Recognizing Human Activities through Deep Learning: A Convolutional Neural Network Approach

B. SANTOSH KUMAR

Associate Professor, Department of MCA, Wesley PG College, Secunderabad, India.

Abstract - Human activity recognition (HAR) plays a pivotal role in a myriad of applications, ranging from healthcare monitoring to smart environments. This paper explores the application of Convolutional Neural Network (CNN) deep learning models for the purpose of human activity recognition. CNNs, renowned for their effectiveness in image-related tasks, are employed to analyze temporal sequences of sensor data, providing a novel approach to capturing and interpreting complex patterns associated with human activities. The study commences with the preprocessing of sensor data, transforming raw input into a format suitable for training a CNN model. Temporal dependencies within the data are preserved, allowing the model to discern nuanced patterns indicative of various human activities. The CNN is then trained on labeled datasets, encompassing diverse human activities, to learn and generalize patterns inherent in the sensor data. The model's architecture is optimized to balance complexity and efficiency, ensuring its effectiveness in real-time applications.

Keywords: Human activity recognition (HAR), Convolutional Neural Network (CNN), Deep learning models, Sensor data preprocessing, Temporal sequences.

I. INTRODUCTION

In the field of artificial intelligence and pattern recognition, Human Activity Recognition (HAR) is a major endeavor that seeks to identify and categorize human activities or behaviors from sensor data inputs [1]. The use of computational methods covers several fields, including healthcare, sports analytics, security systems, and human-computer interaction. The significance of precisely identifying and comprehending human actions has generated considerable attention owing to its potential to improve customized services, healthcare surveillance, and the development of intelligent systems [2]. Nevertheless, despite notable advancements in approaches for Human Activity detection (HAR), there are still persistent difficulties that need the development of creative solutions to ensure reliable and precise detection.

The complexities inherent in human behavior provide significant difficulties within the domain of activity recognition. The presence of diverse human behaviors, contextual influences, and the ever-changing character of real-world situations provide challenges in attaining reliable and accurate categorization [3]. Furthermore, the use of conventional machine learning methods often faces constraints in accurately capturing the intricate temporal and spatial connections present in sensor data with a large number of dimensions. Therefore, there is an urgent demand for advanced approaches that can effectively manage this inherent unpredictability and extract significant patterns from many data sources in order to achieve precise activity detection [4].

Deep Neural Networks (DNNs) have emerged as a promising solution for resolving the complexities connected with Human Activity Recognition [5]. The intrinsic capacity of deep neural networks (DNNs) to independently acquire hierarchical representations from unprocessed input makes them a tempting option for problems related to activity detection [6]. These networks have a hierarchical structure with linked nodes, allowing them to analyze complex patterns and abstract aspects from sensor data [7]. As a result, they are capable of capturing subtle temporal and spatial correlations. The capacity of deep neural networks (DNNs) to effectively adjust and comprehend intricate patterns embedded within the data allows them to overcome the difficulties arising from the diversity and intricacy of human activities. Consequently, DNNs provide a powerful resolution to the issue of human activity recognition (HAR).

In this context, the implementation of deep learning models that have been tuned particularly to the domain of HAR has a substantial amount of potential. Researchers and practitioners use DNNs because of their versatility and representation learning power. This enables them to develop unique architectures, maximize model performance, and explore creative ways for feature extraction and fusion. Leveraging the power of deep neural networks to good use is a step in the right direction toward the goal of precise, real-time human activity identification [8]. This would have consequences for a wide range of industries and fields, paving the way for revolutionary advances in intelligent systems.

This research investigates Deep Neural Networks for Human Activity Recognition, including new designs, optimization methods, and feature extraction and fusion methodologies. DNNs are used to improve activity recognition systems' accuracy, efficiency, and real-time performance. This study explores HAR-specific deep learning approaches to enhance intelligent systems that can properly recognize and comprehend human behaviors across applications and domains.

II. LITERATURE

Wen Qi et al [9] proposed the introduction of a novel and effective framework known as Fast and Robust Deep Convolutional Neural Network (FR-DCNN), specifically designed for the purpose of human activity recognition (HAR) using smartphone data. The effectiveness of the raw data collected from inertial measurement unit (IMU) sensors is enhanced by the integration of several signal processing algorithms and a signal selection module. In addition, this approach presents a rapid computational method for developing the DCNN classifier with the incorporation of a data compression module. The empirical assessments carried out on a dataset consisting of 12 varied activities demonstrate that the FR-DCNN model developed in this study exhibits superior performance in terms of both computational efficiency and accuracy for recognition tasks.

Cheng Xu et al [10] introduced a deep learning framework named InnoHAR, developed through the fusion of an inception neural network and a recurrent neural network. This model seamlessly processes end-to-end waveform data from multi-channel sensors. Inception-like modules are employed to extract multi-dimensional features utilizing convolution layers based on diverse kernels. By integrating the Gated Recurrent Unit (GRU), the model effectively models time series features, leveraging the intrinsic characteristics of the data to accomplish classification tasks. Extensive experimental validation conducted across three prevalent public datasets for Human Activity Recognition (HAR) consistently demonstrates the superior performance of our proposed method. Notably, the model exhibits robust generalization capabilities, surpassing existing state-of-the-art techniques.

Andrey Ignatov et al [11] developed a framework for the real-time categorization of human activities using a deep learning-based approach that is independent of user input. The methodology entails the use of Convolutional Neural Networks (CNNs) with the purpose of extracting localized characteristics, in conjunction with fundamental statistical features that preserve global time series information. Furthermore, a study is undertaken to investigate the impact of the length of time series on the accuracy of recognition. The time series duration is limited to a maximum of 1 second, which enables the categorization of continuous real-time activities. The effectiveness of the suggested approach is evaluated by using two commonly used datasets, namely WISDM and UCI. These datasets consist of labeled accelerometer data from 36 and 30 users, respectively. The assessment also includes a cross-dataset analysis. The results reveal that the proposed model delivers state-of-the-art performance measures while using minimum computer resources and eliminating the need for human feature engineering.

Abdulmajid Murad et al [12] demonstrated the use of deep recurrent neural networks (DRNNs) for the purpose of developing recognition models that are capable of effectively capturing complex dependencies present in input sequences of different lengths. The study examines several topologies, namely unidirectional, bidirectional, and cascaded, which are based on long short-term memory (LSTM) deep recurrent neural networks (DRNNs). The effectiveness of these structures is evaluated using a range of benchmark datasets. The empirical data presented in this study demonstrate that the models described herein exhibit superior performance compared to standard machine learning methods, such as support vector machines (SVM) and k-nearest neighbors (KNN). Furthermore, the proposed models exhibit enhanced performance in comparison to other deep learning approaches such as deep belief networks (DBNs) and convolutional neural networks (CNNs).

Song-Mi Lee et al [13] proposed a novel approach using a one-dimensional Convolutional Neural Network (CNN) to accurately detect and classify human activities. This technique involves the analysis of triaxial accelerometer data collected from volunteers' cellphones. The collection of data pertaining to three unique human actions, namely walking, running, and staying motionless, is carried out via the use of the accelerometer sensor that is included inside cellphones. In order to enhance the learning process of the one-dimensional convolutional neural network (1D CNN), the acceleration data along the x, y, and z axes are transformed into vector magnitude data. This transformed data is then used as the input during the network's training phase.

Charissa Ann Ronao et al [14] presented a novel approach utilizing a deep convolutional neural network (convnet) for performing efficient and effective human activity recognition (HAR) using smartphone sensors. The proposed method leverages the inherent characteristics of activities and 1D time-series signals, while also offering an automated and data-adaptive feature extraction process from raw data. Experimental evidence demonstrates that convolutional neural networks (ConvNets) do, in fact, extract pertinent and more intricate features when additional layers are added. However, the disparity in the complexity level of these features diminishes with the inclusion of each subsequent layer. The use of a broader temporal range for local correlation and the adoption of a smaller pooling size have been shown to provide advantageous outcomes. Convolutional neural networks (CNNs) have shown exceptional performance in classifying moving activities, even those that are very similar and traditionally considered difficult to identify.

Pichao Wang et al [15] introduced an easy and effective methodology aimed at converting the spatio-temporal data included in 3D skeletal sequences into multiple 2D pictures referred to as Joint Trajectory Maps (JTM). Convolutional Neural Networks (ConvNets) are then used to extract distinctive characteristics with the aim of achieving real-time recognition of human actions. The efficacy of this approach has been evaluated using three well-established public benchmarks, specifically the MSRC-12 Kinect gesture dataset (MSRC-12), G3D dataset, and UTD multimodal human action dataset (UTD-MHAD). The results have demonstrated exceptional performance, thereby establishing a novel state-of-the-art benchmark.

III. PROPOSED MODEL

Human Activity Recognition (HAR) using neural networks encompasses many essential stages. To start, it is essential to gather labeled data that encompasses time-series sensor values with their accompanying activity labels. The data should be preprocessed via the use of cleaning, normalization, and segmentation techniques to ensure its relevance and accuracy within appropriate time periods. The features are extracted from the preprocessed data in order to give relevant input to the neural network. Propose an appropriate neural network design, often a sequential model including of input layers, dense layers for feature extraction, and maybe recurrent layers to account for temporal relationships.

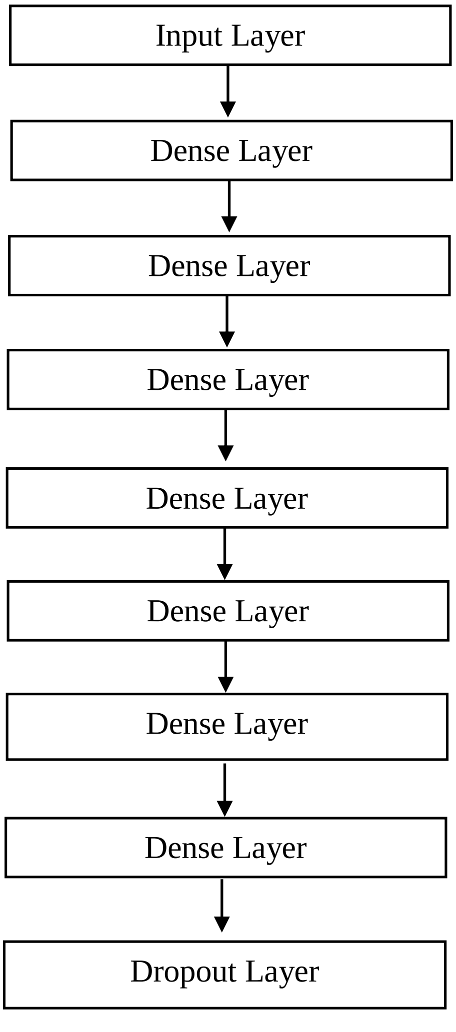

Figure 1: Proposed model Architecture

The identification of human activity, often referred to as Human Activity identification (HAR), is often accomplished via the use of sequential models. This approach entails the analysis of time-series data in order to accurately identify various activities, including but not limited to walking, running, and sitting. In this discussion, we will analyze the fundamental components often used in a sequential model for Human Activity Recognition (HAR).

- **Input Layer:**

The input layer serves as the primary layer of the neural network, responsible for receiving and representing the unprocessed input data. In the context of Human Activity Recognition (HAR), the data often comprises time-series sensor measurements, such as data from accelerometers or gyroscopes. Each individual feature or sensor channel has the potential to correlate to a distinct dimension inside the input layer. For example, if the model utilizes accelerometer data in three dimensions (x, y, z), each dimension may be considered as a distinct feature inside the input layer. The primary function of the input layer is to receive and transmit the input data to the succeeding levels in order to facilitate further processing.

- **Dense (Fully Connected) Layer:**

The dense (fully connected) layer is a kind of layer in a neural network that connects every neuron from the previous layer to every neuron in the current layer. Dense layers are very prevalent inside neural networks, being the most often used layer type. In a dense layer, every neuron is linked to each neuron in the preceding layer, resulting in a completely connected configuration.

The user's text does not provide any information to rewrite. In the domain of Human Activity Recognition (HAR), it is common practice to use one or more densely connected layers in order to effectively capture intricate connections and patterns within the input data. The neurons within these layers undergo a learning process in order to extract pertinent properties that are essential for the identification of activities. The quantity of neurons inside the thick layers and the selection of activation functions are pivotal factors that significantly impact the model's capacity to acquire knowledge and extrapolate information from the input data.

- **Dropout Layer:**

The dropout layer is a regularization method used in neural networks to mitigate the issue of overfitting. During the training process, a proportion of the neurons is stochastically set to zero, resulting in a phenomenon known as "dropout". This technique is used to enhance the network's ability to acquire more resilient and robust characteristics.

The user's text does not provide any information to rewrite in an academic manner. In the context of hierarchical attention networks (HAR), the inclusion of a dropout layer subsequent to one or more dense layers might enhance the model's ability to generalize well on input that has not been previously seen. It is particularly crucial to consider this aspect when working with limited datasets or intricate models that may be susceptible to overfitting. The inclusion of a dropout layer in the model facilitates a decrease in the model's dependence on individual neurons, hence fostering a more generalized comprehension of the input data.

In brief, a common approach for human activity detection involves using a sequential model. This model generally commences with an input layer that receives time-series sensor data. Subsequently, one or more dense layers are used for the purpose of extracting and representing features. Additionally, dropout layers may be included into the model to facilitate regularization. The design and layout of these layers are contingent upon the specific attributes of the input data as well as the intricacy of the recognition job at hand.

Hyperparameter tuning

Hyperparameter tuning, sometimes referred to as hyperparameter optimization, is a methodical exploration of various hyperparameter values in order to identify the optimal combination that yields the maximum performance for a given model. The objective of this procedure is to achieve an equilibrium between underfitting and overfitting by the identification of hyperparameter values that exhibit strong generalization capabilities towards data that has not been previously seen.

The techniques used for the optimization of hyperparameters:

Grid search: Grid search is a method that entails defining a grid of hyperparameter values and then training the model using every conceivable combination. The method is characterized by its systematic and comprehensive nature.

Random Search: Random search is a hyperparameter optimization technique that differs from grid search in that it randomly selects hyperparameter values from predetermined ranges. The aforementioned methodology has the potential to exhibit more efficiency in comparison to grid search, particularly in scenarios when the search space is extensive.

Bayesian Optimization: Bayesian optimization use probabilistic models to provide predictions about the performance of various hyperparameter setups. The process involves the selection of novel hyperparameter values by leveraging the model's predictions in conjunction with an acquisition function.

Gradient-Based Optimization: Gradient-based optimization approaches may be used in certain scenarios to identify optimum hyperparameters. Nevertheless, this phenomenon is often seen in the context of deep learning and neural network structures.

IV. EXPERIMENTAL RESULTS

This section provides a comprehensive analysis of the results obtained from the simulations conducted using the proposed methodology. The dataset used in this study was acquired via the online platform Kaggle. The dataset underwent processing according to the recommended technique. The Human Activity Recognition database was constructed using data collected from 30 research participants who engaged in various activities of daily living (ADL). During the study, individuals wore a waist-mounted smartphone equipped with inertial sensors. The aim of this task is to categorize various actions into one of the six predefined activity types.

The dataset includes information for each entry.

- The triaxial acceleration is derived from the accelerometer, which measures the overall acceleration, as well as the predicted body acceleration.
- The triaxial angular velocity may be obtained using the gyroscope.
- The dataset consists of a 561-feature vector including variables from both the time-based and frequency domains.
- The label assigned to its action.
- The name and identification of the experimenter subject.

The sample data extracted from the dataset is shown in Figure 2.

	tBodyAcc-mean()-X	tBodyAcc-mean()-Y	tBodyAcc-mean()-Z	tBodyAcc-std()-X	tBodyAcc-std()-Y	tBodyAcc-std()-Z	tBodyAcc-mad()-X	tBodyAcc-mad()-Y	tBodyAcc-mad()-Z	tBodyAcc-max()-X	tBodyAcc-max()-Y	tBodyAcc-max()-Z
0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Figure 2: Sample data from Dataset

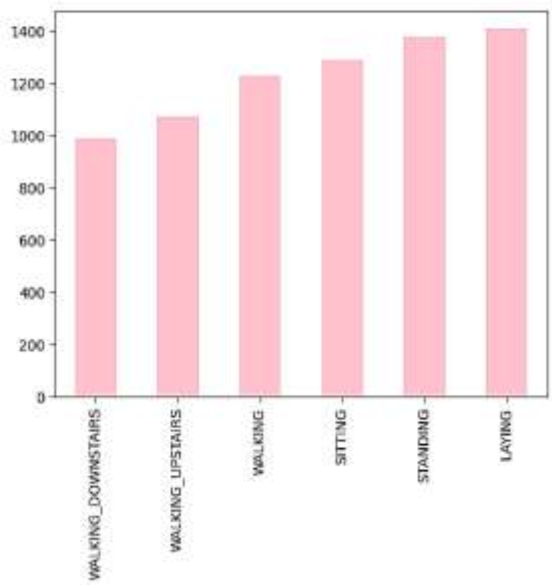

Figure 3: Activity count in Dataset

In Figure 3, a visual representation illustrates the distribution of various activities performed by individuals in the dataset. The depicted activities encompass a range of daily movements and postures, including walking downstairs, walking upstairs, walking, sitting, standing, and laying.

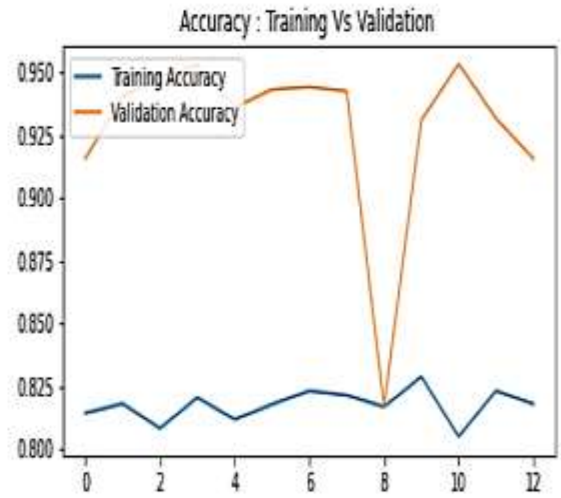

Figure 4: Training and validation Accuracy

Figure 4 shows the Training and validation Accuracy. Training accuracy and validation accuracy are often used performance measures in the field of machine learning for evaluating the efficacy of a model during the training phase. The metric of training accuracy evaluates the performance of a model on the training dataset, including instances that the model has been exposed to and acquired knowledge from. The metric denotes the ratio of accurately categorized cases in the training phase. In contrast, the assessment of validation accuracy involves assessing the model's ability to generalize to unfamiliar data by measuring its performance on a distinct validation dataset.

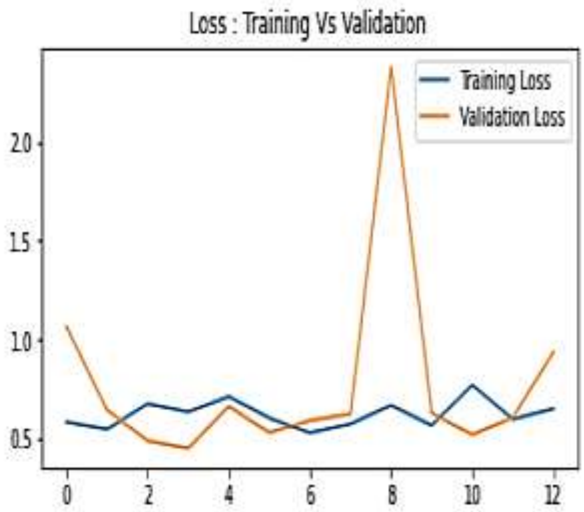

Figure 5: Training and validation Loss

Figure 5 shows the Training and validation Loss. The measurements of training and validation loss play a crucial role in the training and assessment processes of machine learning models. In the training phase, a model is presented with a dataset that has been labeled, and its parameters are repeatedly modified in order to minimize the training loss. The training loss measures the disparity between the model's anticipated outputs and the actual labels associated with the training data. The objective of this procedure is to enhance the model's ability to effectively apply learned knowledge to data that it has not been previously exposed to. Simultaneously, the evaluation loss is computed using a distinct dataset that was not used throughout the training process, so offering an autonomous assessment of the model's efficacy when presented with new, unfamiliar samples. It is important to monitor both the training and validation loss in order to evaluate the model's capacity to learn from the training data while avoiding overfitting, which refers to excessively fitting the training data, and to measure its effectiveness on unseen data.

Accuracy:

At the 11th epoch, the suggested model reached its highest level of accuracy, which was “0.9532”.

V. CONCLUSION

In conclusion, this paper has presented a compelling exploration into the realm of human activity recognition (HAR) using Convolutional Neural Network (CNN) deep learning models. With HAR being integral to applications spanning healthcare monitoring to smart environments, the study has demonstrated the efficacy of CNNs in capturing and interpreting complex patterns within temporal sequences of sensor data associated with human activities. The research commenced with a meticulous preprocessing of sensor data, transforming raw input into a format conducive to training a CNN model. The preservation of temporal dependencies within the data enabled the model to discern nuanced patterns indicative of various human activities. Training the CNN on labeled datasets encompassing diverse activities facilitated the learning and generalization of patterns inherent in the sensor data. Optimizing the model's architecture struck a balance between complexity and efficiency, ensuring its effectiveness in real-time applications. The proposed CNN model achieved a commendable accuracy of 95.32%, underscoring its ability to accurately recognize and classify a wide array of human activities.

VI. REFERENCES

[1] Voicu, Robert-Andrei, Ciprian Dobre, Lidia Bajenaru, and Radu-Ioan Ciobanu. "Human physical activity recognition using smartphone sensors." *Sensors* 19, no. 3 (2019): 458.

[2] Wannenburg, Johan, and Reza Malekian. "Physical activity recognition from smartphone accelerometer data for user context awareness sensing." *IEEE Transactions*

on Systems, Man, and Cybernetics: Systems 47, no. 12 (2016): 3142-3149.

[3] Saha, Jayita, Chandreyee Chowdhury, Ishan Roy Chowdhury, Suparna Biswas, and Nauman Aslam. "An ensemble of condition based classifiers for device independent detailed human activity recognition using smartphones." *Information* 9, no. 4 (2018): 94.

[4] Gumaei, Abdu, Mohammad Mehedi Hassan, Abdulhameed Alelaiwi, and Hussain Alsalman. "A hybrid deep learning model for human activity recognition using multimodal body sensing data." *IEEE Access* 7 (2019): 99152-99160.

[5] Chen, Yufei, and Chao Shen. "Performance analysis of smartphone-sensor behavior for human activity recognition." *Ieee Access* 5 (2017): 3095-3110.

[6] Zhou, Baoding, Jun Yang, and Qingquan Li. "Smartphone-based activity recognition for indoor localization using a convolutional neural network." *Sensors* 19, no. 3 (2019): 621.

[7] Nweke, Henry Friday, Ying Wah Teh, Mohammed Ali Al-Garadi, and Uzoma Rita Alo. "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges." *Expert Systems with Applications* 105 (2018): 233-261.

[8] Almaslukh, Bandar, Abdel Monim Artoli, and Jalal Al-Muhtadi. "A robust deep learning approach for position-independent smartphone-based human activity recognition." *Sensors* 18, no. 11 (2018): 3726.

[9] Qi, Wen, Hang Su, Chenguang Yang, Giancarlo Ferrigno, Elena De Momi, and Andrea Aliverti. "A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone." *Sensors* 19, no. 17 (2019): 3731.

[10] Xu, Cheng, Duo Chai, Jie He, Xiaotong Zhang, and Shihong Duan. "InnoHAR: A deep neural network for complex human activity recognition." *Ieee Access* 7 (2019): 9893-9902.

[11] Ignatov, Andrey. "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks." *Applied Soft Computing* 62 (2018): 915-922.

[12] Murad, Abdulmajid, and Jae-Young Pyun. "Deep recurrent neural networks for human activity recognition." *Sensors* 17, no. 11 (2017): 2556.

[13] Lee, Song-Mi, Sang Min Yoon, and Heeryon Cho. "Human activity recognition from accelerometer data using Convolutional Neural Network." In *2017 ieee international conference on big data and smart computing (bigcomp)*, pp. 131-134. IEEE, 2017.

[14] Ronao, Charissa Ann, and Sung-Bae Cho. "Human activity recognition with smartphone sensors using deep learning neural networks." *Expert systems with applications* 59 (2016): 235-244.

[15] Wang, Pichao, Zhaoyang Li, Yonghong Hou, and Wanqing Li. "Action recognition based on joint trajectory maps using convolutional neural networks." In *Proceedings of the 24th ACM international conference on Multimedia*, pp. 102-106. 2016.