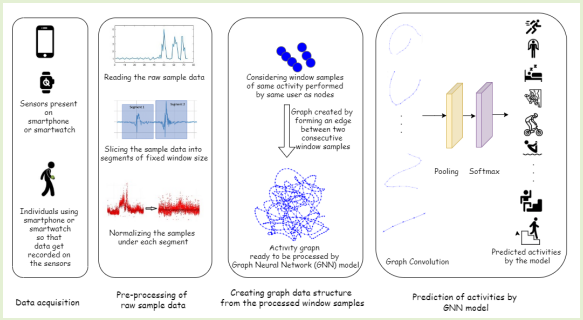
A New Framework for Smartphone Sensor-Based Human Activity Recognition Using Graph Neural Network

Riktim Mondal, Debadyuti Mukherjee, Pawan Kumar Singh, *Member, IEEE*, Vikrant Bhateja, *Senior Member, IEEE*, and Ram Sarkar, *Senior Member, IEEE*

***Abstract*—Automatic human activity recognition (HAR) through computing devices is a challenging research topic in the domain of computer vision. It has widespread applications in various fields such as sports, healthcare, criminal investigation and so on. With the advent of smart devices like smartphones, availability of inertial sensors like accelerometer and gyroscope can easily be used to track our daily physical movements. State-of-the-art deep neural network models like Convolutional Neural Network (CNN) do not need any additional feature extraction for such applications. However, it requires huge amount of data for training which is time consuming, and requires ample resource. Another limiting factor of CNN is that it considers only the features of an individual sample for learning without considering any structural information among the samples. To address the aforesaid issues, we propose an end-to-end fast Graph Neural Network (GNN) which not only captures the individual sample information efficiently but also the relationship with other samples in the form of an undirected graph structure. To the best of our knowledge, this is the first work where the time series data are transformed into a structural representation of graph for the purpose of HAR using sensor data. Proposed model has been evaluated on 6 publicly available datasets, and it achieves nearly 100% recognition accuracy for all the 6 datasets. Source code of this work is available at https://github.com/riktimmondal/HAR-Sensor.**



***Index Terms*— Human activity recognition, graph neural network (GNN), message passing, smartphone sensors, deep learning.**

I. Introduction

IN RECENT years with the availability of smart devices like smartphones, our society has witnessed a tremendous change in the field of multimedia communication. Not only this, these devises are also used in monitoring our daily activities. Sensors like accelerometers, gyroscopes, proximity and magnetometers, etc. which come as inbuilt sensors in the smart devices can help to measure our physical activities. This has led to an interesting domain of research called Human Activity Recognition (HAR) which finds its application in healthcare [1], sports activities [2], smart home systems [3], and criminal surveillance [4].

HAR can be defined as the task of distinguishing a physical activity completed by an individual subject to a hint of development inside a specific domain. Daily activities, for example, *Walking, Sitting, Standing,* and *Climbing steps* etc. are considered as conventional physical events and build our class of actions.To record these events or change in events, sensors, for example, tri-axial accelerometer and gyroscopes, capture information while the movement is being performed. Different physiological signals, for example, heartbeat, breath, and so forth and natural signals, for example, temperature, time, pressure, and so on are also recorded which help to

Manuscript received July 18, 2020; revised August 5, 2020; accepted August 5, 2020. Date of publication August 11, 2020; date of current version April 16, 2021. The associate editor coordinating the review of this article and approving it for publication was Prof. Tohru Kawabe. *(Corresponding author: Pawan Kumar Singh.)*

Riktim Mondal, Debadyuti Mukherjee, and Ram Sarkar are with the Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India (e-mail: riktim97@gmail.com; debadyuti23@gmail.com; rsarkar@ieee.org).

Pawan Kumar Singh is with the Department of Information Technology, Jadavpur University, Kolkata 700106, India (e-mail: pawankrsingh@ieee.org).

Vikrant Bhateja is with the Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial Group of Professional Colleges (SRMGPC), Lucknow 226028, India which is currently affiliated under Dr. A. P. J. Abdul Kalam Technical University, Lucknow 226031, India (e-mail: bhateja.vikrant@ieee.org).

Digital Object Identifier 10.1109/JSEN.2020.3015726

develop a better HAR system. An important application of HAR is in healthcare domain, particularly in helping restoration, physiotherapist, and old age caring management and psychological weakness.

Traditional HAR strategies have gained a significant improvement in finding noteworthy data from scores of low-level readings. However, such models are fruitful for information gathered in controlled environment as well as for limited number of activities. Recognition of complex human activities are restricted due to the naive feature extraction procedures and restriction in field awareness. The simple feature extraction procedures lower the performance of the learning algorithms and do not work well for some closely related human activities such as *Walking* and *Jogging*. On the other hand, deep learning models have the abilities to automatically learn features from the raw data with the higher precision [5].

In the earlier days, traditional machine learning models like Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) were used for HAR, where additional time and frequency domain features were needed to be extracted from the raw sensor data. Micucci *et al.* in [6] used SVM, and obtained an accuracy of 98.71% on uniMiB-SHAR dataset. KNN was used in the work reported in [7] and Random Forest (RF) was used in [8]. Both the works considered WISDM dataset [9] for experimentation, and achieved an accuracy of 96.2% and 98.09% respectively. Dadkhahi and Marlin [10] introduced tree based firm cascade linear classifier and achieved showed 97% accuracy on PAMAP2 dataset [11]. A Variational Deep Embedding (VaDE) clustering model was introduced in [12] to cluster different activities obtaining accuracy of 84.46% on HHAR dataset [13]. For USC-HAD dataset, a classical model of SVM was used by [14] to produce an accuracy of 95% where they performed feature extraction followed by classification using SVM to predict motion category for each subject.

Later with the availability of computation resources Deep Neural Network (DNN) based models like Convolutional Neural Network (CNN) [15], Recurrent Neural Network (RNN) [16] etc. become state-of-the-art models for various classification tasks including HAR. The authors of [17] used a CNN based approach to achieve an accuracy of 91.7% by extracting features from 3-D raw accelerometer data. Deep RNN (DRNN) was introduced by [18] to study different types of sequential models on various datasets. Authors of [19] used combination of multi-channel Deep CNN (DCNN) and multilayer backpropagation on PAMAP2 dataset to extract features and obtained an accuracy of 90.8%. A new model based on Hierarchical deep long short-term memory unit named as H-LSTM was introduced in [20] to perform HAR on HHAR dataset [13] where they achieved 90% accuracy. Although satisfactory results are obtained using such networks but it needs a huge amount of labeled dataset which is time consuming and hard to annotate correctly. Besides, most of these models require heavy computation power and take a very long time to train the model. Another important factor which limits the success of these models is that at the time of learning, these models learn only from the input feature of each sample independently without considering any structural and relational information among the samples. This is one of the key reasons for the conventional DNN based models to perform poorly when human activities show close structural resemblance.

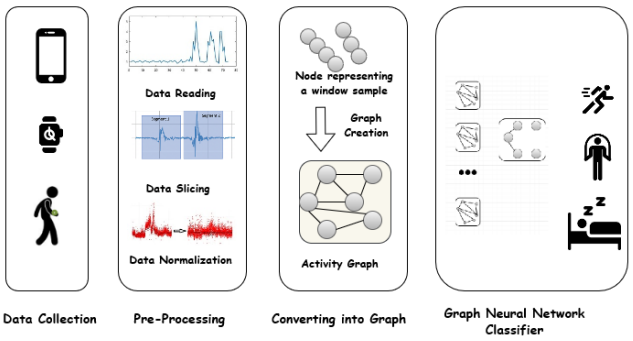


Fig. 1. Generic framework for transforming time-series based sensor data into a graph structure for HAR using GNN based approach.

A. Motivation

The motivation of this work is to represent such unstructured time-series sensor data into a form of graph (i.e., a structured way of representing the data) where each node corresponds to a single sample of activity with undirected edges connecting other samples belonging to the same activity. This very structural representation of nodes with edges connecting other nodes helps to capture the information from both individual samples as well as from the neighboring samples. To the best our knowledge, this is the first attempt where the concept of Graph Neural Network (GNN) is applied for the task of HAR using sensor data. GNNs are Artificial Neural Network (ANN) or connectionist network models which can capture the structural dependence in a graph through sub-graphs and nodes via message passing technique between the nodes of the graph. Although GNN [21] is a very recent concept, researchers have started applying it in various domains like Recommender Systems, Image classification and Language Modeling.

B. Contributions

In this article, a new end-to-end trainable GNN based framework is proposed for sensor-based HAR problem. The schematic diagram of the framework is shown in Fig. 1. Our work also gives a complete perspective on HAR framework and describes different issues related to the GNN structure. It further endeavors to decrease the computational needs and improve the performance by using graph based strategies. The data are classified at the graph level, i.e., each node of a graph belongs to the same activity performed by a particular user. The main *contributions* of this article are listed below:

1) A generic dataset preparation framework is proposed to transform any sensor based HAR data into a graph structure for training with graph based models.
2) A very fast and effective GNN model built upon GraphConv message passing technique [22] is developed which learns from neighborhood features along with its

own features that correspond to both local and global feature learning for graph based classification.

3) Extensive experiments are performed on 6 standard datasets to produce state-of-the-art results with very fast training and evaluation on low-resource systems.

II. Data Preparation

In this section, we describe the datasets considered here for experimentation. We have considered 6 publicly available HAR datasets namely MobiAct [23], WISDM [9], MHEALTH [24], PAMAP2 [11], HHAR [13] and USC-HAD [25] for evaluating our proposed GNN based model. A tabular description of the HAR datasets on the basis of number of activity classes, range of number of graphs obtained per activity as well as sensors used for data collection is provided in Table I.

The data pre-processing and preparation techniques for the 6 HAR datasets have been discussed below.

1) At first we have divided our dataset user-wise, and for each user we have further divided the samples on the basis of activities performed by a user. Hereafter, such sub-divisions will be referred to as slices.
2) The features are normalized for each slice so that all the values of a feature are centralized. Normalization is done by subtracting a feature value by the mean of all the values of that feature present in the slice and later by dividing by the standard deviation of the values.
3) Since each slice contains a very large number of samples, we have further pre-processed the samples into non-overlapping windows. Each window has the maximum size of 128 samples i.e., we consider a window of 6.4 seconds duration for the WISDM, MobiAct datasets, 2.56 seconds duration for the MHEALTH dataset, 1.28 seconds for PAMAP2, HHAR and USC-HAD datasets.
4) For each window, we take the aggregate feature value as the mean of all the sample values present in the window for the corresponding feature.
5) Next we have considered each window as a node and created a graph by taking a maximum number of 600 nodes per graph. Since each slice contains less than 600 nodes so, there is one graph per slice. From the given nodes, edges of the graph are obtained as follows:

To predict an activity, we need to know its nature i.e., how the feature values change with time. So, two consecutive windows can be used as an estimate of two consecutive time-periods. Following this logic, we have connected every two consecutive nodes by an undirected edge. Therefore, for each graph the nodes corresponding to the first and last windows have degree 1, and the rest of the nodes have degree 2. Some samples of the generated graphs of MobiAct, WISDM, MHEALTH, PAMAP2, HHAR and USC-HAD datasets are shown in the Fig. 2. The source code for generation of graph attributes i.e., nodes, edges, feature values of a node, labels of node and graph can be found at.[1]

[1]https://github.com/riktimmondal/HAR-Sensor

TABLE I
DESCRIPTION OF THE HAR DATASETS USED IN THIS EXPERIMENT(HERE A, G, M AND E DENOTE ACCELEROMETER, GYROSCOPE, MAGNETOMETER AND ECG RESPECTIVELY WHICH ARE THE SENSORS USED TO COLLECT DATA)

Dataset	Number of Activity classes	Activities Present	Graph range per activity	Sensor(s) used
MobiAct	13	*Standing, Walking, Jogging, Jumping, Stairs Up, Stairs Down, Sit Chair, Car Step In,Car Step Out, Forward Lying, Front Knees Lying, Sideward Lying, Back Sitting Chair*	49-200	A
WISDM	6	*Walking, Jogging, Upstairs, Downstairs, Sitting, Standing*	23-26	A
MHEALTH	12	*Standing still, Sitting and relaxing, Lying down, Walking, Climbing stairs, Waist bends forward, Frontal elevation of arms, Knees bending(Crouching), Cycling, Jogging, Running, Jump front and back*	10	A, G, M, E
PAMAP2	12	*Lying, Sitting, Standing, Ironing, Vacuum cleaning, Descending stairs, Ascending stairs, Walking, Nordic walking, Cycling,Running and Rope jumping*	4-8	A, G, M
HHAR	6	*Biking, Sitting, Standing, Walking, Stair Up,Stair down*	27-33	A
USC-HAD	12	*Walking Forward, Walking Left, Walking Right, Walking Upstairs, Walking Downstairs, Running Forward, Jumping Up, Sitting, Standing, Sleeping, Elevator Up, Elevator Down*	14	A, G

III. Graph Neural Network (GNN)

A graph G is represented by nodes (V) and edges (E) as $G=(V, E)$ where each node contains feature vectors, X_v *for* $v \in V$ and each edge contains feature vectors X_e *for* $e \in E$. Here E represents the connection between two closest neighboring nodes. Classification using GNN is performed at two levels. First, for node classification where each node $v \in V$ of G is labeled as y_v and one needs to predict the labels of each node v by finding an embedding vector h_v such that $y_v=f(h_v)$ where f is a differentiable approximation function.

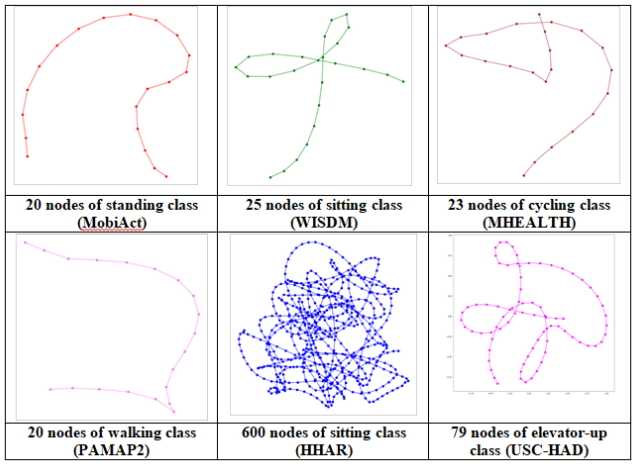


Fig. 2. One sample graph showing an activity class generated for each of the 6 HAR datasets.

f is used to map the relation of similarity between graph nodes in simple embedding dimension with the actual graph nodes present in the real graph. In embedding dimension, many information is stored in compressed manner through h_v. Second, besides using node information, structure of the graph is also considered for graph classification. Here, a set of graphs $g_1,g_2..,g_n \in G$ with their respective labels $y_1,y_2,..,y_n \in Y$ are provided, where the representation vector h_g needs to be found for every g for predicting each graph label through $y_g = f_1(h_g)$. Here f_1 is similar to f but it stores the entire graph representation instead of nodes. In some situations, the feature vector at edges i.e. X_e are also considered where directed edge graphs with additional edge attributes are present. In our case, we have considered a graph with undirected edges containing only node feature attributes.

GNN uses node features X_v and graph structure to learn the embedding vector of node h_v and the embedding vector of the entire graph h_g. Most of the GNNs work on the principle of message passing technique to collect information from its neighbors and pass it along. It is a neighborhood strategy where the representation of a node is updated iteratively by the aggregated representation of all of its neighboring nodes. So, after k number of iterations, a node representation $h_v^{(k)}$ contains the structural information of all the neighboring nodes which are k-hops away. Here, k number of iterations actually corresponds to kth layer as well, so at $k>0$layer, the final representation vector of a node is represented as:

$$h_v^{(k)} = NON-LINEARITY(COMBINE(h_v^{(k-1)}, AGGREGATE(h_u^{(k-1)}) : u \in N(v))) \quad (1)$$

Here *NON-LINEARITY* is a nonlinear function like Sigmoid, ReLU, ELU; *N(v)* represents all the neighboring nodes of v directly connected to it with an edge. We initialize $h_v^{(0)} = X_v$. The *COMBINE* and the *AGGREGATE* functions can be "add", "mean", "max", "concatenation" as per types of GNN models.

In our present work, we have used ***GraphConv*** as the message passing layer where the representation vector at layer $k>0$ is represented as,

$$h_v^{(k)} = \sigma(h_v^{(k-1)}) * W_1^{(k)} + \sum_{u \in N(v)} h_u^{(k-1)}) * W_2^{(k)}) \quad (2)$$

where $W_1^{(k)}$ and $W_2^{(k)}$ represent weight matrices need to learn by model, σ represents component wise *NON-LINEARITY* function, $\sum$ is the *AGGREGATE* function.

For node classification this $h_v^{(K)}$ vector can be directly used at final *Kth* layer for label prediction. Whereas for entire graph classification all the nodes v present in the graph G, its embedding vectors $h_v^{(K)}$ are aggregated to get the entire graph level prediction using a *READOUT* function.

$$h_G^{(K)} = READOUT(h_v^{(K)} : v \in G) \quad (3)$$

Here *READOUT* is the final aggregation function used to combine all nodes' $v \in G$ embedding vector through operations like "add", "mean" etc. For our task, we have considered both the "mean" and the "max" functions as the aggregation functions.

A. Proposed GNN Based HAR Model

Our network consists of a block of ***GraphConv*** layer with two inputs: (a) X which is the feature matrix of each node of shape[V,D] where V is the total number of nodes in the graph and D is the dimensionality of the input features vector; and (b) the adjacency matrix A of shape [V,V] or edge list E of shape [2,L] consisting of all edges present in the entire graph in the form of pair(V_1,V_2) where V_1 and V_2 are two nodes connected by an edge and L is the total number of edges in the entire graph.

Output of the ***GraphConv*** layer is passed through ReLU function to introduce non-linearity. It is followed by a ***Top-KPooling*** layer which takes X and E as inputs to reduce the number of graph nodes and convert the graphs into sub-graphs which are coarse version of the entire graph capable to represent the high-dimensional representation of the same. This output (O_1) is passed on to another block of the same GraphConv-ReLU-TopKPooling layers whose output (O_2) is added with the output O_1. The final $O_1 + O_2$ output is passed through a linear layer with ReLU activation, dropout layer (dropout rate = 0.5). Then, a linear layer with dimension equal to that of the number of classes of the problem under consideration, with ***Log Softmax*** as the activation function is used to produce the final probability vector Z.

$$LogSoftmax(Z_i) = log(\frac{e^{Z_i}}{\sum_{j=1}^{C} e^{Z_j}}) \quad (4)$$

where, Z_i is the probability/log of its value of ith element in the last linear layer vector and $\sum_{j=1}^{C} e^{Z_j}$ is the sum of all probability values of all jth elements including i in the Z vector for C number of classes. Our proposed GNN model is shown in Fig. 3. We have used Negative Log Likelihood (NLL) function as our classification objective function which needs to be minimized and can be represented as follows,

$$NLL(Z) = -\sum_{i=1}^{C}(y_i * LogSoftmax(Z_i)) \quad (5)$$

where y_i represents the ground truth label of the ith graph.

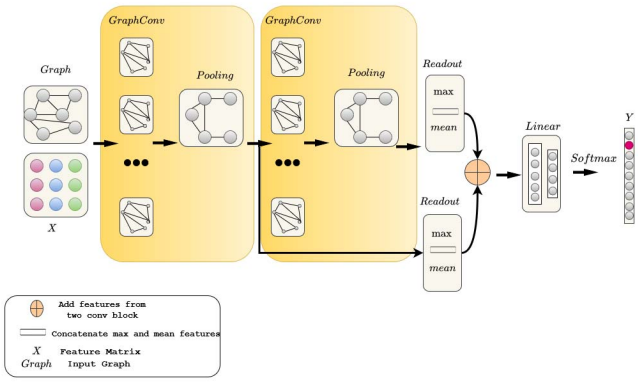


Fig. 3. Proposed GNN model applied for HAR using sensors data.

TABLE II
DISTRIBUTION OF TRAINING AND TESTING SETS AND TIME TAKEN BY OUR MODEL FOR TRAINING AND TESTING FOR EACH DATASET

Dataset	No. of Graph in Train set	No. of Graph in Test set	Time to train (sec)	Time to test (sec)
MobiAct	712	306	5.3675	0.1599
WISDM	125	54	1.5330	0.0483
MHEALTH	84	36	0.7232	0.0257
PAMAP2	58	26	2.02928	0.0264
HHAR	122	53	6.8431	0.0781
USC-HAD	117	51	3.4871	0.0764

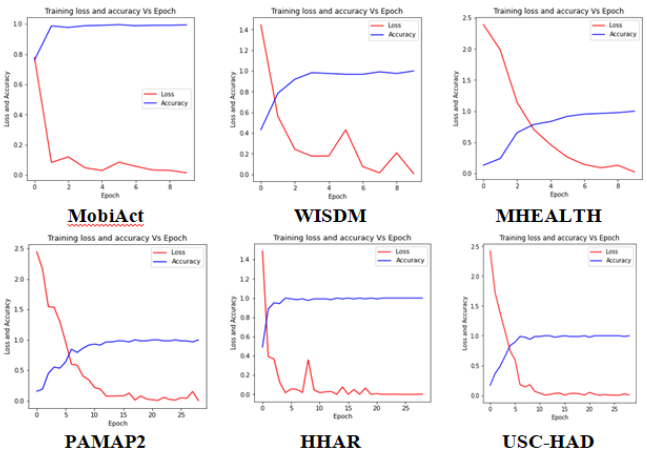


Fig. 4. Plot showing the variation of **training accuracy and training loss** against **epoch sizes** for 6 HAR datasets used in our experiment. This plot is from one particular experiment of our setup out of a total 5 experiments with randomly shuffled datasets.

IV. EXPERIMENTATION

For each of 6 HAR datasets, mentioned earlier, we have split the dataset into train and test sets in the ratio of 70% and 30% upon which the number of graphs obtained for training and testing for each dataset is shown in Table II. We have performed our experiments 5 times, and in each time the dataset is randomly shuffled and split into the said train (70%) and test (30%) sets. Then the model performance is evaluated on the test set. We have displayed the plots of training accuracy and training loss with respect to each epoch for a single experiment for all datasets in Fig. 4.

TABLE III
BEST ACCURACY ON TRAIN AND TEST SETS OBTAINED ON A SINGLE EXPERIMENT OUT OF 5 EXPERIMENTS FOR ALL 6 HAR DATASETS

Dataset	Best train accuracy of single iteration of experiment	Best test accuracy from 30% data of single iteration of experiment
MobiAct	99.86%	100%
WISDM	100%	100%
MHEALTH	100%	100%
PAMAP2	100%	100%
HHAR	100%	100%
USC-HAD	100%	100%

TABLE IV
MEAN TEST ACCURACY CONSIDERING 5 EXPERIMENTS FOR BOTH 30% AND 50% TEST DATASETS

Dataset	Mean Accuracy for (70-30) split	Mean Accuracy for (50-50) split
MobiAct	1.0 ± 0.0	0.9992 ± 0.0015
WISDM	1.0 ± 0.0	1.0 ± 0.0
MHEALTH	1.0 ± 0.0	0.9933 ± 0.0081
PAMAP2	0.9615 ± 0.0486	0.9523 ± 0.0451
HHAR	0.9962 ± 0.0075	0.9931 ± 0.0090
USC-HAD	1.0 ± 0.0	1.0 ± 0.0

A. Setting of Model Parameters

We have trained the model with mini batch Gradient Descent Algorithm [26] with batch size of **16**. Adam optimizer [27] is used with initial learning rate of **0.01** and weight decay of **5e-4**. All the experiments are performed on a machine with Xeon Octa core processor, 8 GBGPU and 32 GB system memory.

B. Model Efficiency

The model is trained with maximum **30** epochs for each dataset. As we claim the model is very fast for training and evaluating, so in Table II we display the time taken for training and testing respectively for each dataset for a single iteration of experiment.

C. Results and Discussion

Here, we show that such a simple model based on GNN is capable of performing classification of human activities with very high accuracy even after considering every activity (from each subject) and every data sample. Most of the researchers have claimed to achieve high accuracy but considered less number of activity classes. Considering one particular experiment out of total 5 experiments, where we have obtained the best results for 70% train and 30% test datasets is shown in Table III. We have also calculated the mean accuracy with standard deviation from all 5 experiments for both 70-30 and 50-50 train and test set splits of the said datasets as shown in Table IV.

To establish the robustness of our model for different datasets, where the number of activities performed varies from minimum 6 to maximum 13, we have shown the effect of progressive addition of training data to the model in order

TABLE V
COMPARISON OF PROPOSED MODEL WITH EXISTING MODELS FOR 6 HAR DATASETS

Dataset	Author	Year	Methodology or Classifier Used	Accuracy(%)
MobiAct	Zheng et al.[28]	2018	TASG applied with SVM	90.5
	Xu et al.[29]	2019	CNN with LSTM Model	98.9
	Zerkouk et al.[30]	2020	Autoencoder with CNN and LSTM	98
	Proposed	**2020**	**GNN model**	**100**
WISDM	Quispe et al.[7]	2018	KNN	96.2
	Zhang et al.[31]	2019	U-Net	96.4
	Burns et al.[32]	2020	Deep triplet embedding	91.3
	Proposed	**2020**	**GNN model**	**100**
MHEALTH	Gumaei et al.[33]	2019	Hybrid Deep Learning Model	99.6
	Chen et al.[34]	2019	Recurrent Convolutional & Attention	94.05
	Burns et al.[32]	2020	Deep triplet embedding	99.9
	Proposed	**2020**	**GNN model**	**100**
PAMAP2	Dadkhahi et al.[10]	2017	Tree structured cascaded network	97
	Priyadharshini et al.[19]	2019	MCDCNN	90.8
	Zhang et al.[35]	2019	HMVAN	94.4
	Proposed	**2020**	**GNN model**	**100**
HHAR	Qin et al.[36]	2019	SSUI	96.41
	Gupta et al.[37]	2020	FECM	89.9
	Qin et al.[38]	2020	GASF Concatenation	96.74
	Proposed	**2020**	**GNN model**	**100**
USC-HAD	Mohammed et al.[39]	2018	LSTM	96.31
	Tahir et al.[40]	2020	SMO	73.33
	Singh et al.[41]	2020	Deep ConvLSTM	94.06
	Proposed	**2020**	**GNN model**	**100**

to obtain a better classification model (see Fig. 5). We have considered a single experiment out of a total 5 experiments where we have iterated the splitting ratio for train set 10% to 90% (test is done on other 90% to 10% data accordingly) in steps of 10% to show improvement of our model for increase in training data. Even for imbalanced dataset where we have considered two cases,(1) when imbalance is in number of graph per class (obtained mean accuracy is 99.37%), and (2) when imbalance is in number of nodes in the graphs of a class (obtained mean accuracy is 99.09%).

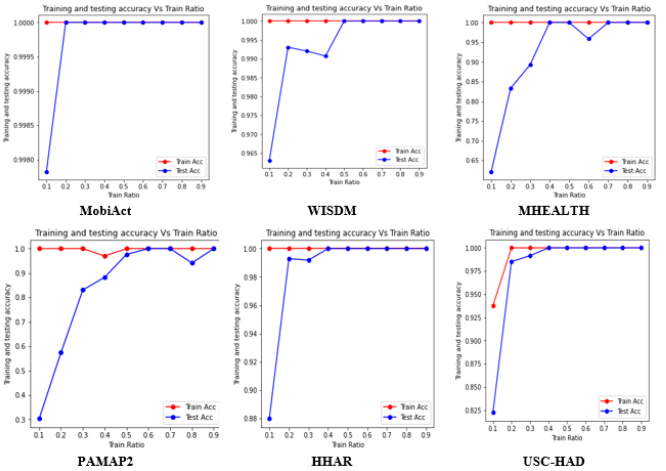


Fig. 5. Progressive **accuracies of train and test sets** with **increasing training data** to the GNN model for 6 HAR datasets.

D. Comparison With Existing HAR Methods

We have also compared our methods with some recent methods. From Table V, it is evident that our model performs better in comparison to existing methods which are considered as state-of-the-art for HAR on sensor data.

V. CONCLUSION

Sensor-based HAR is an active research field since the last decade. However, there are certain perspectives of this HAR problem which can be used to improve the framework and alter the way people make interactions with their smartphones. To this end, we have designed a GNN based model with message passing technique to use the structural information obtained from the time-series sensor data through graph representation. We have set a new benchmark result on 6 standard and publicly available sensor-based HAR datasets. We have also contributed by developing a generic framework to convert time series sensor data into graphical data to be used for training any GNN models on HAR datasets. We have also verified robustness of the model on imbalanced data. For real life applications in future, we can use the framework in online mode where with new collection of sensor data, new nodes and new graph can be created for real time prediction. Also by using transfer learning approach, we can transfer knowledge from one data source to another. Since, here we have used simple GNN model which does not work for input graph whose embeddings cannot be represented as injective function and even different depth and width of the graph may affect in results, so in future, we plan to use other graph models like spline convolution, attention based convolution etc. Besides, we have considered only node feature attributes, in our future work, we will work upon edge feature attributes as well.

REFERENCES

[1] V. Osmani, S. Balasubramaniam, and D. Botvich, "Human activity recognition in pervasive health-care: Supporting efficient remote collaboration," *J. Netw. Comput. Appl.*, vol. 31, no. 4, pp. 628–655, Nov. 2008.

[2] Y.-L. Hsu, S.-C. Yang, H.-C. Chang, and H.-C. Lai, "Human daily and sport activity recognition using a wearable inertial sensor network," *IEEE Access*, vol. 6, pp. 31715–31728, 2018.

[3] S. Ramasamy Ramamurthy and N. Roy, "Recent trends in machine learning for human activity recognition–A survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 4, p. e1254, Jul. 2018.

[4] M. Ehatisham-ul-Haq *et al.*, "Authentication of smartphone users based on activity recognition and mobile sensing," *Sensors*, vol. 17, no. 9, p. 2043, Sep. 2017.

[5] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Skeleton-based action recognition with multi-stream adaptive graph convolutional networks," 2019, *arXiv:1912.06971*. [Online]. Available: http://arxiv.org/abs/1912.06971

[6] D. Micucci, M. Mobilio, and P. Napoletano, "UniMiB SHAR: A dataset for human activity recognition using acceleration data from smartphones," *Appl. Sci.*, vol. 7, no. 10, p. 1101, Oct. 2017.

[7] K. Montero Quispe, W. Sousa Lima, D. M. Batista, and E. Souto, "MBOSS: A symbolic representation of human activity recognition using mobile sensors," *Sensors*, vol. 18, no. 12, p. 4354, Dec. 2018.

[8] K. H. Walse, R. V. Dharaskar, and V. M. Thakare, "Performance evaluation of classifiers on WISDM dataset for human activity recognition," in *Proc. 2nd Int. Conf. Inf. Commun. Technol. Competitive Strategies (ICTCS)*, 2016, pp. 1–7.

[9] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explorations Newslett.*, vol. 12, no. 2, pp. 74–82, Mar. 2011.

[10] H. Dadkhahi and B. M. Marlin, "Learning tree-structured detection cascades for heterogeneous networks of embedded devices," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 1773–1781.

[11] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *Proc. 16th Int. Symp. Wearable Comput.*, Jun. 2012, pp. 108–109.

[12] Z. Jiang, Y. Zheng, H. Tan, B. Tang, and H. Zhou, "Variational deep embedding: An unsupervised and generative approach to clustering," 2016, *arXiv:1611.05148*. [Online]. Available: http://arxiv.org/abs/1611.05148

[13] A. Stisen *et al.*, "Smart devices are different: Assessing and mitigating-mobile sensing heterogeneities for activity recognition," in *Proc. 13th ACM Conf. Embedded Netw. Sensor Syst.*, 2015, pp. 127–140.

[14] O. Politi, I. Mporas, and V. Megalooikonomou, "Human motion detection in daily activity tasks using wearable sensors," in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2014, pp. 2315–2319.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[16] T. Mikolov, S. Kombrink, L. Burget, J. Cernocky, and S. Khudanpur, "Extensions of recurrent neural network language model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 5528–5531.

[17] W. Xu, Y. Pang, Y. Yang, and Y. Liu, "Human activity recognition based on convolutional neural network," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 165–170.

[18] A. Murad and J.-Y. Pyun, "Deep recurrent neural networks for human activity recognition," *Sensors*, vol. 17, no. 11, p. 2556, Nov. 2017.

[19] J. M. H. Priyadharshini, S. Kavitha, and B. Bharathi, "Comparative analysis of multilayer backpropagation and multi-channel deep convolutional neural network for human activity recognition," in *Proc. AIP Conf.*, 2019, Art. no. 030014.

[20] L. Wang and R. Liu, "Human activity recognition based on wearable sensor using hierarchical deep LSTM networks," *Circuits, Syst., Signal Process.*, vol. 39, no. 2, pp. 837–856, Feb. 2020.

[21] J. Zhou *et al.*, "Graph neural networks: A review of methods and applications," 2018, *arXiv:1812.08434*. [Online]. Available: http://arxiv.org/abs/1812.08434

[22] C. Morris *et al.*, "Weisfeiler and Leman go neural: Higher-order graph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 4602–4609.

[23] G. Vavoulas, C. Chatzaki, T. Malliotakis, M. Pediaditis, and M. Tsiknakis, "The MobiAct dataset: Recognition of activities of daily living using smartphones," in *Proc. Int. Conf. Inf. Commun. Technol. Ageing Well e-Health*, 2016, pp. 143–151.

[24] O. Banos *et al.*, "mHealthDroid: A novel framework for agile development of mobile health applications," in *Proc. Int. Workshop Ambient Assist. Living*. Cham, Switzerland: Springer, 2014, pp. 91–98, doi: 10.1007/978-3-319-13105-4_14.

[25] M. Zhang and A. A. Sawchuk, "USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors," in *Proc. ACM Conf. Ubiquitous Comput.*, 2012, pp. 1036–1043.

[26] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*. [Online]. Available: http://arxiv.org/abs/1609.04747

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[28] Z. Zheng, J. Du, L. Sun, M. Huo, and Y. Chen, "TASG: An augmented classification method for impersonal HAR," *Mobile Inf. Syst.*, vol. 2018, pp. 1–10, Dec. 2018.

[29] J. Xu, Z. He, and Y. Zhang, "CNN-LSTM combined network for IoT enabled fall detection applications," in *Proc. J. Phys., Conf.*, 2019, vol. 1267, no. 1, Art. no. 012044.

[30] M. Zerkouk and B. Chikhaoui, "Spatio-temporal abnormal behavior prediction in elderly persons using deep learning models," *Sensors*, vol. 20, no. 8, p. 2359, Apr. 2020.

[31] Y. Zhang, Z. Zhang, Y. Zhang, J. Bao, Y. Zhang, and H. Deng, "Human activity recognition based on motion sensor using U-Net," *IEEE Access*, vol. 7, pp. 75213–75226, 2019.

[32] D. M. Burns and C. M. Whyne, "Personalized activity recognition with deep triplet embeddings," 2020, *arXiv:2001.05517*. [Online]. Available: http://arxiv.org/abs/2001.05517

[33] A. Gumaei, M. M. Hassan, A. Alelaiwi, and H. Alsalman, "A hybrid deep learning model for human activity recognition using multimodal body sensing data," *IEEE Access*, vol. 7, pp. 99152–99160, 2019.

[34] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, and F. Nie, "A semi-supervised recurrent convolutional attention model for human activity recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1747–1756, May 2020.

[35] X. Zhang, Y. Wong, M. S. Kankanhalli, and W. Geng, "Hierarchical multi-view aggregation network for sensor-based human activity recognition," *PLoS ONE*, vol. 14, no. 9, Sep. 2019, Art. no. e0221390.

[36] Z. Qin *et al.*, "Learning-aided user identification using smartphone sensors for smart homes," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7760–7772, Oct. 2019.

[37] A. Gupta, H. P. Gupta, B. Biswas, and T. Dutta, "A fault-tolerant early classification approach for human activities using multivariate time series," *IEEE Trans. Mobile Comput.*, early access, Feb. 13, 2020, doi: 10.1109/TMC.2020.2973616.

[38] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K.-K.-R. Choo, "Imaging and fusing time series for wearable sensor-based human activity recognition," *Inf. Fusion*, vol. 53, pp. 80–87, Jan. 2020.

[39] S. Ashry, R. Elbasiony, and W. Gomaa, "An LSTM-based descriptor for human activities recognition using IMU sensors," in *Proc. 15th Int. Conf. Informat. Control, Autom. Robot.*, vol. 1, 2018, pp. 494–501.

[40] S. Badar ud din Tahir, A. Jalal, and M. Batool, "Wearable sensors for activity analysis using SMO-based random forest over smart home and sports datasets," in *Proc. 3rd Int. Conf. Advancements Comput. Sci. (ICACS)*, Feb. 2020, pp. 1–6.

[41] S. P. Singh, A. Lay-Ekuakille, D. Gangwar, M. Kumar Sharma, and S. Gupta, "Deep ConvLSTM with self-attention for human activity decoding using wearables," 2020, *arXiv:2005.00698*. [Online]. Available: http://arxiv.org/abs/2005.00698

Riktim Mondal received the B.E. degree from the Department of Computer Science and Engineering, Jadavpur University. His research interests include computer vision, pattern recognition, natural language processing, and human–computer interface.

Debadyuti Mukherjee is currently pursuing the bachelor's degree with the Department of Computer Science and Engineering, Jadavpur University. His research interests include computer vision, pattern recognition, and machine learning.

Pawan Kumar Singh (Member, IEEE) received the B.Tech. degree in information technology from the West Bengal University of Technology in 2010 and the M.Tech. degree in computer science and engineering and the Ph.D. (Engineering) degree from Jadavpur University (JU) in 2013 and 2018, respectively. He is currently working as an Assistant Professor with the Department of Information Technology, JU. He has published more than 50 research papers in peer-reviewed journals and international conferences. His current research interests include computer vision, pattern recognition, handwritten document analysis, image and video processing, machine learning, and artificial intelligence. He is a member of The Institution of Engineers, India, and the Association for Computing Machinery (ACM); and a Life Member of the Indian Society for Technical Education (ISTE), New Delhi, and the Computer Society of India (CSI). He also received the RUSA 2.0 Fellowship for pursuing his postdoctoral research in JU in 2019.

Vikrant Bhateja (Senior Member, IEEE) is working as an Associate Professor with the Department of ECE, SRMGPC, Lucknow. His research interests include digital image and video processing, computer vision, medical imaging, machine learning, and pattern analysis and recognition. He has around 155 quality publications in various international journals and conference proceedings. He is an Associate Editor of IJSE and IJACI. He has edited more than 22 volumes of conference proceedings with Springer Nature. He is the EiC of the IJNCR journal (IGI Global).

Ram Sarkar (Senior Member, IEEE) received the B.Tech. degree in computer science and engineering from the University of Calcutta in 2003 and the M.E. degree in computer science and engineering and the Ph.D. (Engineering) degree from Jadavpur University in 2005 and 2012, respectively. He joined the Department of Computer Science and Engineering, Jadavpur University, as an Assistant Professor, in 2008, where he is currently working as an Associate Professor. His current research interests include image processing, pattern recognition, and machine learning. He received the Fulbright-Nehru Fellowship (USIEF) for his postdoctoral research in the University of Maryland, College Park, USA, from 2014 to 2015.