different input

The confusion matrix of the best model is shown in Fig. 8. LSTM achieved better results for all activities. This is due to the following two main reasons. First, the LSTM is designed specifically to capture sequence of data over time which is an ideal case with HAR applications. Second, the feature extraction is automatically and implicitly achieved by the LSTM units making it a perfect model for raw data inputs.

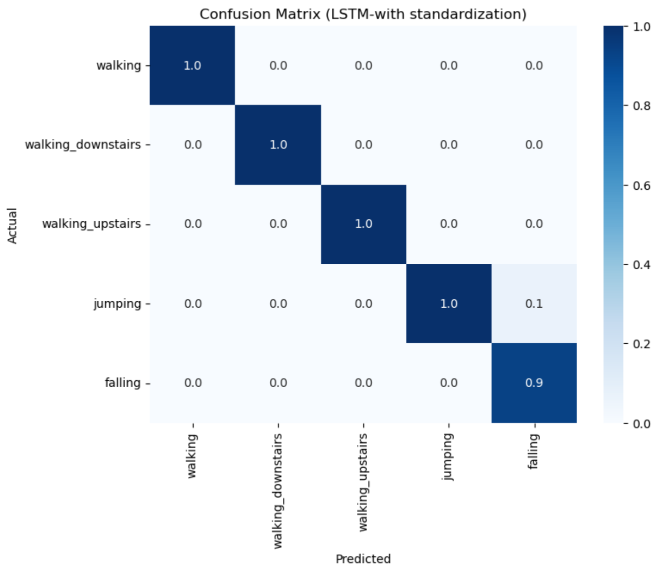


Fig. 8. Confusion Matrix of the best model (LSTM with standardization)

V. Conclusion and Future Work

In conclusion, this paper presented an effective Human Activity Recognition system that operates on IMU sensory data. This study compared the performance of several ML and DL models. Different categories of IMU sensor data were employed for model training. These include accelerometer data, gyroscope data and a combined set of accelerometer and gyroscope data. ML models were trained on statistical features (mean and standard deviation), whereas DL models were trained directly on the raw data.

Results show that the proposed LSTM model with standardized raw accelerometer and gyroscope data outperformed all the different variations of the ML and DL models proposed in this work. The results also indicate that standardization is effective in the case of multiple features. However, no conclusive finding is drawn from the case of a single feature as it purely depends on the feature itself.

Future work on this topic includes deploying the classification model onto a microcontroller board in the form of AI on the Edge. This requires taking multiple factors into consideration such as the model size and complexity, quantization and the resource constraints of the microcontroller. In addition, more activities can be sampled in order to cover a wider range of applications.

References

[1] C. Jobanputra, J. Bavishi and N. Doshi, "Human Activity Recognition: A Survey," Procedia Computer Science, vol. 155, p. 698–703, 2019.

[2] S. Aarthi and S. Juliet, "A Comprehensive Study on Human Activity Recognition," in 2021 3rd International Conference on Signal Processing and Communication (ICPSC), 2021.

[3] A. Franco, A. Magnani and D. Maio, "A multimodal approach for human activity recognition based on skeleton and RGB data," Pattern Recognition Letters, vol. 131, p. 293–299, 3 2020.

[4] O. Dehzangi and V. Sahu, "IMU-Based Robust Human Activity Recognition using Feature Analysis, Extraction, and Reduction," in 2018 24th International Conference on Pattern Recognition (ICPR), 2018.

[5] B. Mahesh, "Machine Learning Algorithms -A Review," International Journal of Science and Research (IJSR), pp. page: 381-386, 2020.

[6] J. Shen and H. Fang, "Human Activity Recognition Using Gaussian Naïve Bayes Algorithm in Smart Home," Journal of Physics: Conference Series, vol. 1631, p. 012059, 9 2020.

[7] A. Ayman, O. Attalah and H. Shaban, "An Efficient Human Activity Recognition Framework Based on Wearable IMU Wrist Sensors," in 2019 IEEE International Conference on Imaging Systems and Techniques (IST), 2019.

[8] S. Chung, J. Lim, K. J. Noh, G. Kim and H. Jeong, "Sensor Data Acquisition and Multimodal Sensor Fusion for Human Activity Recognition Using Deep Learning," Sensors, vol. 19, p. 1716, 4 2019.

[9] Y. Chen and Y. Xue, "A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer," in 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015.

[10] A. Barna, A. K. M. Masum, M. E. Hossain, E. H. Bahadur and M. S. Alam, "A study on Human Activity Recognition Using Gyroscope, Accelerometer, Temperature and Humidity data," in 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019.

[11] E. Ramanujam, T. Perumal and S. Padmavathi, "Human Activity Recognition With Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review," IEEE Sensors Journal, vol. 21, p. 13029–13040, 6 2021.

[12] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors and Tutorials" IEEE Communications Surveys, vol. 15, p. 1192–1209, 2013.

[13] M. Shoaib, S. Bosch, O. Incel, H. Scholten and P. Havinga, "Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors," Sensors, vol. 16, p. 426, 3 2016.