Machine Learning Model for Human Activity Analysis

Alia Rifat, Pratiksha Pradip Pandao, and B. Shoban Babu

ABSTRACT

Human Activity Recognition is an active subject of research and scientific progress in which several models have been presented for identifying and categorizing activities using Machine Learning utilizing various methodologies. The purpose of human activity recognition is to look at activities in video or still photos. Human activity recognition systems are motivated by this fact, and their goal is to appropriately classify input data into its underlying activity category. Human activities are classified as (a) gestures, (b) atomic actions, (c) human-to-object or human-to-human interactions, (d) collective actions, (e) behaviors, and (f) events, depending on their complexity. Today, health informatics is a critical field for improving healthcare efficiency by streamlining the collecting, storage, and retrieval of critical patient health data. In this paper, an intelligent smart healthcare system is provided that uses machine learning approaches to deliver ubiquitous human activity recognition (HAR) in an automated manner. The goal is to model and recognize activities of everyday living in a precise and efficient manner. Furthermore, for HAR purposes, we focus on a dataset collecting body motion and vital sign recordings from volunteers of various profiles while performing various physical activities. This research has demonstrated that identifying human activity from sensor data is extremely difficult, even with the availability of a number of machine learning approaches. When it comes to machine learning techniques, there is no one-size-fits-all approach.

Keywords: IoT, LeenaBOT, machine learning, smart healthcare.

A. Rifat*
LeenaBOT Robotics pvt Ltd, India.
(e-mail: jodhyy@gmail.com)
P. P. Pandao
LeenaBOT Robotics pvt Ltd, India.
(e-mail: p3pandao@gmail.com)
B. S. Babu
SV Engineering College Tirupati India.
(e-mail: bshobanbabu@gmail.com)

*Corresponding Author

I. INTRODUCTION

The difficulty of identifying sequences of accelerometer data captured by specialized harnesses or smart phones into recognized well-defined movements is known as human activity recognition. The ability to classify what activity a person is engaged in at any given time allows computers to provide support and instruction to that person prior to or throughout a task. The challenge arises from the diversity of our movements as we go about our daily tasks [1]. Hand-crafting features from time series data based on fixed-sized windows and training machine learning models, such as ensembles of decision trees, are two traditional approaches to the problem. The problem is that this type of feature engineering necessitates extensive knowledge in the field.

Deep learning approaches such as recurrent neural networks and one-dimensional convolutional neural networks, or CNNs, have recently been shown to produce state-of-the-art results on demanding activity recognition tasks using feature learning on raw data rather than data feature engineering. One of the most exciting areas of research is human activity identification utilizing smartphone sensors such as accelerometers. One of the time series categorization problems is the HAR. Various machine learning and deep learning models were used in this project to achieve the best final result. In the same way, we may utilize the Recurrent Neural Network (RNNLSTM)'s (long short-term memory) model to distinguish various human movements such as standing, climbing upstairs and downstairs, and so on. The LSTM model is a recurrent neural network that can learn order dependency in sequence prediction issues. This model is employed because it aids in the recall of values over arbitrary time spans. Walking, Upstairs, Downstairs, Sitting, and Standing are examples of human activity recognition activities [2].

The goal of this study is to look at a dataset of sensor data from human behaviors from a large number of people and see whether we can evaluate it, draw conclusions, and predict the activity using Machine Learning. We also try to see if we can recognize the participants based on their walking styles and derive additional conclusions. The possibilities of such a study can be used to scenarios such as activity detection, tiredness monitoring, individual identification, and probable deployment in extremely sensitive and secure workplaces, among others. Multiple applications of these models have resulted from the employment of new and improved technologies in this field by various researchers and developers [3]. The actions going place in real time can be tracked in a very effective and optimal manner thanks to such highly trained models. Unusual or suspicious behavior can be dealt with using simple measures that ensure peace and harmony in the living world. With the support of regular monitoring, this can also be very valuable in developing a smart home environment as well as a smart healthcare

service. Many security issues may be managed with attention, and the potential for harm can be reduced [4]. People's psychological well-being can also be ensured by such successful implementation of these models in day-to-day life, without fear of harm from such activities.

The camera module can be used to implement the human activity recognition model, which captures the raw data that is used as an input to the recognition system. After feature extraction, activity categorization is done by producing different frames of such input data. This action is then classified as either normal or suspicious, and an alarm is forwarded to the appropriate authorities. Fig. 1 depicts Human Activity Recognition Methods.
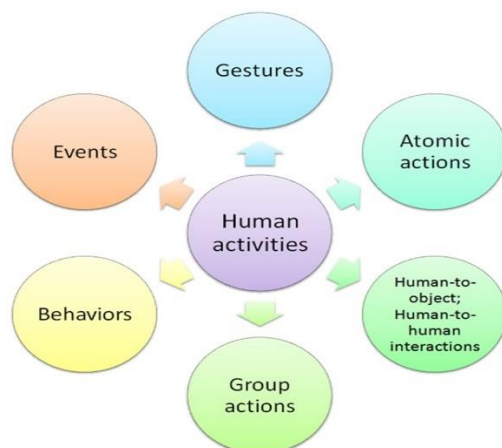

Fig 1. Human Activity Recognition Methods

II. PROBLEM DEFINITION

Following an assessment of the town's activities, it was discovered that many of them are of an unethical nature and are detrimental to people's well-being. Many other activities were also noticed, all of which had an influence and had consequences as a result of the absence of monitoring and sufficient care. These acts not only harm tangible items, but they also have a psychological influence on people. A lot of money is lost as a result of such acts. People, particularly youngsters and the elderly, must face consequences as a result of a lack of effective monitoring and care, which can be as terrible as the death of their loved ones. The faith in security infrastructure has been harmed as a result of numerous similar acts. To address all of these issues, we developed a Human Activity Recognition system based on machine learning. A camera module has been included in this system, which aids in the capture of real-time action sequences. Segmentation further divides this action sequence across multiple frames. Datasets extract features from input data, and the nature of activity is determined. Depending on the nature of the activity, it is regarded as either normal or suspicious. If the identified activity is of an abnormal type, an instant alarm signal is issued to the authority, and the adverse implications of such action can be minimized. In this research, an attempt is made to propose a viable remedy to such societal challenges. We've come to a conclusion with various solutions, which will help researchers and developers improve the system further.

III. LITERATURE SURVEY

Reference [5] employed principal component analysis to de-noise the signal before using a short-time Fourier transform to extract features from channel state information CSI. These characteristics were then fed into classifiers like the random forest (RF) and hidden Markov model (HMM). The findings were compared to those of a DL technique based on an LSTM, which did not require de-noising or feature extraction because these tasks were handled by the model. The RF and HMM models, respectively, scored 64.67 percent and 73.33 percent. The LSTM model had a 90.5 percent accuracy rating, which was almost 17 percent higher than the HMM model and about 26 percent higher than the RF model. Although the researchers acknowledged that LSTMs were substantially slower to train, these results suggest that DL models can outperform traditional approaches.

Reference [6] classified four separate activities using CSI data obtained from numerous access points. A DL model is contrasted to an SVM technique, which uses numerous inputs into a convolutional neural network (CNN), which is then integrated into a fully connected layer and ultimately into a classifier. The results suggest that the DL approach improves data modelling accuracy by learning to model the data better. The CSI data is converted into spectrographs, which are then separated into time periods in this project.

Reference [7] Using WiFi data, researchers compared an LSTM to one of the most prevalent univariate time series models, the Auto-Regressive Integrated Moving Average (ARIMA) model. The root mean square error (RSME) of LSTM predictions was reduced by 80.9 percent to 93.4 percent, demonstrating once again that the DL technique is more promising in this area.

Reference [8] introduced a DL approach to HAR known as the Auto-Encoder Long-term Recurrent Convolution Network (AE-LRCN). To learn representations and reduce inherent noise, an autoencoder was utilized. The raw CSI time window was turned into a latent space of 256 features, whereas our proposed solution used a CAE to compress the CSI time window into a latent space of 12 features, in addition to removing noise.

Reference [9] A robust HAR system based on coordinate transformation, PCA (CT-Principal Component Analysis), and online SVM has been developed. To limit the effect of orientation fluctuations, the CT-PCA technique is applied. Their OSVM is self-contained and only utilizes a small amount of data from an unknown location.

Reference [10] proposed a HAR technique using smartphone accelerometer data to alter position and orientation. They employ the magnitude of their acceleration as the fourth data dimension. In the model, they also use generic SVM and location specific SVM.

Reference [11] many academics rely on generic machine learning algorithms since they require a large amount of training data. As a result, the focus of our paper is on generic ways to analyzing and detecting human behavior.

IV. PROPOSED SYSTEM ARCHITECTURE

A. General System

Our system constitutes of the following:

1) Hardware

The Camera Module is the major hardware component that has been employed to capture the real-time actions that have been conducted. The video is captured at a minimum of 30 frames per second, which is then stripped down into individual frames to make 30 frames per second. A batch of 16 frames is extracted at a time to examine the activity in progress. The processing unit on which the machine learning model runs is another important component.

2) Cloud

The frames taken from the hardware using the Camera module are examined further in the processing unit to look for any unusual behaviour. When this type of activity is detected, the cloud appears. Cloud is then notified of the questionable conduct, and it can warn the user or the appropriate authority. There are two methods for notifying the authority:

a) Android application-based notification;
b) GSM module for SMS.

We use the cloud as an efficient storage solution because the data we acquire on a daily basis cannot be entirely stored on a physical device. Data stored in the cloud can be accessed from anywhere.

B. Android Application

On the Android application, you may see the data that is collected on the cloud. They can be real-time data or data that has been previously stored in the cloud. The authority or user can be notified by an Android application in a variety of methods, including notification alerts, vibrations, or flashing lights. This Android software will not only warn the user, but it will also alert the security or police authorities in the area where the action is occurring. On the app, you can send an alert through notification or SMS [12]. Along with the alarm, the location of the suspicious behavior can be sent to the appropriate authority. The user is also given a button that, when pressed, will sound an alarm near the camera or surveillance gadget.

V. DIAGRAMS

Fig. 2 and 3 depict the system architecture and flow chart.
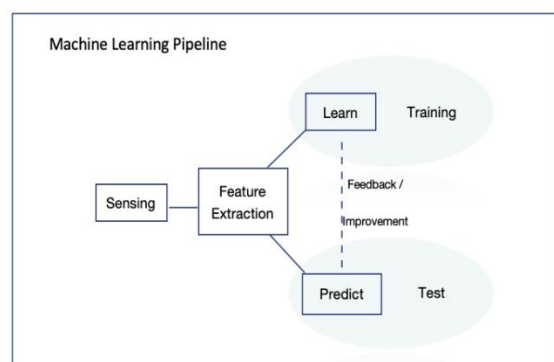
A. System Architecture

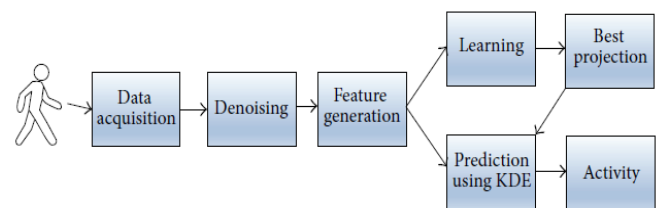
Fig 2. System architecture.

B. Flowchart

Fig 3. Flowchart.

VI. DATASETS

This was a 6-category classification issue. The classes were chosen to represent the usual range of activities that a person might perform on a regular basis: sit, stand, walk, lay, fall, and empty. The lay class consisted of a person lying down on a desk, which was supposed to represent a person lying down on a bed [13]. A person falling and remaining on the ground was simulated in the fall class. These two exercises were chosen to demonstrate how this DL approach could be applied in a practical setting. There were two sets of data obtained. The first, Set A, collected all of the samples of each activity on the same day; the second, Set B, gathered three samples of all of the activities on the same day over seven days. Set A included 18 samples of each activity over the course of 160 seconds; the first and last 55 seconds were of the empty room, while the middle 60 seconds were of the activity. Set B was gathered in a different way, with three random sequences of each activity lasting 50 seconds each, totaling 970 seconds of CSI data every day [14]. Set B comprised a total of 110 samples and was used to assess the setup's temporal robustness.

The centre 40 seconds of the samples were subsectioned and dissected into two 20-second samples to guarantee that only the relevant activity was done in each sample. Each of the 20-second samples was successively stacked into 50 percent overlapping 4 second windows, which was used as the input data for the CAE, as mentioned above in the pre-process section. Both datasets are accessible upon request from the relevant author [15].

VII. EXPERIMENT DESIGN

The NVIDIA GeForce GTX 1080 graphics card with 8GB of memory was used for all training and testing. The operating system was Ubuntu 16.04.3, Python was version 3.6, and TensorFlow was version 1.9. The CAE was trained with a mini-batch size of 128 whereas the TCN and LSTM were trained with a batch size of 32 due to the small amount of training samples. Set A 5-fold cross-validation was employed with 70% of the samples being utilized for the training set, 10% for the validation set, and 20% for the test set. SetB was tested using 7-fold cross-validation, with 5 days of training samples, 1 day of validation samples, and 1 day of testing samples.

VIII. RESULT ANALYSIS

Understanding human activities has become one of the most popular study issues in computer vision. The type and

amount of data used by each method is determined by the underlying algorithm's ability to handle diverse and/or large-scale data. Due to crowded backgrounds, complex camera motion, substantial intraclass variances, and data collecting challenges, developing a completely automated human activity detection system is a difficult undertaking [16].

The camera first detects the presence of a human, then the activity that is taking place. The deep learning model utilized is ResNet, which generally employs 2D kernels, however 3D kernels are enabled to improve the efficiency of activity recognition. The kinetics400 dataset, which identifies 400 different types of activities, was used to train the activity recognition model. In addition, 400 films of each action are available to improve accuracy [17].

For activity identification, the ResNet model for human activity recognition with 2D CNN employs 150 layers. This 2D CNN ResNet model handles tasks such as detection, segmentation, and captioning, but our model additionally includes a 3D CNN that processes videos rather than images, increasing the chances of action recognition success. Detection, summarization, optical flow, segmentation, and captioning are some of the activities that can be completed [18]. As a result, as shown in table I, better accuracy in human activity recognition is attained. The LSTM is successful, with an overall average accuracy of 80.56 percent.

TABLE I: LSTM CONFUSION MATRIX

	Empty	Sit	Stand	Walk	Lay	Fall
Empty	100	0	0	0	0	0
Sit	0	100	0	0	0	0
Stand	0	0	83.3	0	16.7	0
Walk	0	0	0	100	0	0
Lay	0	0	33.3	0	66.7	0
Fall	50	0	0	0	16.7	33.3

IX. CONCLUSION

We have suggested a machine learning-based Human Behavior Recognition system that deals with identifying activity as normal or suspicious based on its nature. Long short-term memory (LSTM), Bi-directional Long short-term memory (Bi-directional LSTM), Convolutional Neural Network (CNN) [19], and Convolutional LSTM were used to classify the majority of the activities. Our best performing model has an average classification accuracy of more than 89 percent based on our analysis of 14 different activities. Despite the fact that earlier findings revealed that six human activities were employed in the research, our study followed numerous 06 human activities that are more generic than those used in past significant studies [20]. If more than 15 or 20 human activities are used, the results may differ. Researchers in the future should look into the effects of more human activity. However, our findings only provide the necessary generalization for non-hand-oriented activity recognition scenarios. If any unusual activity is detected, an immediate alert notification is issued to authority, allowing for the reduction of further depressing repercussions. In the future, we'd like to investigate the performance of different machine learning techniques. In this experiment, we learned that simple machine learning algorithms can perform well with correct parameter tweaking, and that statistical testing can be used to establish the relevance of the results.

X. FUTURE SCOPE

Human Activity Recognition can help with smart home monitoring, healthcare services, security surveillance, and childcare, among other things. We may upgrade our programme in the future by employing object activity recognition, which allows us to track and analyse the activities that objects conduct. The use of combined huge datasets can be used to identify activities occurring at a slower rate of time. The system should be able to detect even the tiniest or most minute alterations [21]. If the actor performing the anomalous action is not captured in the first place, the data of the actor can be preserved, and the individual identified [22]. Activities that repeat themselves should be saved in order to save time and space throughout the recognition process. Such a paradigm can likewise be implemented in the government authority sector. Much more work can be done in terms of improving accuracy and dealing with difficulties such as optical identification and image background clutter.

REFERENCES

[1] Kanade P, Prasad JP. Machine Learning Techniques in Plant Conditions Classification and Observation. *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* IEEE. 2021; 729-734.

[2] Turaga P, Chellappa R, Subrahmanian VS, Udrea O. Machine recognition of human activities: a survey. *IEEE Transactions on Circuits and Systems for Video Technology*. 2008; 18(11): 1473-1488.

[3] Lee D, Yang M-H, Oh S. Fast and accurate head pose estimation via random projection forests. *Proceedings of the IEEE International Conference on Computer Vision (ICCV '15)*. IEEE. 2015:1958-1966.

[4] Yousefi S, Narui H, Dayal S, Ermon S, Valaee S. A survey on behavior recognition using wifi channel state information. *IEEE Comms Mag.* 2017; 55(10).

[5] Li H, Ota K, Dong M, Guo M. Learning human activities through wi-fi channel state information with multiple access points. *IEEE Com Mag.* 2018; 56(5).

[6] Qolomany B, Al-Fuqaha A, Benhaddou D, Gupta A. Role of deep lstm neural networks and wi-fi networks in support of occupancy prediction in smart buildings. *IEEE 19th Int. Conf. on HPCC; IEEE 15th Int. Conf. on SmartCity; IEEE 3rd Int. Conf. on DSS.* 2017: 50-577.

[7] Zou H, Zhou Y, Yang J, Spanos CJ. Towards occupant activity driven smart buildings via wi-fi enabled iot devices and deep learning. *Energy and Buildings*. 2018; 177: 12-22.

[8] Kanade P, Prasad JP. Arduino Based Machine Learning and IoT Smart Irrigation System. *International Journal of Soft Computing and Engineering (IJSCE)*. 2021; 10(4): 1-5.

[9] Agarwal I, Kushwaha AKS, Srivastava R. (2015). *Weighted Fast Dynamic Time Warping Based Multiview Human Activity Recognition Using a RGB-D Sensor*. [Internet] 2015. Available from: https://ieeexplore.ieee.org/ document / 7490046.

[10] Sorkun MC, Danişman AE, Durmaz İncel Ö. *Human Activity Recognition With Mobile Phone Sensors: Impact Of Sensors And Window Size*. [Internet] 2015. Available from: https://ieeexplore.ieee.org /document/ 8404569.

[11] Kanade P, Prasad JP, Kanade S. IOT based Smart Healthcare Wheelchair for Independent Elderly. *European Journal of Electrical Engineering and Computer Science*. 2021; 5(5): 4-9.

[12] Käse N, Babaee M, Rigoll G. *Multi-view human activity recognition using motion frequency*. [Internet] 2017. Available from: https://sigport.org /sites/default /files /docs/ICIP_PaperID1443_Final.pdf.

[13] Kanade P, Prasad JP. Mobile and Location Based Service using Augmented Reality: A Review. *European Journal of Electrical Engineering and Computer Science*. 2021; 5(2): 13-18.

[14] Chen Z, Zhu Q, Soh YC, Zhang L. Robust human activity recognition using smartphone sensors via ct-pca and online svm. *IEEE Transactions on Industrial Informatics*. 2017; 13(6): 3070-3080.

[15] Sun L, Zhang D, Li B, Guo B, Li S. Activity recognition on an accelerometer embedded mobile phone with varying positions and

orientations. *International conference on ubiquitous intelligence and computing. Springer,* 2010: 548–562.

[16] Jahangiri A, Rakha HA. Applying machine learning techniques to transportation mode recognition using mobile phone sensor data," *IEEE transactions on intelligent transportation systems.* 2015; 16(5): 2406-2417.

[17] Kanade P, Kanade S. Medical Assistant Robot ARM for COVID-19 Patients Treatment – A Raspberry Pi Project. *International Research Journal of Engineering and Technology.* 2020; 7(10): 105-111.

[18] Ji S, Xu W, Yang M, Yu K. 3D *Convolutional Neural Networks for Human Action Recognition.* [Internet] 2012. Available from: https://ieeexplore.ieee.org /document / 6165309.

[19] Simonyan K, Zisserman A. *Two-Stream Convolutional Networks for Action Recognition in Videos.* [Internet] 2014. Available from: https://arxiv.org/abs/1406.2199.

[20] Ng JH-Y, Hausknecht M, Vijayanarasimhan S, Vinyals O, Monga, Toderici RG. *Beyond Short Snippets: Deep Networks for Video Classification.* [Internet] 2015. Available from: https://arxiv.org/abs/1503.08909.

[21] Kanade P, Alva P, Prasad JP, Kanade S. Smart Garbage Monitoring System using Internet of Things (IoT). *2021 5th International Conference on Computing Methodologies and Communication (ICCMC).* IEEE. 2021: 330-335.

[22] Kanade P, Alva P, Kanade S, Ghatwal S. Automated Robot ARM using Ultrasonic Sensor in Assembly Line. *International Research Journal of Engineering and Technology (IRJET).* 2020; 7(12): 615-620.