

HAR-GCNN: Deep Graph CNNs for Human Activity Recognition From Highly Unlabeled Mobile Sensor Data

Abduallah Mohamed*, Fernando Lejarza†, Stephanie Cahail‡, Christian Claudel§, Edison Thomaz¶

The University of Texas at Austin

{*abduallah.mohamed, †lejarza, ‡stephaniecahail, §christian.claudel, ¶ethomaz}@utexas.edu

Abstract—The problem of human activity recognition from mobile sensor data applies to multiple domains, such as health monitoring, personal fitness, daily life logging, and senior care. A critical challenge for training human activity recognition models is data quality. Acquiring balanced datasets containing accurate activity labels requires humans to correctly annotate and potentially interfere with the subjects’ normal activities in real-time. Despite the likelihood of incorrect annotation or lack thereof, there is often an inherent chronology to human behavior. For example, we take a shower after we exercise. This implicit chronology can be used to learn unknown labels and classify future activities. In this work, we propose HAR-GCCN, a deep graph CNN model that leverages the correlation between chronologically adjacent sensor measurements to predict the correct labels for unclassified activities that have at least one activity label. We propose a new training strategy enforcing that the model predicts the missing activity labels by leveraging the known ones. HAR-GCCN shows superior performance relative to previously used baseline methods, improving classification accuracy by about 25% and up to 68% on different datasets. Code is available at <https://github.com/abduallahmohamed/HAR-GCNN>.

Index Terms—deep graph convolutional neural network, human activity recognition, deep learning

I. INTRODUCTION

Human Activity Recognition (HAR) has been an active research field recently, covering a wide range of applications in health monitoring and fitness [1], [2]. Technological advances regarding inertial sensors with longer battery life spans and improved computing capabilities have enabled gathering larger volumes of continuous data to be used for HAR [3]. Despite these recent advances, “ground truth annotation”, which involves labeling sensor readings, remains a critical challenge [3]. Such annotation tasks are typically performed manually and occur either in real-time or *post hoc* once the activity has been completed. In the HAR studies, such annotation tasks can be expensive, onerous, prone to human error, and can even condition the user and interfere in the activity itself [4]. Thus, the training data likely contains a significant amount of unreliable and missing labels. These incorrect or missing labels create gaps in the data which have a detrimental impact on model development and training. Thus, a paramount aspect of models for HAR is their ability to learn in the presence of missing labels, while achieving high accuracy in classifying human activity. A variety of deep and

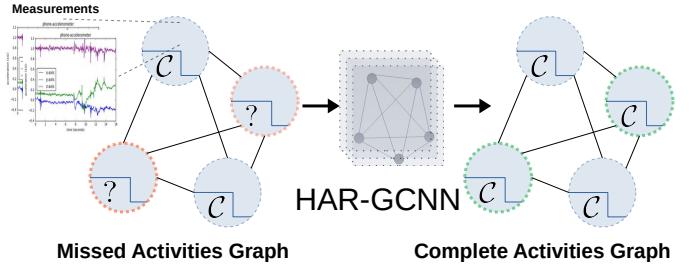


Fig. 1. The input to HAR-GCNN is a graph comprised of partially labelled sensor measurements that are chronologically ordered activities. The model predicts the classes C of the unlabeled activities. Each input node contains a period of sensor measurements alongside a label or multi-labels. The output is the activity class(es) for each node.

machine learning methods have been previously proposed for single and multi-label classification of human activity [5]–[7]. The missing label(s) can be predicted based on the readings of multiple sensors such as accelerometer, gyroscope, location, etc [6]. However, fewer approaches leverage the context of “neighboring” activities for predicting a missing label which has been shown to improve HAR [7], [8].

We hypothesize that human activities follow a chronological correlation which can provide informational context to improve HAR. The benefit of this hypothesis, if valid, is that one can leverage known or correctly labeled activities to predict the surrounding ones. In order to investigate the validity this assumption, we formulated the HAR problem comprising a sequence of chronologically ordered activities, some of which were correctly labeled while others were not labeled at all. We used deep learning as a data-driven approach to discover this chronological correlation. We first employed Recurrent Neural Nets (RNNs) which is the most straightforward deep model used to handle time series data. Nonetheless, our literature survey revealed that Convolutional Neural Networks (CNNs) can often be more powerful than RNNs in performing HAR [9]. Yet, both approaches CNNs and RNNs are structurally not geared towards directly leveraging the correlation between the sequential activities. For example, RNNs treat each time step separately by fusing it into the “neural memory”, which is the only component that partially correlates the data. On a similar note, CNNs only make use of neighbouring activity information based on the chosen kernel

size. Based on the shortcomings of these architectures, we posited that graph-based structures more adequately capture the aforementioned features of the problem at hand. Graph representations in HAR allow for modeling each activity as a node, and the graph edges can directly model the relationship between these activities. A suitable tool for learning such graphs is deep Graph CNNs (GCNNs) [10]. GCNNs behave as CNNs but weight the nodes based on the value of the edges to exploit the correlation between sequential activities.

In this work we show that the chronological correlation indeed exists by experimenting with two commonly employed HAR datasets. One was collected in the wild, while the other was collected in a scripted manner. Our results show that the proposed models benefit from this correlation and use it to predict the neighbouring missing activities, improving performance relative to RNNs and CNNs benchmarks. In the following sections, we discuss prior works, as well as the datasets utilized for training and validation. We provide details of our formulation including the graph structure used to model human activities. Then, we introduce our proposed HAR-GCNN, highlighting details of its architecture and implementation, and define the comparison baselines used to benchmark our approach. In the experiments section, we evaluate our model's performance relative to other deep neural network architectures employing widely used HAR datasets.

II. RELATED WORK

A large number of traditional supervised machine learning techniques have been developed in the literature for HAR from mobile sensor data. Such models include, for example, logistic regression, k -nearest neighbors, decision trees, and multi-layer perceptrons (MLP) [5], [11], [12]. These models typically exhibit good performance when trained on controlled datasets that are fully and accurately labeled, which may not be the case for in-the-wild environments [13]. Furthermore, these models often require substantial feature extraction which can be time-consuming and rely heavily on domain knowledge [6].

Addressing these shortcomings, a variety of prior works have explored unsupervised and semi-supervised learning techniques on human activity data. For example, [14] developed a network of convolutional autoencoders on an unlabeled dataset to extract useful features, which were then used to complement their supervised learning mechanism. [14] proposed Deep Auto-Set, a deep learning classification model trained on raw multi-modal sensory segments. Furthermore, [2] implemented semi-supervised, active learning techniques operated both online and offline. [2] demonstrated superior performance relative to fully supervised approaches on HAR datasets where ground truth labels are scarce. More specifically, [2] evaluated the performance of different pool-based and stream-based active learning frameworks on various datasets relevant to human activity [15] and contrasted classification accuracy against classical machine learning approaches.

Besides these models, a great number of deep learning frameworks have been proposed for HAR [6]. For example, different CNN architectures [16]–[18] resulted in significantly

higher accuracy than prior machine learning techniques applied on HAR. Further, [19] explored normalization and sensor data fusion using CNNs and demonstrated that these techniques can further improve the performance of deep learning models. [7] investigated deep convolutional recurrent models including deep feed-forward neural networks (DNN), CNNs, and different variants of Long Short-Term Memory (LSTM) cells. Their results show that recurrent networks significantly outperform convolutional networks on activities that are short in duration but follow a natural chronology. Such performance gains are likely because recurrent models can contextualize to improve recognition. [20] reported improved HAR using ensembles of deep LSTM learners relative to previously reported recurrent neural networks. [21] developed a deep RNN performing extensive hyper-parameters optimization and showed superior performance compared to other traditional machine learning models.

Several variants of graph-based models have been recently proposed leveraging spatial and temporal properties in the data for collective HAR (i.e., predicting the activity of an entire group as opposed to a single subject). Numerous of these models have been reported for computer vision applications to (deeply) learn the interactions between individual participants for a given sequence of video clips and to predict the overall activity or outcome of a group of people [22]–[26]. Similarly, [27] proposed a context-aware, semi-supervised graph propagation algorithm with a Support Vector Machine (SVM) classifier to address individual HAR. The results in [27] suggest that exploiting contextual data from neighboring activities (i.e., nodes) can result in improved performance even relative to fully supervised approaches, likely by discriminating outliers and erroneously labeled data.

GCNNs have recently gained significant popularity for various applications of semi-supervised learning [10], [28]. GCNNs leverage dependencies between the features and labels of nodes in a given graph resulting in improved predictive performance, particularly when a significant number of the training labels are missing. Considering these desirable characteristics, as well as the aforementioned advantages of models that can contextualize, GCNNs are an appealing modeling strategy that to the best of our knowledge has not been previously reported for HAR. In light of this, the main contributions of this work are:

- A formulation for HAR exploiting the chronological context of the activities embedded within a graph structure.
- A novel mechanism that trains HAR-GCNN to learn to predict the missing labels in the input graph with high accuracy.
- Extensive computational experiments evaluating the effect of the percentage of missing labels, as well as the effect of the number of activities on the prediction accuracy. Experiments performed on Extra-Sensory [15] and PAMAP [29] datasets, which are among the most commonly used in the HAR literature.
- Benchmarks of our proposed approach against previously reported deep architectures including CNN and LSTM

models.

III. PROBLEM FORMULATION & DATASETS

Given a set of time series sensory measurements $\mathcal{F} = \{f_t | t \in T\}$ sampled over a window of time steps T , it is of interest to learn the classes of activities $\mathcal{C} = \{c_t | t \in T\}$ associated with each of such measurements. The measurements are collected from a variety of sensors such as accelerometers, gyroscopes, etc. Each interval of measurements is associated with multiple labels corresponding to different activities (i.e., multi-label classification). Our problem formulation uses a set of prior and posterior measurements with known activities to predict the multi-label or single label classes of those unknown activities. In other words, using $(f_{t-m}, c_{t-m}), \dots, (f_{t-1}, c_{t-1}), (f_{t+1}, c_{t+1}), \dots, (f_{t+m}, c_{t+m})$ and f_t to predict c_t , where m is the number of neighbour activities to be used. In this way, we consider a sequence of activities represented by sensor measurements over a time horizon including the associated labels, whether they are known or not, and predict the class of the unknown labels exploiting the observed sequential order. One important remark is that the datasets under consideration are recorded *in-the-wild*. That is, an *in-the-wild* dataset contains sensor readings from the users who were not previously instructed to perform a given set of activities (which is typically the case in more controlled environments). *In-the-wild* settings in turn reduce the bias within the collected data and results in a more natural chronology of activities from which our proposed model can learn. The Extra-Sensory [15] dataset is an example of such *in-the-wild* measurements. The dataset contains over 300,000 minutes of labeled sensor recordings from smartphones and smartwatches worn by a total of 60 study participants. The behavioural activities in this dataset can be categorized by 51 non-exclusive labels such as sitting, sleeping, strolling, cooking, etc. For this type of the data the HAR task at hand is a *multi-label classification problem*. Further, the dataset contains 224 different raw features obtained from mobile sensor readings recorded over 20 seconds window every one minute. To further validate the advantages of the framework proposed herein, we also conduct experiments using the PAMAP dataset [29]. The PAMAP dataset while collected in a more controlled environment provides a useful instance to validate our framework, particularly when labels are artificially hidden during training. The PAMAP dataset consists of data from 9 subjects, wearing 3 inertial measurement units and a heart rate monitor and performing 12 exclusive activities (i.e., the HAR task is a single-label classification problem) with 52 raw features per instance. The PAMAP dataset is being used herein as an example of a scripted dataset, which likely means that the chronological order of the recorded activities indeed contains some amount of bias and which we further emphasize in the numerical experiments section. **Training/Testing details:** Both datasets were split into training and test sets with a ratio of 2:1. The sequence of the recorded

activities was kept and no randomization was used, which is crucial for learning from the natural sequential order of activities. The Extra-Sensory dataset serves to show that the activities can be predicted with higher accuracy by exploiting such implicit order. Conversely, the PAMAP dataset is employed to show that if the sequence of activities is scripted *a priori*, HAR-GCNN results in almost perfect classification performance as it can quickly learn the underlying script. We note that no test data was used to train the proposed model, and the data sets were kept separate to prevent any data leaking.

IV. HAR-GCNN METHOD DESCRIPTION

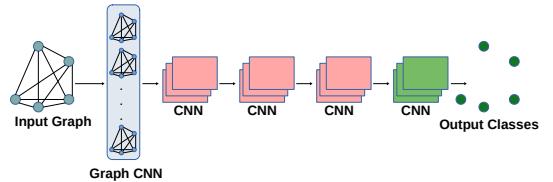


Fig. 2. HAR-GCNN model description. The CNN is just a single layer convolution with an activation function.

Constructing the activity graph: We model the set of activities as a graph, which is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \{v | v \in \mathcal{V}\}$ is a set of vertices. Each vertex $v = [f_t, c_t]$ contains both measurements over a time window and the associated multi-label class, with the exception of the nodes whose activity label is missing and we thus want to predict, for which c is set to zero. The graph edges are fully connected, $\mathcal{E} = \{e_{ij} | e \in \mathcal{E}; i, j \in |\mathcal{V}|\}$, where \mathcal{E} is the set of all the graph edges. The weight of these edges is set to one, so that predicting the activity of all nodes is equally important. Our proposed model allows using any number of nodes during training. For example, the graph can consist of three nodes with one of them missing its label. In the experiments section we study the effect of the number of input labelled neighbour nodes on predicting a missing activity class. Further, we note that framework allows for predicting more than one missing label, which we explore in our numerical experiments.

Model Description: First, the **GCNN layer** takes as input the aforementioned activity graph \mathcal{G} . The GCNN model has a layer-wise structure defined as $\text{GCNN}(\mathcal{V}^{(l)}, A) = \sigma(A_{\text{norm}} \mathcal{V}^{(l)} W^{(l)})$ Where A is the adjacency matrix that defines the edges of the activity graph and $\sigma(\cdot)$ denotes an activation function. The GCNN operates as an ordinary CNN except that it weights the kernel by the value of the normalized adjacency matrix A_{norm} defined as:

$$A_{\text{norm}} = I - \hat{D}^{-\frac{1}{2}}(A + I)\hat{D}^{-\frac{1}{2}}$$

where \hat{D} is the degree node matrix of $A + I$, and I is the identity matrix. This normalization approach was introduced by [9]. More in-depth information about graph CNNs can be found in [10]. The output of the GCNN is a graph embedding that represents the whole information of the sensor measurements and their corresponding known labels.

The second step in HAR-GCNN is the **CNN output layers**, for which we use a sequence of CNNs (single layer convolutions with an activation function). This architecture was chosen mainly because performance can deteriorate when the depth of the GCNNs increases [30]. That is, it would be inefficient to use a sequence of graph CNNs to predict each activity label. The output of these CNN layers is directly regressed against the proper loss function (Binary Cross Entropy or Cross-Entropy) to predict the labels \mathcal{C} . Figure 2 describes the distinct steps included in our architecture.

Implementation¹ Details: We employ the same GCNN implementation as the one in the pioneering work of [10]. The chosen implementation is easier to interpret and deploy, relative to similar architectures available in PyTorch. The model consists of a single GCNN layer that outputs a graph embedding. Then, three CNN layers are used to process the embedding and a final CNN layer is included with a sigmoid (softmax for PAMAP) activation function. We used PReLU [31] activation functions for intermediate layers. The total model parameters size is $\sim 15k$ for the Extra-Sensory dataset and $\sim 5k$ for the PAMAP dataset.

V. BASELINE METHODS

In order to benchmark the performance of our approach, we compared it to the aforementioned, more commonly encountered deep architectures. Mainly, we contrast HAR-GCNN to CNNs [16]–[18] and LSTMs [7], [32] that were previously used for HAR. **CNN baseline** The CNN baseline, shown in Figure 3, is a sequence of five CNN layers that have the same depth as HAR-GCNN model. The reason for choosing this architecture is that \mathcal{V} represents a set of an arbitrary number of activities, and a CNN can learn a kernel that is agnostic to the width and height of the input. The total parameters size is also $\sim 15k$ for the Extra-Sensory and $\sim 5k$ for the PAMAP datasets, thus allowing for a fair comparison with HAR-GCNN. The input to this base model is the \mathcal{V} itself, which can be interpreted as an image of width equal to the sensor readings, and of height equal to the number of nodes or activities in the graph. All layers used PReLU [31] activation function except the last layer with a sigmoid (softmax for PAMAP) function. The intermediate three layers use a residual connection to improve performance during training. **LSTM baseline** The inputs to the LSTM, shown in Figure 3, are the graph nodes treated as a sequence of time steps. The model parameter size is also $\sim 15k$ and $\sim 5k$ for both Extra-Sensory and PAMAP respectively, same as HAR-GCNN model and CNN base model. The 1D-CNN layers use PReLU activation function except for the last layer which uses a sigmoid (softmax for PAMAP) activation function.

VI. TRAINING MECHANISM

Besides keeping the number of parameters consistent across models, we trained all models using the same settings. We used Binary Cross Entropy (BCE) as the loss function in the multilabel case of the Extra-Sensory dataset, which defined as:

¹Code: <https://github.com/abduallahmohamed/HAR-GCNN.git>

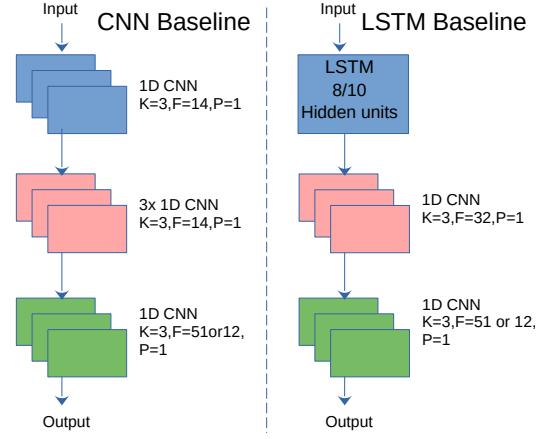


Fig. 3. CNN and LSTM baselines. Both baselines share the same depth and the same number of parameters as in our model. K is the kernel size, P is the padding size and F is the features. The output layer changes its features based on the number of predicted classes. All CNNs are single layer convolution followed by an activation function.

$\text{BCE loss}_j = -[\text{class}_j \cdot \log x_j + (1 - \text{class}_j) \cdot \log(1 - x_j)]$ For the single-label PAMAP dataset we used the Cross-Entropy(CE) loss, which is defined as: $\text{CE loss}(x, \text{class}) = -\log\left(\frac{\exp(x[\text{class}])}{\sum_j \exp(x[j])}\right)$ where j is the class number and x is the predicted class. We artificially imposed missing labels in both datasets to simulate the case of missing labels. Our training framework randomly hides a percentage of the known activity labels \mathcal{C} within the input graph \mathcal{G} , which serve as the missing labels we want to predict. Moreover to make the model generalize better, we add random noise (Gaussian with 0 mean and 1 standard deviation) to disturb the measurements vector \mathcal{F} , which helps generalize the performance of HAR-GCNN for unseen data while also accounting for possible measurement errors in the raw sensor data. In all of our experiments, we used a 50% probability for either hiding a label or adding noise to the measurements. Further, we restricted the percentage of hidden nodes to be no greater than 66%, so that we are guaranteed that at least 33% of the nodes have their original labels. The disturbed data were generated once and were used in all of our experiments across different models for consistency. All of our results are reported on the withheld test set. We report the mean of three repetitions of our experiments, which were conducted with a controlled random seed to ensure consistent settings. Figure 4 shows the data-flow and the dimension of the data alongside the training mechanism.

VII. EXPERIMENTS & DISCUSSION

We considered two settings for the amount of missing class labels \mathcal{C} : 33% and 66% in the input graph \mathcal{G} . For example, the case of 66% missing labels corresponds to $\frac{2}{3}$ of the measurements \mathcal{F} being known but without labels \mathcal{C} , which we seek to predict. This method evaluates the model's ability to generalize to unseen instances because the models were trained under different settings as described in Section VI.

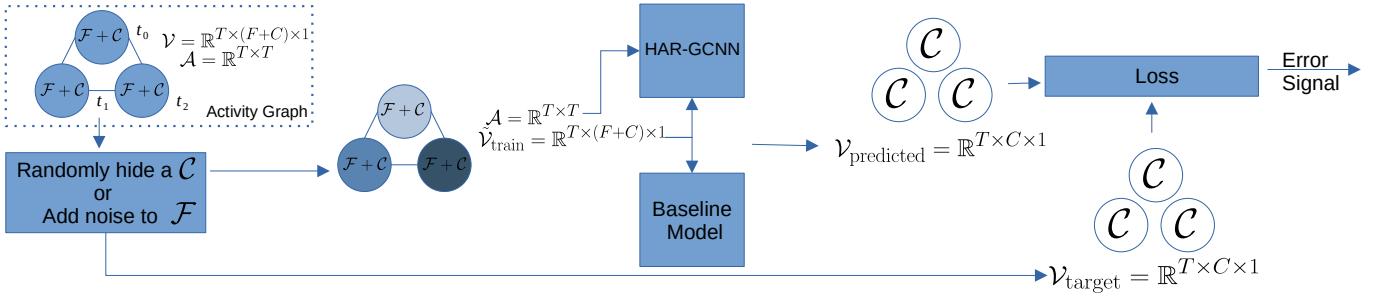


Fig. 4. HAR-GCNN data flow diagram for model training as described in VI. \mathcal{F} represents the sensor measurements with dimension $F = 224, 52$ for the Extra-Sensory dataset and PAMAP dataset, respectively. \mathcal{C} is the set of multi-label activity classes with dimensions $C = 51, 12$ for the Extra-Sensory dataset and PAMAP dataset, respectively. T is the number of time steps. \mathcal{A} stands for adjacency matrix, \mathcal{V} stands for the graph vertices.

TABLE I

F-1 SCORE OVER ACCURACY OF HAR-GCNN VERSUS BASELINE MODELS (THE HIGHER THE BETTER). THE EXTRA-SENSORY RESULTS ARE $F_1^{\text{macro}} / \text{ACC}_{\text{mean}}$, AND PAMAP RESULTS ARE F_1 / ACC . 3, 5, 10 AND 25 CHRONOLOGICALLY ORDERED ACTIVITIES WERE USED OVER A RANGE OF MISSING LABELS ON ACCURACY OVER F-1 SCORE. THE AVERAGE OF THREE RUNS IS REPORTED, AND BOLDED NUMBERS SHOW THE BEST.

# of Activities	% missing labels	Extra-Sensory Dataset			PAMAP Dataset		
		HAR-GCNN	CNN	LSTM	HAR-GCNN	CNN	LSTM
3	33%	0.781 / 99.52	0.621 / 99.23	0.464 / 98.43	0.999 / 99.92	0.975 / 97.52	0.973 / 97.26
	66%	0.792 / 99.52	0.531 / 98.69	0.377 / 97.95	0.999 / 99.94	0.902 / 90.22	0.903 / 90.26
5	33%	0.814 / 99.41	0.715 / 99.41	0.504 / 98.67	0.998 / 99.76	1.000 / 99.96	0.990 / 99.05
	66%	0.880 / 99.58	0.659 / 99.25	0.459 / 98.37	1.000 / 99.98	0.996 / 99.57	0.950 / 95.05
10	33%	0.813 / 99.14	0.744 / 99.46	0.522 / 98.64	0.999 / 99.91	0.998 / 99.81	0.997 / 99.65
	66%	0.868 / 99.52	0.691 / 99.28	0.483 / 98.37	1.000 / 99.98	0.997 / 99.75	0.969 / 96.87
25	30%	0.748 / 98.76	0.784 / 99.50	0.532 / 98.73	0.998 / 99.79	1.000 / 99.99	0.998 / 99.75
	66%	0.838 / 99.41	0.724 / 99.23	0.494 / 98.46	1.000 / 99.99	0.997 / 99.68	0.981 / 98.10

For the Extra-Sensory dataset, we used the macro F_1 score which is defined as: $F_1^{\text{macro}} = \frac{1}{N} \sum_{n \in N} F_1^n$ where N is the total number of classes. The mean accuracy is defined as: $\text{Acc}_{\text{mean}} = \frac{1}{N} \sum_{n \in N} \text{Acc}_n$. For the PAMAP dataset we use the ordinary F1 and accuracy scores.

A. Performance of HAR-GCNN Against Baselines

Table I shows the main results comparing the performance of HAR-GCNN versus the baseline models. We explain the main results by focusing on the three activities case (i.e., first row of Table I) but note that the observations generalize for all other instances in Table I. In the three activities case, 33% missing labels imply that only one out of the three activities does not have a label. An important result is that HAR-GCNN, in the case of 33% or 66% missing labels, outperforms both CNN and LSTM baselines. The performance improvement is about 25% and 68% relative to CNN and LSTM models respectively for the Extra-Sensory dataset. For the PAMAP dataset HAR-GCNN has a $\sim 2\%$ higher F-1 score than both CNN and LSTM, and almost reaching the upper performance bound of 1.00. We posit that the observed performance improvements stem from the proposed graph formulation, in which the graph edges can more deeply capture the relationship between labels and their features. Therefore, if at least one activity is correctly classified, then the remaining nodes will have a higher likelihood of being correctly labeled as well. We note that the CNN baseline significantly outperforms the LSTM model across the two datasets considered. CNN-based approaches appear to be more

suitable to this type of problem, as they do not accumulate errors sequentially while classifying the measurements and have a more global view of the entire state. Unlike in recurrent nature of methods like LSTM, the error in previous predictions propagates to the future ones. These findings are in line with prior works [9].

B. Effect of The Cardinality of The Graph Nodes

We also explored the effect of the number of chronologically ordered activities on the classification accuracy. From Table I it is noticeable that the results discussed in the previous section are consistent for an increasing number of activities in the input graph. From the results obtained using the Extra-Sensory dataset, the performance of all base models improves as more graph nodes are considered. The performance of HAR-GCNN is still consistently better than the base models and does not improve substantially when the number of activities is increased from 5 to 25. These results imply that HAR-GCNN learns a proper representation that has a constant performance irrelevant from the number of activities. The results obtained using the PAMAP dataset in Table I show that the performance across all models (HAR-GCNN and baseline models) does not change significantly with the graph cardinality. This constant performance can be attributed to the PAMAP dataset, which was collected from users being asked to perform tasks in a pre-specified manner. Models trained on this dataset might readily infer such pre-specified sequence of activities and thus resulting in very high F-1 score and accuracy.

VIII. CONCLUSION

In this paper, we presented HAR-GCNN, a deep graph model to predict missing activity labels leveraging the context of chronological sequences of these activities. We proposed an approach for modelling the activities as nodes in a fully connected graph. We introduced a new training mechanism that enables the model to learn the implied natural chronological order of human activities. To gain a quantitative understanding of our proposed strategy, we benchmarked our design against baseline deep models sharing the similar structural properties as HAR-GCNN. Our results indicate that HAR-GCNN has superior performance in terms of F-1 score and accuracy compared to the chosen baselines. Further, our results suggest that learning a chronology of activities to predict the missing label is likely a key driver for such performance improvements. Our experimental results reveal that HAR-GCNN has a stable performance, independently from the number of chronologically ordered activities considered within the input graph.

REFERENCES

- [1] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, “Convolutional neural networks for human activity recognition using mobile sensors,” in *6th International Conference on Mobile Computing, Applications and Services*. IEEE, 2014, pp. 197–205. 1
- [2] R. Adaimi and E. Thomaz, “Leveraging active learning and conditional mutual information to minimize data annotation in human activity recognition,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, Sep. 2019. [Online]. Available: <https://doi-org.ezproxy.lib.utexas.edu/10.1145/3351228> 1, 2
- [3] A. Bulling, U. Blanke, and B. Schiele, “A tutorial on human activity recognition using body-worn inertial sensors,” *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, pp. 1–33, 2014. 1
- [4] H. Kwon, G. D. Abowd, and T. Plötz, “Handling annotation uncertainty in human activity recognition,” in *Proceedings of the 23rd International Symposium on Wearable Computers*, 2019, pp. 109–117. 1
- [5] J. Mantičjarvi, J. Himberg, and T. Seppanen, “Recognizing human motion with multiple acceleration sensors,” in *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236)*, vol. 2. IEEE, 2001, pp. 747–752. 1, 2
- [6] 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. 1, 2
- [7] N. Y. Hammerla, S. Halloran, and T. Plötz, “Deep, convolutional, and recurrent models for human activity recognition using wearables,” *arXiv preprint arXiv:1604.08880*, 2016. 1, 2, 4
- [8] L. Cao, Y. Wang, B. Zhang, Q. Jin, and A. V. Vasilakos, “Gchar: An efficient group-based context-aware human activity recognition on smartphone,” *Journal of Parallel and Distributed Computing*, vol. 118, pp. 67–80, 2018. 1
- [9] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, “Social-stgcn: A social spatio-temporal graph convolutional neural network for human trajectory prediction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14424–14432. 1, 3, 5
- [10] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016. 2, 3, 4
- [11] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011. 2
- [12] S. Pirttikangas, K. Fujinami, and T. Nakajima, “Feature selection and activity recognition from wearable sensors,” in *International symposium on ubiquitous computing systems*. Springer, 2006, pp. 516–527. 2
- [13] Y. Vaizman, N. Weibel, and G. Lanckriet, “Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–22, 2018. 2
- [14] A. A. Varamin, E. Abbasnejad, Q. Shi, D. C. Ranasinghe, and H. Rezatofighi, “Deep auto-set: A deep auto-encoder-set network for activity recognition using wearables,” in *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, ser. MobiQuitous ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 246–253. [Online]. Available: <https://doi.org/10.1145/3286978.3287024> 2
- [15] Y. Vaizman, K. Ellis, and G. Lanckriet, “Recognizing detailed human context in the wild from smartphones and smartwatches,” *IEEE Pervasive Computing*, vol. 16, no. 4, pp. 62–74, 2017. 2, 3
- [16] R. Grzeszick, J. M. Lenk, F. M. Rueda, G. A. Fink, S. Feldhorst, and M. ten Hompel, “Deep neural network based human activity recognition for the order picking process,” in *Proceedings of the 4th international Workshop on Sensor-based Activity Recognition and Interaction*, 2017, pp. 1–6. 2, 4
- [17] F. Moya Rueda, R. Grzeszick, G. A. Fink, S. Feldhorst, and M. Ten Hompel, “Convolutional neural networks for human activity recognition using body-worn sensors,” in *Informatics*, vol. 5, no. 2. Multidisciplinary Digital Publishing Institute, 2018, p. 26. 2, 4
- [18] F. Cruciani, A. Vafeiadis, C. Nugent, I. Cleland, P. McCullagh, K. Votis, D. Giakoumis, D. Tzovaras, L. Chen, and R. Hamzaoui, “Feature learning for human activity recognition using convolutional neural networks,” *CCF Transactions on Pervasive Computing and Interaction*, vol. 2, no. 1, pp. 18–32, 2020. 2, 4
- [19] S. Münzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, and R. Dürichen, “Cnn-based sensor fusion techniques for multimodal human activity recognition,” in *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017, pp. 158–165. 2
- [20] Y. Guan and T. Plötz, “Ensembles of deep lstm learners for activity recognition using wearables,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 1–28, 2017. 2
- [21] M. Inoue, S. Inoue, and T. Nishida, “Deep recurrent neural network for mobile human activity recognition with high throughput,” *Artificial Life and Robotics*, vol. 23, no. 2, pp. 173–185, 2018. 2
- [22] L. Lu, Y. Lu, R. Yu, H. Di, L. Zhang, and S. Wang, “Gaim: Graph attention interaction model for collective activity recognition,” *IEEE Transactions on Multimedia*, vol. 22, no. 2, pp. 524–539, 2019. 2
- [23] J. Zhang, F. Shen, X. Xu, and H. T. Shen, “Temporal reasoning graph for activity recognition,” *IEEE Transactions on Image Processing*, vol. 29, pp. 5491–5506, 2020. 2
- [24] J. Wu, L. Wang, L. Wang, J. Guo, and G. Wu, “Learning actor relation graphs for group activity recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9964–9974. 2
- [25] D. Singh and C. K. Mohan, “Graph formulation of video activities for abnormal activity recognition,” *Pattern Recognition*, vol. 65, pp. 265–272, 2017. 2
- [26] J. Tang, X. Shu, R. Yan, and L. Zhang, “Coherence constrained graph lstm for group activity recognition,” *IEEE transactions on pattern analysis and machine intelligence*, 2019. 2
- [27] M. Stikic, D. Larlus, and B. Schiele, “Multi-graph based semi-supervised learning for activity recognition,” in *2009 international symposium on wearable computers*. IEEE, 2009, pp. 85–92. 2
- [28] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, “Graph neural networks: A review of methods and applications,” *arXiv preprint arXiv:1812.08434*, 2018. 2
- [29] A. Reiss and D. Stricker, “Introducing a new benchmarked dataset for activity monitoring,” in *2012 16th International Symposium on Wearable Computers*. IEEE, 2012, pp. 108–109. 2, 3
- [30] G. Li, M. Müller, A. Thabet, and B. Ghanem, “Deepgcn: Can gcn go as deep as cnns?” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 9266–9275. 4
- [31] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” *CoRR*, vol. abs/1502.01852, 2015. [Online]. Available: <http://arxiv.org/abs/1502.01852> 4
- [32] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, “Convolutional, long short-term memory, fully connected deep neural networks,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 4580–4584. 4