An Intelligent Robust One Dimensional HAR-CNN Model for Human Activity Recognition using Wearable Sensor Data

Dr. P. Deepan, Associate Professor-AIML, St. Martins Engineering College, deepanp87@gmail.com

Dr. R. Santhosh kumar, Associate Professor - AIDS, St. Martins Engineering College, santhoshkumar.aucse@gmail.com

Dr. B. Rajalingam, Associate Professor-CSE, St. Martins Engineering College, rajalingam35@gmail.com

Dr. P. Santosh Kumar Patra, Professor - CSE, St. Martins Engineering College, principal@smec.ac.in

S. Ponnuthurai , Assistant Professor-CSE, Roever Engineering College, ponnuthuraiselvaraj@gmail.com

Abstract— **One of the biggest new trends in artificial intelligence is the ability to recognise people's movements and take their actions into account. It can be used in a variety of ways, including for surveillance, security, human-computer interaction, and content-based video retrieval. There have been a number of researchers that have presented vision-based techniques to human activity recognition. Several challenges need to be addressed in the creation of a vision-based human activity recognition system, including illumination variations in human activity recognition, interclass similarity between scenes, the environment and recording setting, and temporal variation. To overcome the above mentioned problem, by capturing or sensing human actions with help of wearable sensors, wearable devices, or IoT devices. Sensor data, particularly one-dimensional time series data, are used in the work of human activity recognition. Using 1D-Convolutional Neural Network (CNN) models, this works aims to propose a new approach for identifying human activities. The Wireless Sensor Data Mining (WISDM) dataset is utilised to train and test the 1D-CNN model in this dissertation. The proposed HAR-CNN model has a 95.2%of accuracy, which is far higher than that of conventional methods.**

Keywords— Human activity, IoT, wireless data, 1D-Convoluion, vision and sensor.

I. INTRODUCTION

In many parts of the globe, human activity recognition (HAR) is a prominent area of investigation. Industry automation, sports, medical technology, security, smart cities, and smart homes are just a few of the applications for which it may be applied. HAR is vital in human-centered applications including health detection, driving behaviour tracking, gait detection, fall detection, and other personalised services. Human activity recognition may be separated into two categories, as seen in Figure 1, namely, visual based and sensor-based human activity recognition [1]. A camera system is utilised in the vision-based HAR to track human activity as well as changes in the surrounding landscape. Techniques such as marker extraction and structural modelling, as well as motion segmentation and extraction of actions and tracking are employed in this method to achieve its results [2]. Researchers use a wide variety of cameras, ranging from simple RGB cameras to more complex systems, such as the fusion of many cameras to provide stereo vision or depth cameras, which make use of infrared LEDs to determine the depth of an image [3].

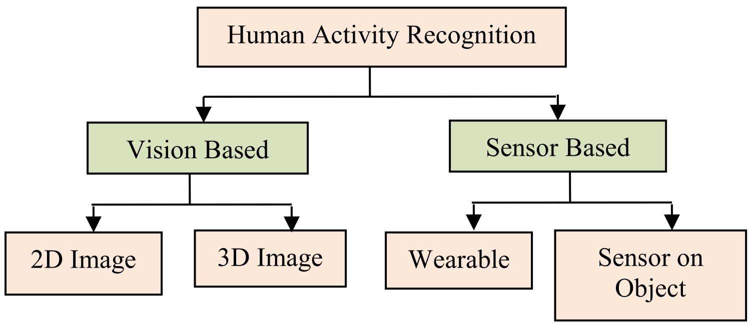

Fig. 1. Types of human activity recognition

The vision based HAR can be classified as 2Dimages and 3D images. Several researchers [4,5] have been published on vision-based activity recognition system. However, the vision based HAR system met the following issues while attempting to classify human action recognition.

- Recognition of variations in human behaviour due to lighting conditions
- There is a lot of inter class similarity between scenes.
- The setting of Environment and record
- Temporal variation
- Obtaining and labelling training data

The tracking of a person's activities using a network of sensors and linked devices is similar to sensor-based human activity recognition. They create data in the form of a succession of state changes or parameter values that occur over a time period of several seconds to minutes. The use of contact detectors, radio frequency identification (RFID), accelerometers, motion detectors, noise detectors, radar, and

ISBN: 978-1-6654-7436-8/22/$31.00 ©2022 IEEE

other sensors may all be used to identify people, objects, and the surrounding environment. Among the sensor-based technologies available, there are two categories to consider: wearables and sensors on objects [6].

1.1 Organization the paper

The following is an outline of the research work: The findings of past relevant study are discussed in Section 2. Section 3 explores into the construction of a human activity detection system, with a particular emphasis on the convolutional neural network approach of recognising activities. Section 4 describes the experimental setup used to evaluate the proposed approach, as well as the findings obtained from it. Section 5 presents a summary of the findings and discusses possible avenues for further research.

II. RELATED WORK

The HAR field has been doing a substantial amount of research over the course of the last decade to better understand and recognise human behaviour for a wide range of applications. To begin, the researchers used manual methods to discern between the activities of daily life in smart homes. These handmade characteristics are generated once the dataset has been segmented into specified activity sequences or windows [7]. Researchers have looked into a range of aspects in order to design effective activity recognition systems. Yala et al. [8] first presented the standard, time-dependent, sensor-dependent, and sensor-dependent extension feature vector extraction methods that are elaborated upon as well. Subsequently, a final classification is made using a classification technique, such as SVM or Random Forest. Aminikhanghahi et al. [9] recently evaluated multiple forms of sensor flow segmentations, inspired by prior work. They did, however, include a list of handcrafted characteristics. The number of seconds since midnight, the day of the week, and the duration between sensor changes have all been investigated as temporal aspects. Location, among other spatial factors, was assessed. However, metrics like the number of events that occurred in the window and the identification of the sensor are often included in the segments that came before it. In their study, He and colleagues [10] suggested a high-precision HAR system that makes use of discrete cosine transform, principal component analysis, and support vector machine analysis.

In the beginning of this research study, the researchers classified the sensor data using traditional machine learning methods such as decision trees, support vector machines (SVMs), and naive bayes. The features of the acceleration and angular velocity data were recovered with the use of a gradient histogram and a Fourier descriptor that were based on the centroid feature described in [11]. This feature was used to determine the properties of the data. The author [12] used two different classifiers: SVM and k-NN, which are both neural networks. Ronao and colleagues [13] developed a monitoring system that made use of six inertial measurement units. In order to categorise the activities based on a number of network features that passed the statistical test, the random forest (RF) classifier was employed. At the conclusion of the round, an overall accuracy of 84.6 percent was obtained. The authors of study [14] discuss a wearable wireless accelerometer-based activity identification system and its use in medical detection. The system was designed to monitor a patient's activity levels. In order to accomplish the task of feature selection, a combination of Relief-F and sequential forward floating search (SFFS) was utilized. Finally, by utilising naive bayesian and k-NN models, we were able to classify the various activities that were carried out and make comparisons between them.

When it comes to tasks involving human activity identification, machine learning algorithms may depend substantially on the manual extraction of features. The most prevalent kind of constraint is one that is imposed by human domain knowledge. Deep learning algorithms have been used by researchers in order to address this issue. These algorithms are able to automatically extract and present low-level original temporal properties with abstract sequences from raw sensor data during the training phase of the deep learning process. In recent developments in the area of pattern recognition, time series based 1D convolutional neural network deep learning models (also known as HAR-CNN) are being applied to the problem of identifying human activities.

III. PROPOSED WORK

In this section, we present ways for improving the classification accuracy of the HAR by using the 1D-convolutional neural network deep learning model. The processes of proposed HAR-CNN model consist of dataset collection, pre-processing of data, extracting the features from the data and finally classify and recognize the human activity. The proposed HAR-CNN architecture is depicted in Figure 3.

3.1 *Contribution of this research work*

The Summary of the main contribution is as follows:

- Dataset Collection
- Pre-processing
- Training the data using proposed HAR-CNN
- Evaluate metrics
- Recognizing the human activity

3.2 *Data Collection*

We utilised data from our Wireless Sensor Data Mining Lab to create the Human Activity Recognition (HAR) dataset [15].

Fig. 2. Sample Image for Wearable Sensor

ISBN: 978-1-6654-7436-8/22/$31.00 ©2022 IEEE

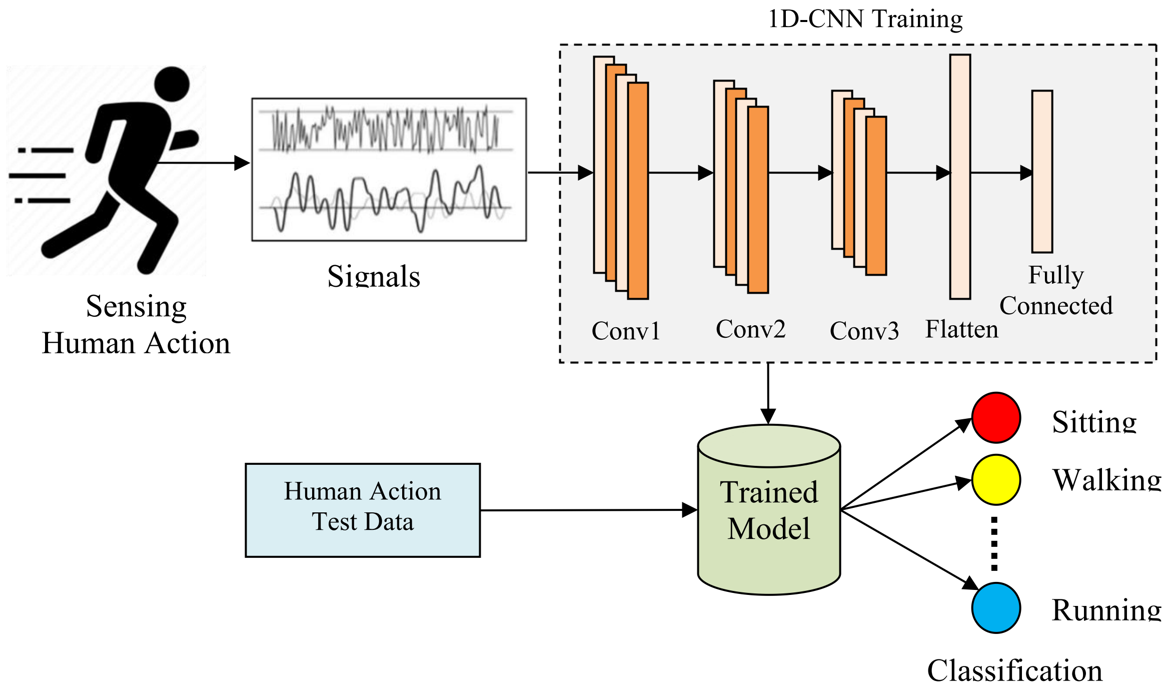

Fig. 3. Architecture of 1D HAR-CNN Model

A total of 36 people provided their contact information through smartphone to compile this research. To gather data, a tri-axial accelerometer is incorporated in the device (x axis-y axis-z axis). It takes 20 seconds to capture a sample, and the sampling frequency is set at 20 Hz (20 samples per second). Figure 2 shows sample image for wireless sensor. There are 1098203 records (rows) of data present in the six human activities in the Wireless Sensor Data Mining (WISDM) dataset.

3.3 Pre-Processing

The raw sensor data must be transformed first before the classifier method can be applied. The raw data from the accelerometer is shown in Figure 4, and it reveals that each of the three axes is represented by a value. Segmenting the data into 10-second intervals without overlap has been done. This is because he thought that the 200 readings in a 10-second interval were sufficient.

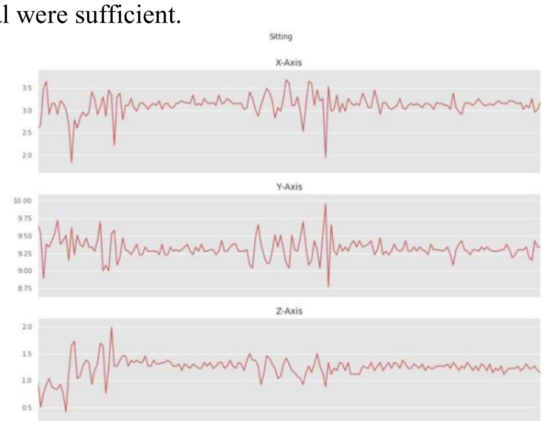

Fig. 4. Sample signal information for sitting activity

Then they created features based on 200 raw accelerometer readings for each segment. It is possible to generate a total of 43 features. Six different extraction processes constitute the basis for all of these variations. Analyses are performed on each axis to determine the average, standard deviation, average absolute difference, and duration between peaks. In addition to these results, it is feasible to derive the Average Resultant Acceleration and the Binned Distribution.

	user-id	activity	timestamp	x-axis	y-axis	z-axis
1098194	19	Sitting	131623091524000	8.5	-1.3	2.5
1098195	19	Sitting	131623131471000	8.7	-1.3	2.4
1098196	19	Sitting	131623172578000	8.8	-1.3	2.2
1098197	19	Sitting	131623251466000	9.1	-1.4	1.9
1098198	19	Sitting	131623291475000	9.0	-1.5	1.8
1098199	19	Sitting	131623331483000	9.0	-1.6	1.7
1098200	19	Sitting	131623371431000	9.0	-1.5	1.7
1098201	19	Sitting	131623411592000	9.1	-1.4	1.7
1098202	19	Sitting	131623491487000	9.0	-1.5	1.7
1098203	19	Sitting	131623531465000	8.9	-1.3	1.6

Fig. 5. Attributes for Sitting Activity

ISBN: 978-1-6654-7436-8/22/$31.00 ©2022 IEEE

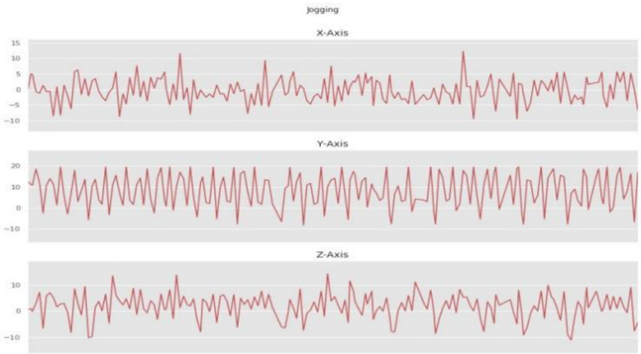

Fig. 6. Sample signal information for Jogging activity

	user-id	activity	timestamp	x-axis	y-axis	z-axis
0	33	Jogging	49105962326000	-0.7	12.7	0.5
1	33	Jogging	49106062271000	5.0	11.3	1.0
2	33	Jogging	49106112167000	4.9	10.9	-0.1
3	33	Jogging	49106222305000	-0.6	18.5	3.0
4	33	Jogging	49106332290000	-1.2	12.1	7.2
5	33	Jogging	49106442306000	1.4	-2.5	-6.5
6	33	Jogging	49106542312000	-0.6	10.6	5.7
7	33	Jogging	49106652389000	-0.5	13.9	7.1
8	33	Jogging	49106762313000	-8.4	11.4	5.1
9	33	Jogging	49106872299000	1.0	1.4	1.6

Fig. 7. Attributes for Sitting Activity

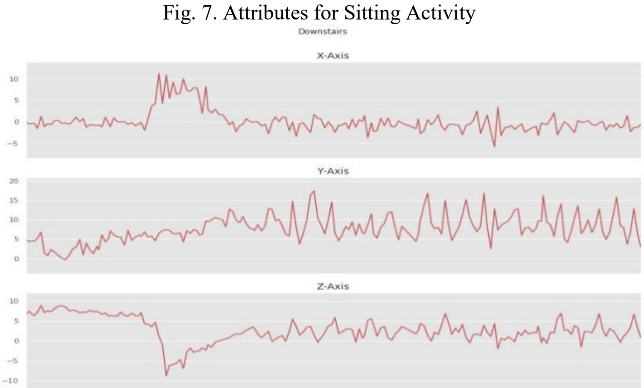

Fig. 8 Sample signal information of Downstairs activity

	user-id	activity	timestamp	x-axis	y-axis	z-axis
49000	20	Downstairs	2854792308000	1.7	9.4	4.1
49001	20	Downstairs	2854842204000	-1.9	8.1	0.5
49002	20	Downstairs	2854892284000	-0.7	7.9	3.4
49003	20	Downstairs	2854942210000	1.6	9.0	4.4
49004	20	Downstairs	2854992229000	2.5	5.9	4.9
49005	20	Downstairs	2855042247000	1.4	6.5	-2.3
49006	20	Downstairs	2855092296000	1.9	8.9	-1.6
49007	20	Downstairs	2855142253000	-0.9	8.6	5.2
49008	20	Downstairs	2855192302000	0.8	10.5	5.6
49009	20	Downstairs	2855242229000	0.6	16.8	8.2

Fig. 9 Attributes for Downstairs Activity

3.4 Proposed HAR-CNN Model

The CNN is a type of neural network whose design is based on the notion of a biological neuron defined as the receptive field, which is used to mimic the connectivity pattern of neurons in the human brain. The CNN model is a feed forward neural network that consists of a stack of filters (convolutional layer) and sub-sampling layers (pooling layer) that alternately repeat themselves, with one or more fully connected neurons (fully connected layer/dense) at the end. Despite the fact that this model can be used in a variety of domains, it works perfectly in image processing. Individual blocks or layers are concatenated together to form the CNN. These layers are being assembled in order to complete a set of tasks. The classic convolutional neural network overall architecture, which includes the following layers, is depicted in Figure 4.In order to build a CNN, the following layers are needed:

- Convolutional Layer
- Pooling Layer
- Output Layer

3.4.1 Convolution layer (Feature Extraction)

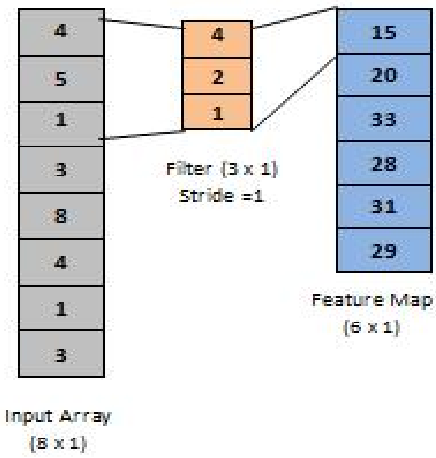

Fig. 10 calculation of 1D- convolution process

ISBN: 978-1-6654-7436-8/22/$31.00 ©2022 IEEE

The convolution layer is one of the fundamental building blocks of a CNN, and it is from this layer that the term "Convolutional Neural Network" is derived. The goal of the convolution layer is to learn feature representations from the inputs by using a sequence of filters or kernels whose parameters must be learned. It uses linear multiplication as a convolution operational activity to extract high-level properties from the sensor signals. The 1D convolutional process of HAR-CNN model is depicted in Figure 10.

3.4.2 *Pooling Layer*

When layers are pooled, the height and breadth of the feature map are decreased but the depth is maintained. Reducing the amount of work necessary to process the data while still extracting the most important characteristics in feature maps is an advantage. Max, Min and average pooling layers are the three types of pooling layers.

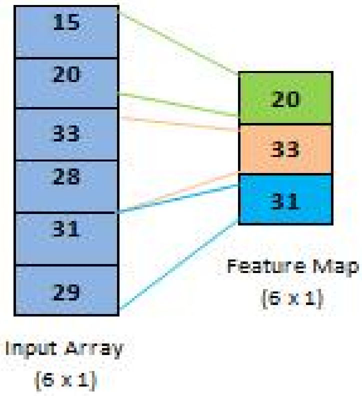

Fig. 11. Pooling Process

Max pooling, as seen in Figure 11, returns the maximum value of the elements in the region of the image that the filter affects, whereas average pooling gives the average value of the elements in that region. It is generally agreed that max pooling improves efficiency since it is better at identifying and isolating dominant features. It helps prevent the issue of overfitting as well.

- The first convolutional layer uses 192 convolutional filters and a kernel size of 64 to process accelerometer data with a stride of 1. As a result of this, its output is sent into the ReLU function.
- A max-pooling layer is used to decrease the feature representation with a stride of 3 and a kernel size of 3x1.
- Convolutional filters with 96 convolutions and a kernel size of 12 are added to the convolutional layer, and the step size is 1.More abstract and hierarchical properties can be learned this way.Its output is confined to the ReLU algorithm.
- Finally, a max-pooling layer with a stride of 3 and a kernel size of 1 reduces the feature representation by 3.
- Finally, the statistical characteristics are flattened and concatenated into the max-pooling layer's output.
- A dropout layer with a dropout rate of 0.5 is implemented in order to prevent overfitting.
- When a softmax layer receives the output from the fully-connected layer, it performs a probability distribution over six different activity classes.

The trainable parameters and non- trainable parameters of proposed HAR 1D-CNN model is shown in Figure 12.

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 69, 64)	2368
conv1d_2 (Conv1D)	(None, 60, 64)	41024
conv1d_3 (Conv1D)	(None, 53, 64)	32832
conv1d_4 (Conv1D)	(None, 48, 64)	24640
global_max_pooling1d_1 (Glob	(None, 64)	0
dropout_1 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 6)	390

Total params: 101,254
Trainable params: 101,254
Non-trainable params: 0

None

Fig. 12. Summary of Trainable Parameter for CNN

ISBN: 978-1-6654-7436-8/22/$31.00 ©2022 IEEE

IV. Experimental Results and Analysis

Human activity recognition using 1D-CNN model is the subject of this section, which examines its performance and effectiveness. The recommended HAR 1D-CNN model is constructed with the help of the integrated development environments Python and Anaconda (IDEs).

4.1 *Evaluation Metrics*

Evaluation measures for the HAR algorithms under consideration include a wide range from which to choose. The accuracy, precision, recall, and F-measure metrics were examined in this paper, and the findings were compared to one another. T_P, T_N, F_P and F_N are the four variables required by the assessment of methods. T_P and T_N are the only two classes that exist when an activity has been successfully detected (TP and TN). When an action is misclassified, it can fall into either the FP or FN categories.

Precision: Precision is one of the most common ways to measure how far a metrics will work. It is used to find out how many correctly predicted events there were out of all predictions. In this way, we can measure the precision.

$$\text{Precision} = \frac{t\_p}{t\_n+t\_n} \quad (1)$$

Recall: Recall is the percentage of occurrences that were successfully predicted over the total number of occurrences.

$$\text{Recall} = \frac{t\_p}{f\_n} \quad (2)$$

Accuracy: The number of occurrences that were given the proper classification is used as a measure of accuracy.

The ratio of the correct classification to the total number of classifications is one way to measure the accuracy of a classification system.

$$\text{Accuracy} = \frac{t\_p+t\_n}{t\_p+t\_n+f\_n+f\_p} \quad (3)$$

F1-Score: The F1 is an attempt to represent accuracy by taking the average of Recall and Precision. The harmonic mean is represented by F1 among the various F scores.

$$\text{F1} = \frac{t\_p+t\_n}{t\_p+t\_n+f\_n+f\_p} \quad (4)$$

4.2 *Result Discussions*

Deep learning models are shown and discussed in this part based on their performance on WISDM datasets. As shown in Table 1, we have increase and vary the training epochs from 25 to 150 for improve the classification accuracy.

TABLE I. Result Comparison

S. No.	Epochs	Accuracy	Precision	Recall	F1-Score
1.	25	77.9	71	78	71
2.	50	83.2	83	82	83
3.	100	89.4	89.1	89	89
4.	150	95.2	95	95	95

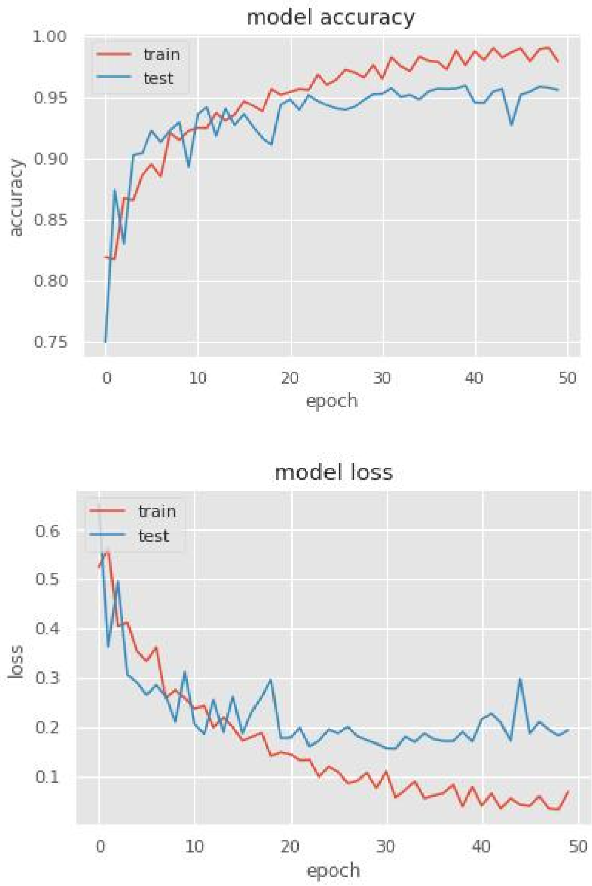

Fig. 13. Performance accuracy and loss of Human Activity Recognition Model

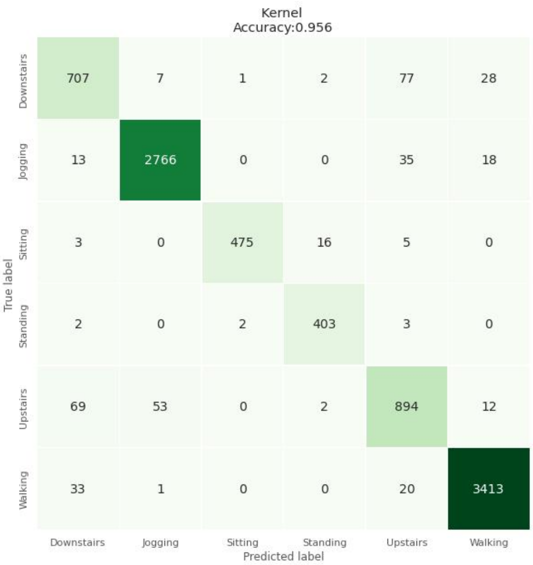

Fig. 14. Confusion matrix of Human Activity Recognition Model

ISBN: 978-1-6654-7436-8/22/$31.00 ©2022 IEEE

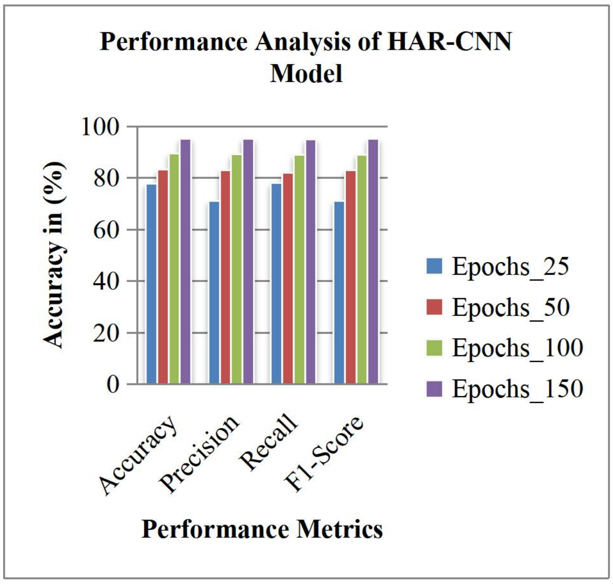

Fig. 15. Performance Analysis of HAR-CNN Model

V. CONCLUSION

With the use of a wearable device or wearable sensor, we proposed a model for human activity identification that may be used to recognise human activities. On the basis of accelerometer, gyroscope, and magnetometer smartphone sensor measurements, a deep-learning model with Convolutional layers was used to classify six distinct human activities (Standing, Sitting, Walking, Jogging, Upstairs and Downstairs). After being examined in a complicated and independent test dataset with accurate findings, the suggested HAR model was tested in a real-world environment to demonstrate that it accurately captured the fundamental features and dynamics of each underlying activity in a variety of circumstances.

Despite the fact that our proposed work and conclusions have yielded good results, there are still areas where improvements may be made in the themes that have been offered. As a result, we must make a contribution to the research in the following ways:

- To capture a real time sensor data to identify the human activity recognition.
- To recognize the human activity for more actions rather than six activity.
- To combine or hybridise the CNN and RNN with the LSTM model in order to increase the recognition accuracy of human activities.

REFERENCES

[1] Cook, D.J.; Krishnan, N.C.; Rashidi, P. Activity discovery and activity recognition: A new partnership. IEEE Trans. Cybern., Vol. 4(3), pp. 820–828, 2013.

[2] C.A. Ronao,and S.B. Cho, "Recognizing human activities from smartphone sensors using hierarchical continuous hidden Markov models,"Int. J. Distrib. Sens. Netw., vol. 13, pp. 1–16, 2017.

[3] K. Davis, E. Owusu, V. Bastani, L. Marcenaro, J. Hu, C. Regazzoni, and L. Feijs, "Activity recognition based on inertial sensors for ambient assisted living," International conference on information fusion (fusion),2016, pp. 371–378.

[4] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, ``A survey on activity detection and classi_cation using wearable sensors," IEEE Sensors J., vol. 17, no. 2, pp. 386_403, Jan. 2017.

[5] Y.S. Lee,and S.B Cho, "Activity Recognition Using Hierarchical Hidden Markov Models on a Smartphone with 3D Accelerometer," International Conference on Hybrid Artificial Intelligent Systems (HAIS'11), pp.460–467, 2011.

[6] Y. Li, D. Shi, B. Ding, and D. Liu, "Unsupervised feature learning for human activity recognition using smartphone sensors," Mining intelligence and knowledge exploration, pp. 99-107, Springer, 2014.

[7] J.Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor based activity recognition: A survey," Pattern Recognition Letters, vol.119, pp. 3–11, 2019.

[8] Yala, N.; Fergani, B.; Fleury, A. Feature extraction for human activity recognition on streaming data. In Proceedings of the 2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA), Madrid, Spain, 2–4 September, pp. 1–6, 2015.

[9] Aminikhanghahi, S.; Cook, D.J. Enhancing activity recognition using CPD-based activity segmentation. Pervasive Mob. Comput. Vol. 5(3), pp. 75–89, 2019.

[10] H. Zhang, Y. Li, Z. Lv, A. K. Sangaiah and T. Huang, "A real-time and ubiquitous network attack detection based on principle component analysis and support vector machine," in IEEE/CAA Journal of Automatica Sinica, vol. 7, no. 3, pp. 790-799, May 2020.

[11] W. Jiang, and Z. Yin, "Machine learning based Human activity recognition using wearable sensors," ACM international conference on Multimedia, pp. 1307-1310, 2015.

[12] F. Herna´ndez, L. F. Sua´rez, J. Villamizar, and M. Altuve, "Human activity recognition on smartphones using a SVM and K-NN" ,Symposium on Image Signal Processing and Artificial Vision (STSIVA),pp. 1-5, 2019.

[13] C.A. Ronao, and S.B. Cho, "Human activity recognition with smartphone sensors using deep neural networks," Expert systems with applications, vol. 59, pp. 235-244, 2016.

[14] A. M. Khan, Y.K Lee, S.Y. Lee, and T.S. Kim, "A Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer," IEEE Trans. Inf. Technol. Biomed, vol. 14, pp. 1166–1172, 2010.

[15] https://www.cis.fordham.edu/wisdm/dataset.php

[16] R. Santhoshkumar, M. kalaiselvigeetha and J. Arunnehru, "SVM – KNN based Emotion Recognition of Human in Video using HOG Feature and KLT Tracking Algorithm", International Journal of Pure and Applied Mathematics 117 (15), 621-634, 2017.

[17] R. Santhoshkumar, M. kalaiselvigeetha and J. Arunnehru, "Activity Based Human Emotion Recognition in video", International Journal of Pure and Applied Mathematics 117 (15), 1185-1194, 2017.

ISBN: 978-1-6654-7436-8/22/$31.00 ©2022 IEEE