# A novel hybrid approach of detecting human activity recognition : Using CNN + Bi-GRU

**Min-Jong Cheon[1], Han-Seon Joo[2], Hyeon-June Jeon[3]**
[1] Hanyang University, Seoul, Republic of Korea
[2] Catholic University of Korea, Bucheon, Republic of Korea
[3] Sungkyunkwan University, Suwon, Republic of Korea

## Abstract

Human Activity Recognition (HAR) technology is about collecting and interpreting information about human motion or gestures and can be applied to various fields such as RGB cameras or Depth cameras. We collected the dataset from the kaggle site for the classification. Machine learning models including Logistic Regression, KNeighborsClassifier, SGDClassifier, SVC, GradientBoostingClassifier, AdaBoostClassifier, XGBClassifier, Extra Tree Classifier, LGBM Classifier, and HistGradientBoostingClassifier. Our proposed model includes CNN + Bi-LSTM and yields 97% accuracy. Our principal finding is that as the dataset consists of both temporal and spatial features, our suggested one, particularly specialized on the CNN and Bi-LSTM outperformed various artificial intelligence models. Furthermore, the bidirectional GRU was more efficient in classification compared to the unidirectional GRU. However, our limitation also includes the format of the given dataset and the low performance from the AdaBoostClassifier.

***Keywords: Deep Learning, Machine Learning, GRU, CNN, HAR***

## 1. Introduction

### 1.1 Background

Human Activity Recognition (HAR) technology refers to a technology that uses a variety of sensors to collect and interpret information related to human motion or gestures. This technology, which compares and analyzes human behavior with AI, can be applied to various fields such as intelligent CCTV, healthcare services, and entertainment. Human Activity Recognition technology can be largely divided into image analysis-based behavior recognition technology and on-body sensor-based behavior recognition technology depending on the type of data being analyzed. Image analysis-based behavior recognition technology is a technology that recognizes human behavior by analyzing images collected from RGB cameras or Depth cameras, and has the advantage of being able to recognize human behavior under various image conditions[1]. This technology can be applied to CCTV images, three-dimensional cameras, images taken with infrared thermal imaging cameras, and virtual composite images, and there is no restriction on the behavior to be recognized. Sensor-based behavior recognition technology is a technology that analyzes various sensor information such as accelerometer, gyroscope, Bluetooth, and sound sensors installed to users and recognizes behavior[2]. Due to the recent widespread distribution of smartphones, devices such as smartphones and smartwatches have built-in sensors, including inertial measuring devices, which allow users to collect data necessary for behavior recognition without having to attach additional sensors. Behavior recognition analysis techniques include methods using AdaBoost, SVM, and random forest, and deep learning-based behavior recognition methods. Recently, with the development of artificial intelligence technology and its advantages, deep learning algorithms such as CNN and RNN have been applied to behavior recognition research[3].

### 1.2 Objective

As the previous research conducted by Ronald Mutegeki, and Dong Seong Han, their suggested model which is composed of CNN and LSTM had deep temporal and spatial features in the model. However, their model achieved 93 % accuracy of the given dataset, which was collected from the UCI dataset, which is available at https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones[4]. Therefore, our aim is to achieve higher accuracy through our suggested model, by adding CNN and Bidirectional GRU for both considering the spatial and temporal features of the given dataset. To sum up, with our suggested model, we could help people with health monitoring, and self monitoring systems.

## 2. Materials and Methods

### 2.1 Data Description

The given dataset is collected from 30 volunteers with an age range of 19 - 48 years. The volunteers performed each behavior which are walking, walking up stairs, walking down stairs, sitting, standing and laying and those behaviors were accumulated through a smartphone, Samsung S II on the waist. Through its embedded gyroscope and accelerometer, 3-axial angular velocity and linear acceleration were calculated with a constant rate of 50 Hz. The dataset consists of 10299 rows and 563 columns. For training the dataset efficiently, we divided the dataset into 70% for the train set and 30% for the test set.

### 2.2 CNN

A Convolutional Neural Network(CNN) performs superior performance with the image dataset compared to the Deep Neural Network(DNN). The CNN is composed of the convolutional layer, pooling layer, and the fully connected layer. The convolutional layer utilizes the feature, which explores the given image and then extracts the features from that. Then, the feature map is generated through the filters and then pooling layers use that feature map as the input one. There are two kinds of pooling layers, which are max pooling and average pooling per each. Those pooling layers minimize the feature map for the size reduction. The convolutional and pooling layers are repeated through those steps and then lastly, the fully connected layer is added for the classification, The fully connected layer is the same as the DNN, which uses the softmax activation function for multi label classification and the sigmoid function for the binary classification[5].

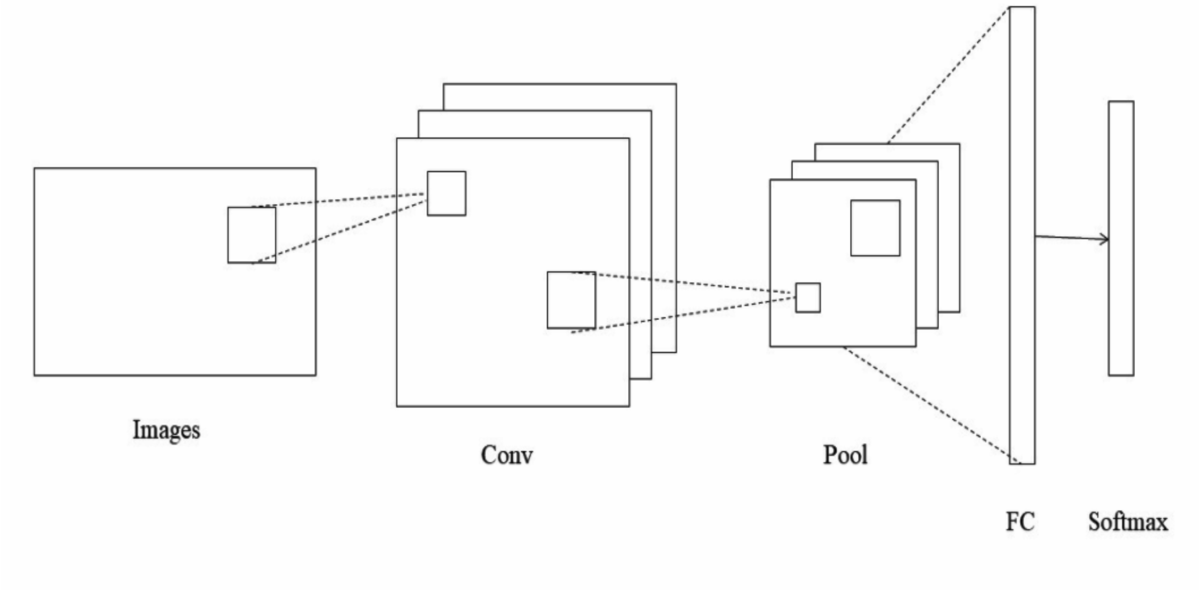

Fig.1 Overall architecture of CNN

## 2.3 GRU

While a Long Short Term Memory(LSTM) which also belongs to the RNN mainly consists of three gates, which are a forget gate, input gate and the output gate[]. However, in the GRU, only a reset gate and an update gate are used in the network. Reset gate uses sigmoid function as output to multiply the value (0,1) by the previous hidden layer for the purpose of properly resetting historical information[]. The update gate feels like a combination of LSTM's forget gate and input gate to determine the percentage of update information in the past and present. In update gate, the output to sigmoid determines the amount of information at this point, subtracts from 1 multiplied by the information of the hidden layer at the previous point, and is similar to input gate and forget gate of each LSTM. The candidate phase is the calculation of the candidates at this point. The key of this step is to multiply the reset gate results without using the information of the past hidden layers. For the hidden layer calculation, the unit calculates the hidden layer at this point by combining the update gate result with the condate result[6].

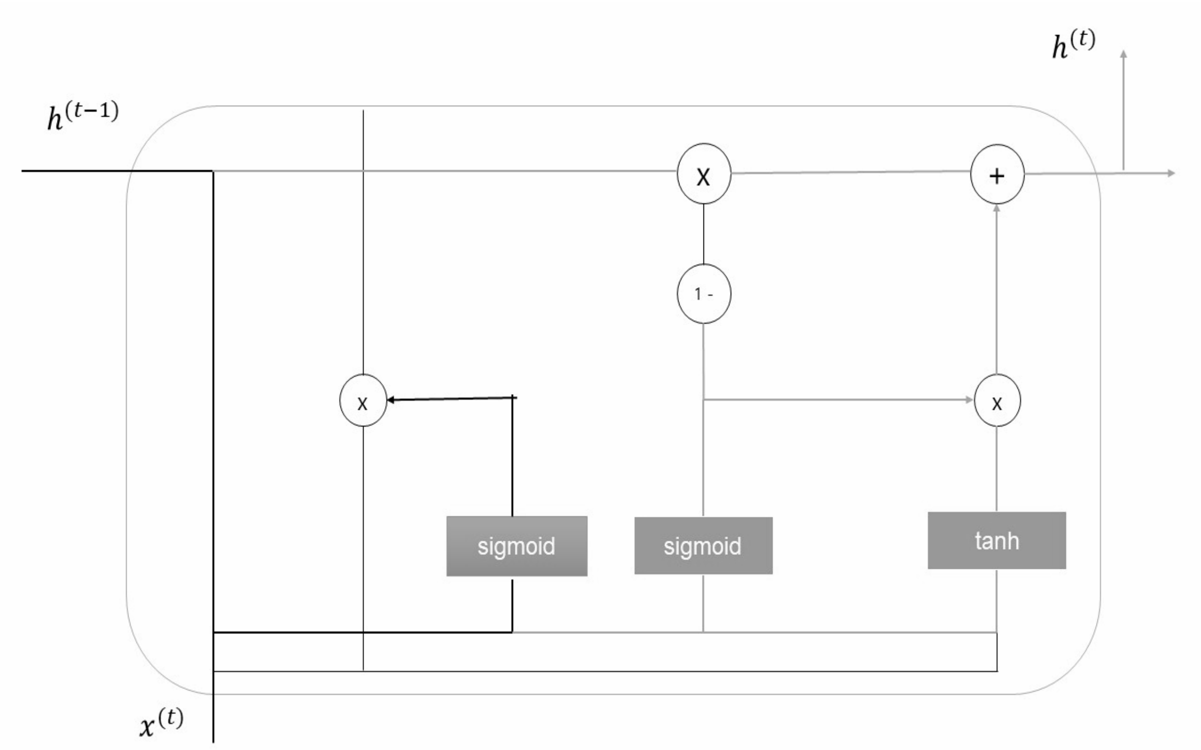

Fig.2 Overall architecture of GRU

## 2.4 Pipeline of the experiment

Firstly, as the input data has 10299 rows and 563 columns, we conducted Principal Component Analysis(PCA). Through the PCA analysis, we down sampled our given dataset into 15 columns. Then, the train test split function was utilized and 70% of the dataset was used as train dataset and the 30% of the dataset was used as test dataset. The standard scaler was also conducted for normalizing the dataset, for fitting the range of each column. Lastly, various machine learning models and deep learning models were used for calculating the accuracy score of the classification.

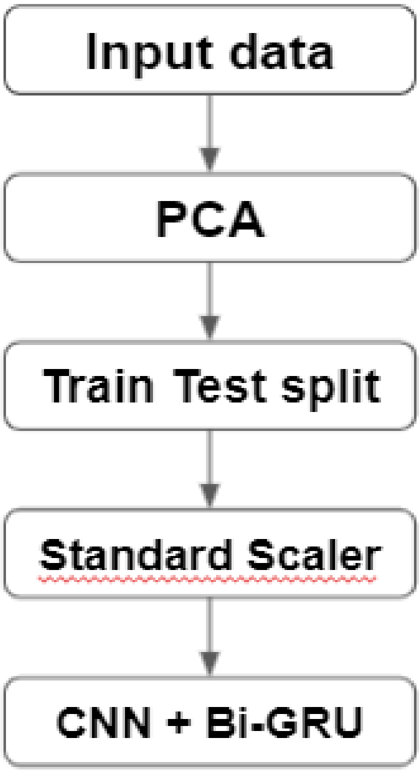

Fig.3 Overall pipeline of the experment

## 3. Result

### 3.1 Machine Learning Models

10 different machine learning models; HIst Gradient Boosting classifier, Light Gradient Boosting Machine (LGBM) classifier, ExtraTrees classifier, XGB classifier, Adaboost classifier, Gradient Boosting classifier, Support Vector Classifier (SVC), Stochastic Gradient Descent(SGD) classifier, KNeighbors classifier and Logistic Regression were utilized for the classification. As the graph below shows, the AdaBoost classifier shows the lowest accuracy which is 0.43, and lgbm yields the highest accuracy which is 0.938.

### 3.2 CNN+Bi=GRU

To improve the accuracy, we added the bidirectional GRU from CNN. Bidirectional GRU allows training the whole parameters in the network, by training the bidirectional ways[]. Therefore, as the graph below shows, our suggested model yields the highest accuracy, 97%. Adding unidirectional GRU only shows 91% accuracy and DNN yields the 90% accuracy with the optimizer adam and sigmoid activation function[7].

## 4. Discussion

### 4.1 Principal Finding

In order to classify the given dataset with substantial accuracy, we suggested the novel hybrid approach which consists of CNN and Bi-GRU. The reason why we combined them is because the dataset consists of the spatial and temporal features. As the result shows, our suggested model outperformed the other ones, especially machine learning models, as the highest accuracy among them was 93% from lgbm, and lowest with 43% from AdaBoostClassifier, while our model yielded 97%.

### 4.2 Limitation

However, there exist some limitations in our research. First of all, even though the dataset is about human activity recognition, the format of the dataset is all about the csv file, which does not have any image related ones. Furthermore, in our result, AdaboostClassifier showed the lowest performance especially compared to the other machine learning models. Therefore, further research should be conducted to solve the downside from the AdaboostClassifier.

## 5. Conclusion

In our research, we conducted the experiment to examine the classification result from the given dataset. Our models include various machine learning models including Logistic Regression, KNeighborsClassifier, SGDClassifier, SVC, GradientBoostingClassifier, AdaBoostClassifier, XGBClassifier, Extra Tree Classifier, LGBM Classifier, and HistGradientBoostingClassifier. Among them, the lowest performance was from the AdaBoostClassifier and the highest one from the LGBM Classifier. However, our suggested one excelled the other ones including both machine learning models and deep learning models with showing the 97% accuracy.

## 6. Reference

[1] Ke, S.-R., Thuc, H., Lee, Y.-J., Hwang, J.-N., Yoo, J.-H., & Choi, K.-H. (2013). A review ON Video-Based human activity recognition. *Computers*, *2*(2), 88–131. https://doi.org/10.3390/computers2020088

[2] Pandya, S. (2021, January 26). *Council post: Understanding the connection: Human activity recognition, safety and productivity*. Forbes. https://www.forbes.com/sites/forbestechcouncil/2021/01/27/understanding-the-connection-human-activity-recognition-safety-and-productivity/?sh=4fabdc1939a8.

[3] Fadelli, I. (2020, August 25). *A 26-layer convolutional neural network for human action recognition*. Tech Xplore - Technology and Engineering news. https://techxplore.com/news/2020-08-layer-convolutional-neural-network-human.html.

[4] UCI machine Learning Repository: Human activity recognition using SMARTPHONES data set. (n.d.). https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones.

[5] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*. https://doi.org/10.1109/icengtechnol.2017.8308186

[6] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

[7] Tao, Q., Liu, F., Li, Y., & Sidorov, D. (2019). Air pollution forecasting using a deep learning model based On 1D convnets and Bidirectional GRU. *IEEE Access*, *7*, 76690–76698. https://doi.org/10.1109/access.2019.2921578