JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Semi-Supervised Learning Framework for Activity Recognition Based on Graph Structures and Label Propagation

[1]Umar Alhassan Umar, [2]Isyaku Ado Ishaq, [3]Abubakar Mohammed Goni, [4]Abdullahi Ibrahim Kurfi

[1234]Department of Computer Science, School of Technology Kano State Polytechnic

[1]Department of Computer Science,

[1]School of Technology Kano State Polytechnic, Kano State, Nigeria,

Abstract: Human Activity Recognition (HAR) has gathered significant research attention within the pervasive computing community. The sensor data collected from human activity monitoring plays a crucial role in various motion analysis tasks and supports numerous applications. However, most current HAR methods rely on supervised learning techniques, which demand a substantial amount of labeled data to achieve effective performance. Data labeling is often tedious, costly, and time-intensive, as it requires scientific experiments or input from skilled human annotators to label previously unlabeled data. Moreover, the annotation process is susceptible to errors and can be unreliable. Developing algorithms that effectively utilize a small amount of labeled data alongside a large volume of unlabeled data is a promising approach to mitigate these challenges. Building on the recent success of semi-supervised learning (SSL) in sensor-based activity recognition, this study leverages SSL to develop a model for human activity recognition that addresses the aforementioned challenges. Specifically, the goal is to design a graph-based semi-supervised activity recognition framework utilizing label propagation (LP). The model's performance was compared with three machine learning (ML) approaches, with the proposed method achieving the highest classification accuracy.

Index Terms - Human Activity Recognition , semi-supervised learning , machine learning , label propagation.

1. INTRODUCTION

Semi-supervised learning is a subset of machine learning that integrates both supervised and unsupervised approaches, utilizing labeled and unlabeled data to train AI models for classification and regression purposes. Although semi-supervised learning is typically applied to use cases similar to those of supervised learning, it is distinguished by techniques that incorporate unlabeled data alongside labeled data during model training, enhancing the learning process beyond traditional supervised methods. Semi-supervised learning methods are particularly valuable in scenarios where acquiring enough labeled data is costly or challenging, while abundant unlabeled data is readily available. In these cases, neither fully supervised nor unsupervised learning approaches offer satisfactory solutions. Human activity recognition (HAR) is a important and burning research area of vesture computing due to its potential to enable novel context-aware applications for elderly care, education, sports, and entertainment [1]. There are mainly two types of HAR: video-based HAR and sensor-based HAR [2]. Video-based HAR examines videos or images containing human motions from the camera, while sensor-based HAR focuses on the motion data from smart sensors such as an accelerometer, gyroscope, Bluetooth, sound sensors, and so on. Due to the thriving development of sensor technology and pervasive computing, sensor-based HAR is becoming more popular and widely used with privacy well-protected [2]. Currently, most smart devices (e.g. smartphones and smartwatches) are embedded with sensors that allow to record different types of data relating to daily living activities [3]. Their advantages in high computational power, small size, and low cost, allow people to interact with these devices as part of their daily living [4, 5]. Obtaining accurate information from human activity-based wearable sensors is significant in pervasive computing [4].However, pervasive sensing became an active research area with the main purpose of extracting knowledge from the data acquired by pervasive sensors [4]. The raw data generated by the sensors are highly fluctuating and oscillatory, which makes it difficult to recognize the underlying patterns using their raw values [4]. For this purpose, HAR systems make use of machine learning techniques, which are useful for building models to describe, analyze, and predict data [4]. The advancement of wearable sensing technology has motivated researchers to develop models for human activity recognition [3]. The recognition of human activities in particular, has become a task of high interest within the field of pervasive computing, especially for medical, military, and security applications [4].

The development of human activity recognition (HAR) algorithms involves the collection of a large amount of dataset which should be annotated by a team or an expert [6]. Machine learning techniques are widely used to recognize activities [4-7]. Most of the activity recognition methods are based on supervised learning techniques which require large amounts of labeled training data [8-9] Supervised algorithms achieve high recognition performance but, they require significant amounts of labeled training data [8]. It is easy to obtain a large number of training examples but difficult, time-consuming, and expensive to gather the corresponding class

labels [6]. In addition, the annotation process is error-prone [6-10] and sometimes even unreliable [11]. Also, obtaining sufficient, accurate, and detailed labels of activities is challenging [8], [12]. These labeling issues prevent the applicability of SL approaches in real-world settings [8]. The main challenge is to design a machine learning model with low annotation effort and maintain adequate performance [6]. Semi-supervised learning is a hot topic aiming to address this issue [7].

Previous research works on activity recognition mostly applied supervised learning techniques [8- 4]. This is because; it might be very hard to discriminate activities in a completely unsupervised context [4], [5]. Supervised learning approaches require labeled sensor data for training, with the labels identifying the activity as well as the start and end times of each activity [13-9]. This is practically very tedious, expensive, time-consuming, unreliable [11] and error-prone [6]. To deal with the limitations of supervised and unsupervised approaches, we build a Graph-based semi-supervised label propagation learning model.

However, the challenge still resides in the lack of sufficient, high-quality activity annotations on sensor data, which most of the existing activity recognition algorithms rely on [14]. In addition, most existing works assume that the true class labels of all incoming instances are immediately available. A more realistic situation is that only a few instances in data streams are labeled. Together with the few labeled data available, we proposed an accurate, cheaper, and fast semi-supervised activity recognition model using label propagation. Thus, we use SSL to address the following issues of the SL algorithm: a) Labeling difficulties of (annotating the start and end of the activity), b) Time-consuming in (labeling large datasets or using slow experiments), c) High cost to (hired team, expert or use special devices for labeling) and d) Expert labeling errors (when the annotator is careless, tired, or some sample are difficult to categorize).

Semi-supervised learning technique designs to automatically exploit unlabeled data in addition to labeled data to improve learning performance, without human intervention [7,15,16]. SSL methods train a classifier based on the large labeled dataset and the small unlabeled dataset only [17]. SSL was introduced to alleviate reliance on large amounts of labeled data by learning the label information from a small set of labeled data samples thus the technique reduces the demand for manual annotations [17]. Many real-world tasks comprise large amounts of unlabeled data but the number of labeled training data is limited [18], because unlabeled data can be easily and cheaply obtained [19,20]. While, labeled data is usually very insufficient [19] because it is difficult, expensive, and time-consuming to gather the corresponding class labels [6]. Semi-supervised learning as a learning technique that is designed to exploit unlabeled data to improve learning performance without human intervention, offers a wide research area [18]. SSL offers the advantage of improving classification accuracy using a few labeled and many unlabeled data for implementing reliable prediction models [21]. Semisupervised learning is subdivided into the Disagreement-based method [22], Low-density separated method [23], Generative method [24], Self-Training [25], and Graph-based method [26].

The graph-based method is an important class of semi-supervised learning techniques [22-27]. It naturally represents data as graphs such that the label information of unlabeled samples can be inferred from the graphs [27]. Graph-based methods semi-supervised learning can be described as a learning method, where a graph is constructed with the nodes corresponding to training instances, while the edges correspond to the relation or similarity between instances [16]; afterward, label information can be propagated according to some criteria [16]. Graph-based methods represent similar samples as connected nodes and construct a graph with these nodes, therefore assuming low density between classes [28]. Similarly, graph-based algorithms are based on manifold assumptions and spread label information on the intrinsic structure of data revealed by both unlabeled and labeled data [29]. These assumptions are sensible since in many real-world problems the neighboring data points or the data points forming the same structure (manifold) are likely to have the same label [19].

There are many researches based on graph-based SSSL techniques. For example, [30] proposed an activity labeling approach based on a graph-based semi-supervised learning algorithm in a health smart home. Another graph-based semi-supervised activity recognition technique was proposed in [29]. The authors use label propagation on a k-nearest neighbor graph to calculate the probability of association of the unlabeled data to each class in this phase. Then we use these probabilities to train the model in a way that each of its hidden states corresponds to one class of activity. Similarly, [28] introduces an activity recognition method that combines small amounts of labeled data with easily obtainable unlabeled data in a semi-supervised learning process. The method propagates information through a graph that contains both labeled and unlabeled data. The method also proposed two different ways of combining multiple graphs based on feature similarity and time. In addition, [17]
introduces a graph-based semi-supervised learning framework for activity recognition that augments labeled datasets through a non-parametric spanning forest construction approach. It does not make prior assumptions on the data distribution; neither requires any pre-defined threshold/parameter for neighborhood exploration. The authors in [30] propose a hybrid semi-supervised anomaly detection model for high-dimensional data that consists of two parts: a deep auto-encoder and an ensemble *k*-nearest neighbor graphs-based anomaly detector.

This study seeks to develop a graph-based semi-supervised activity recognition model using label propagation (LP). When compared with five machine learning (ML) approaches, the results demonstrate that LP achieves superior performance. The main contributions of this research are summarized as follows:

- Develop three semi-supervised activity recognition systems that reduce dependence on large labeled datasets while maintaining high accuracy and ease of training.
- Demonstrate the effectiveness of the proposed graph-based semi-supervised learning algorithms by comparing their performance against logistic regression, KNN, and SVM classifiers.
- Evaluate the proposed graph-based semi-supervised learning algorithms by benchmarking their performance against traditional classifiers, including logistic regression, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM).

2.1 MATERIAL AND METHOD

The proposed methodology can be viewed as a framework consisting of four key components: data preparation, data preprocessing, learning model design, and the final activity classification and evaluation phase

Data Preparation

Prior to applying any machine learning algorithms, it's essential to prepare and comprehend the data. A lack of understanding may lead to potential failures in machine learning outcomes. The dataset utilized in this study was gathered under controlled laboratory conditions using an accelerometer and released by the Wireless Sensor Data Mining (WISDM) Lab. It includes recorded activities from 33 users, with a sampling rate of 20 KHz. The activities documented are Walking, Jogging, Upstairs, Downstairs, Sitting, and Standing. The time series graphical representation of each activity recorded using an accelerometer is shown in Figure 1.

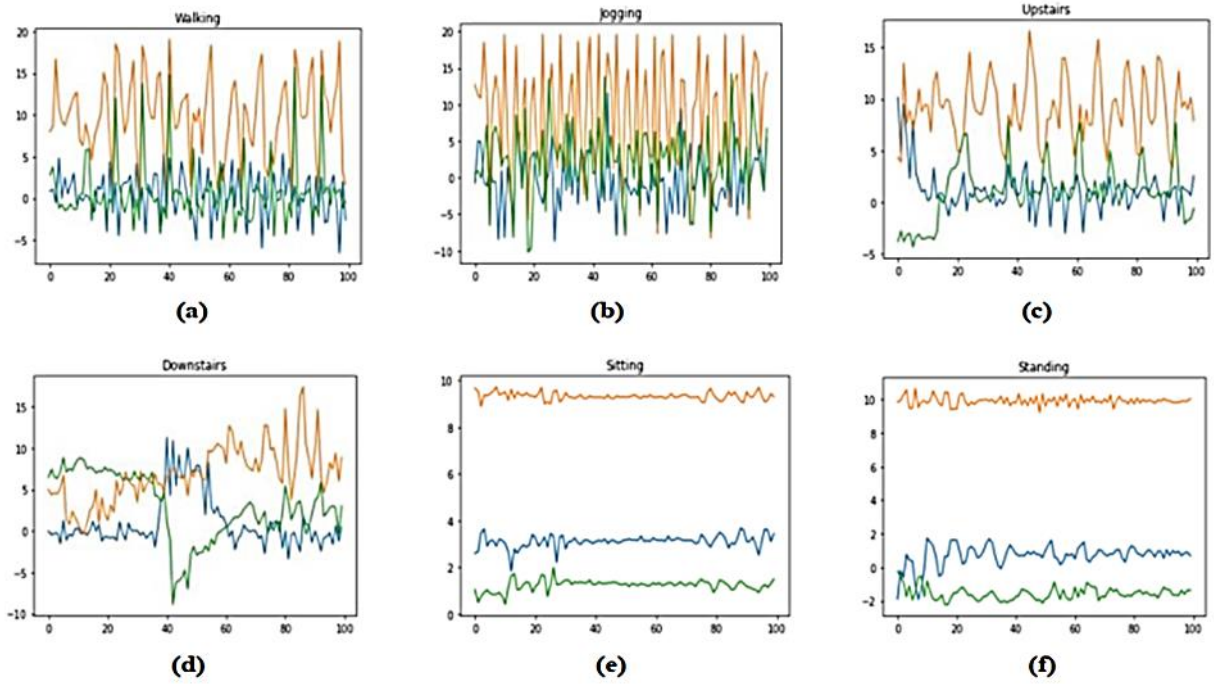

Figure 1. Graphical representation of activities

Data Pre-Preprocessing

The data preprocessing phase includes several key steps: data balancing, data standardization, and label encoding. Additionally, the ability to automatically select a small subset of relevant features from a larger set can improve both computational efficiency and accuracy. Before making predictions, it is essential to balance and scale the features to ensure uniform evaluation. By balancing and standardizing the data, we enhance the efficiency of the proposed activity recognition system. This approach allows us to achieve performance that is comparable to, and in some cases even exceeds, that of fully supervised learning methods, all while utilizing a limited amount of labeled training data.

Table 1. Label assigned to each activity

S/N	ACTIVITY	LABEL
1	Downstairs	0
2	Jogging	1
3	Sitting	2
4	Standing	3
5	Upstairs	4
6	Walking	5

In machine learning, datasets typically contain multiple labels across one or more columns, which can be represented as either words or numbers. To enhance interpretability and make the data more accessible, training data is often labeled using text. In this research, the activity labels were encoded using a Label Encoder. Label encoder converts labels into numeric form to be in the machine-readable form. The numerical value assigned to each activity is shown in Table 1.

Designing learning model

The designing learning model involves data splitting and developing a semisupervised learning model. Data splitting is the process of partitioning data into two portions. One portion of the data is used to develop and train a predictive model and the other to test and evaluate the performance of the model. Data splitting is significant in designing models. This is because; if the data is not split initially the model might be tested with the same data used to train the model. This will also give unreliable accuracy. The normalized and scaled data is initially split into train and test data. The train data is used to train the model while test data is used to test the model. In this work, the available data is split into training and test data in a ratio of 70:30. Unlike supervised learning which used only labeled data to train the model, the semi-supervised learning model used labeled and unlabeled data to train the model. Thus, the training data is further divided into labeled and labeled data. The ratio used for labeled and unlabeled data depends on the experiment conducted. The split ratio for labeled and unlabeled data used in this research work is tabulated in Table 2

Table 2. Split ratio for labeled and unlabeled data

S/N	LABELED DATA	UNLABELED DATA
1	5%	95%
2	4%	96%
3	3%	97%
4	2%	98%
5	1%	99%

Label propagation is a semi-supervised classification algorithm that utilizes a limited amount of manually labelled data to label a significantly larger set of unlabelled data. The process begins with the construction of a weighted undirected graph, where both labelled and unlabelled data serve as nodes, and the relationships between them are represented as edges. When two data points exhibit high similarity, the edge connecting them is assigned a high weight. The label propagation algorithm aims to learn the distribution of labels across the given dataset. It iteratively propagates a small number of manually annotated labels throughout the graph until the algorithm converges or reaches the maximum number of iterations. The results obtained at this point are then utilized to predict the unknown category information for the remaining unlabelled data points. In this manner, the entire training dataset can be labelled effectively. As a result, label propagation is employed to assign labels to the unlabelled data. Our proposed label propagation model is illustrated in Figure 2.

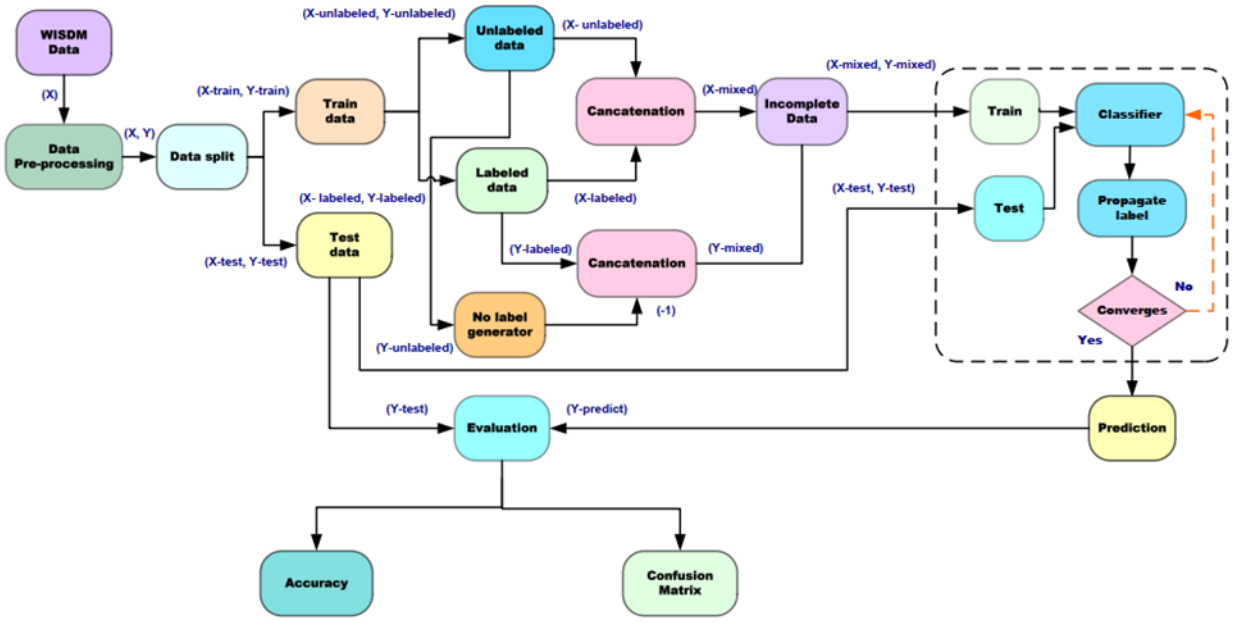

Figure 2. Proposed model

As shown in Figure 2, the WISDM data undergoes preprocessing and is then divided into training and testing datasets. The training data is further categorized into labeled and unlabeled subsets.

$$Training\ data = (Labeled\ data) \cup (Unlabeled) \quad (1)$$

$$Labeled\ data = \{(x_{l1}, y_{l1}), (x_{l2}, y_{l2} \dots \dots x_l, y_l) \quad (2)$$

$$Unlabeled\ data = \{(x_{u1}, y_{u1}), (x_{u2}, y_{u2} \dots \dots x_u, y_u) \quad (3)$$

Also, the class label (Y_L) is given by:

$$Y_L = \{y_{l1}, y_{l2} \dots \dots y_l\} \quad (4)$$

As shown in Figure 2 above, the 'no label' generator replaces all the original labels of the unlabeled data with minus one (-1). Thus the 'no label' blocks create the unobserved labeled (Y_U) which represents as:

$$Y_U = \{y_{u1}, y_{u2} \dots \dots y_u\} \quad (5)$$

Also, the mixed or incomplete data can be represented as:

$$X = \{x_1, x_2 \dots \dots x_{l+u}\} \quad (6)$$

We want to estimate YU from X and YL by constructing a fully connected graph where the nodes are all data points of both labeled and unlabeled data. The edge between any nodes ; is weighted so that the closer the nodes are in local Euclidean distance Dij, the larger the weight wij. The weights are controlled by a parameter gamma.

The Label Propagation model uses the set of labeled data (L) and unlabeled data (U) to construct a graph G = (V, E), where nodes v={1… …n} represent training data and edges E represent the similarity between instances. These similarities are given by a weight matrix W. The weight matrix W can be the k-nearest neighbor matrix or RBF matrix. We decide to use the RBF matrix in this research work.

$$W_{ij} = esp(-\gamma||X_i - X_j||^2) \quad (7)$$

$$W = \begin{bmatrix} W_{ll} & W_{ju} \\ W_{ul} & W_{uu} \end{bmatrix} \quad (8)$$

The probability of jumping from node j to i is:

$$P_{ij} = P(i \to j)\frac{W_{ij}}{\Sigma_{k=1}^{l+u} W_{ij}} \quad (9)$$

The Diagonal matrix can be computed as:

$$D = \begin{bmatrix} D_{ll} & 0 \\ 0 & D_{uu} \end{bmatrix} \quad (10)$$

And

$$D^{-1}W = \begin{bmatrix} D^{-1}W_{ll} & D^{-1}W_{ju} \\ D^{-1}W_{ul} & D^{-1}W_{uu} \end{bmatrix} \quad (11)$$

Then the observe label is:

$$Y_U = D^{-1}WY_L \quad (12)$$

Y_L is the matrix with the actual label of the data set and Y_U is the observed label. The label propagation algorithm is shown in Algorithm 1 below:

Algorithm 1: Label propagation Algorithm

Input: *Labeled data, Unlabeled data, max_iter, and Gamma*

While

1. *Select the set of labeled data (l) and unlabeled data (u)*
2. *Use data points (labeled and unlabeled) to construct a graph G = (V, E) and compute weight matrix W using equation (18).*
3. *Compute the diagonal matrix D*
4. *Iterate:*

$Y_U = D^{-1}WY_L$

Until max_iter is reached

end

Output: *Labels of the unlabeled data*

end

Results

This section showcases the results from the proposed semi-supervised graph-based label propagation classifiers, comparing them with three supervised learning methods: logistic regression, KNN, and SVM, for analysis and evaluation. It provides the result of five experiments carried out when the labeled data is 5%, 4%, 3%, 2%, and 1%. The simulations of this work have been carried out using Python software on a computer with an Intel Core-i3-2.30 GHz CPU and 4GB RAM running Windows 10 with a 64-bit operating system. The parameter used to evaluate the performance of the models is accuracy. In classification problems, Accuracy refers to the number of correct predictions made by the model over all kinds of predictions made. It can also be described as the percentage of correct prediction for the test data. This can simply calculated by dividing the number of correct predictions by the number of total predictions.

The accuracy of the three supervised learning methods and the proposed approach was compared at varying levels of labeled data: 5%, 4%, 3%, 2%, and 1% of the total dataset used for model training. The objective is to leverage the plentiful unlabeled data alongside the limited labeled data to develop accurate, cost-effective, and efficient semi-supervised learning classifiers. The accuracy of Logistic regression (LR), K-nearest neighbor (KNN), Support vector machine (SVM), and Label propagation (LP) classifiers are summarized in Table 3

Table 3. Accuracy recorded for all five experiments

S/N	EXPERIMENT	LABELED DATA	LR	KNN	SVM	LP
1	Experiment 1	5%	90.13	98.91	85.93	99.74
2	Experiment 2	4%	88.96	98.63	81.87	99.63
3	Experiment 3	3%	86.39	97.50	78.04	99.04
4	Experiment 4	2%	83.74	96.29	65.41	98.67
5	Experiment 5	1%	78.89	93.07	61,37	95.56

Discussion

The accuracy recorded for all five experiments carried out is a graphical representation in Figure 3 for easy analysis and interpretation. It was observed from Figure 3 that the accuracy of the activity classification of SL algorithms decreases with the decrease in the number of manually labeled samples. However, the accuracy of the proposed method is slightly dropped. Supervised algorithms typically achieve satisfactory recognition performance in the first experiment when labeled data is 5%. But as labeled data decreases their accuracy gradually decreases because they require significant amounts of labeled activity data for training a classifier. On the other hand, the LP algorithm maintains its performance despite the slight drop in its accuracy as labeled data decreases. In addition, the LP algorithm achieves high accuracy with little labeled data used for the model training. Thus the proposed graph-based semi-supervised label propagation algorithm performs better than the three SL algorithms.

Accuracy recorded for all five experiments

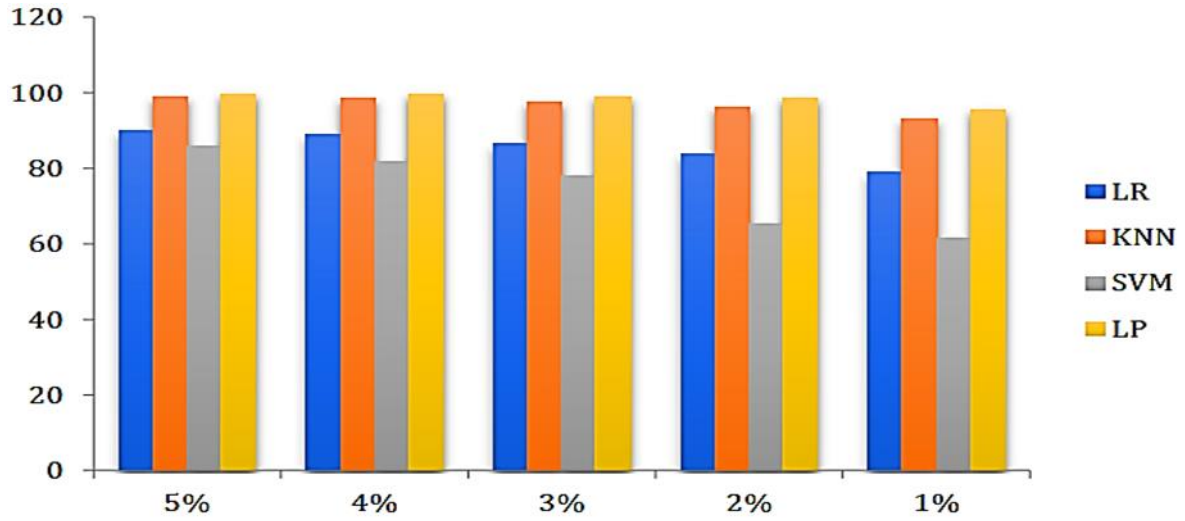

Figure 3. Graphical representation of the Accuracy of all the experiments

Conclusion

This work investigates the application of semi-supervised learning for developing a model for human activity recognition, aiming to alleviate the tedious, costly, and time-intensive challenges associated with data labeling in supervised learning. Although significant research has been conducted on various aspects of human activity recognition recently, challenges still persist within the community. Since collecting labeled data samples is usually difficult, time-consuming, and expensive. Hence, achieving a good learning model with limited labeled samples is a crucial issue. Our proposed semi-supervised learning model addresses this problem by reducing human effort. We use a large amount of unlabeled data, together with the few labeled data, to build an accurate, cheaper, and faster classifier. Thus the problem of insufficient label information could be alleviated. In this research work, our main focus is the Semi-Supervised Learning technique for Human Activity Recognition. In conclusion, our semi-supervised learning (SSL) model demonstrates the ability to achieve higher accuracies than supervised methods with minimal labeling effort. However, the label propagation (LP) approaches do have some limitations, including memory requirements and computational costs, particularly when applied to larger datasets than those utilized in this study. In future work, we plan to implement and compare the performance of our model against other semi-supervised learning algorithms, including the generative method, the disagreement method, and the low-density separation method.

REFERENCES

[1] H. Alemdar, T. L. M. Van Kasteren, and C. Ersoy, (2011). Using active learning to allow activity recognition on a large scale. International Joint Conference on Ambient Intelligence, 105–114, doi: 10.1007/978-3-642-25167-2_12.

[2] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, (2019). Deep learning for sensor-based activity recognition : A urvey. Pattern Recognition Letters, 119, 3–11, doi: 10.1016/j.patrec.2018.02.010.

[3] N. Qamar, N. Siddiqui, M. Ehatisham-Ul-Haq, M. A. Azam, and U. Naeem, (2020). An approach towards position-independent human activity recognition model based on wearable accelerometer sensor. Procedia Computer Science, 177, 196–203, doi: 10.1016/j.procs.2020.10.028.

[4] Ó. D. Lara and M. A. Labrador, (2013). A survey on human activity recognition using wearable sensors. IEEE Communications Surveys and Tutorials, 15(3), 1192–1209, doi: 10.1109/SURV. 2012. 110112.00192.

[5] P. T. Chinimilli, S. Redkar, and W. Zhang, (2017). Human activity recognition using inertial measurement units and smart shoes. Proceedings of the American Control Conference, 15(3), 1462–1467, doi: 10.23919/ACC.2017.7963159.

[6] P. Bota, J. Silva, D. Folgado, and H. Gamboa, (2019). A semi-automatic annotation approach for human activity recognition. Sensors (Switzerland), 19(3), 1–23, doi: 10.3390/s19030501.

[7] D. Guan, W. Yuan, Y. Lee, A. Gavrilov, and S. Lee, (2007). Activity Recognition Based on Semi-supervised Learning. 13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), 469–475.

[8] M. Stikic, D. Larlus, S. Ebert, and B. Schiele, (2011). Weakly supervised recognition of daily life activities with wearable sensors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(12), 2521–2537, 2011, doi: 10.1109/TPAMI.2011.36.

[9] X. Guan, R. Raich, and W. K. Wong, (2016). Efficient multi-instance learning for activity recognition from time series data using an auto-regressive hidden markov model. 33rd International Conference on Machine Learning (ICML), 5, 3452–3473.

[10] H. Kwon et al., (2020). IMUTube: Automatic Extraction of Virtual on-body Accelerometry from Video for Human Activity Recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(3), 1-29., doi: 10.1145/3411841.

[11] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, (2014). Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. IEEE Transactions on Geoscience and Remote Sensing, 53, (6), 3325–3337.

[12] R. Zhao, L., Sukthankar, G., & Sukthankar, (2011). Robust Active Learning Using Crowdsourced Annotations for Activity Recognition. Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence, 74–79.

[13] H. Qian, S. J. Pan, and C. Miao, (2019). Distribution-Based Semi-Supervised Learning for Activity Recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 7699–7706.

[14] A. R. Sanabria and J. Ye, (2020). Unsupervised domain adaptation for activity recognition across heterogeneous datasets. Pervasive and Mobile Computing, 64(April 2020), 101147, doi:10.1016/j.pmcj.2020.101147.
[15] J. E. van Engelen and H. H. Hoos (2020). A survey on semi-supervised learning. Machine Learning, 109(2), 373–440, doi: 10.1007/s10994-019-05855-6.
[16] Z. H. Zhou, (2018). A brief introduction to weakly supervised learning. National Science Review, 5(1), 44–53, doi: 10.1093/nsr/nwx106.
[17] Y. Ma and H. Ghasemzadeh, (2019). LabelForest : Non-Parametric Semi-Supervised Learning for Activity Recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 4520–4527.
[18] Z. H. Zhou and M. Li, (2010). Semi-supervised learning by disagreement. Knowledge and Information Systems, 24(3), 415–439, doi: 10.1007/s10115-009-0209-z.
[19] F. Nie, S. Xiang, Y. Liu, and C. Zhang, (2010). A general graph-based semi-supervised learning with novel class discovery. Neural Computing and Applications, 19(4), 549–555, doi: 10.1007/s00521-009-0305-8.
[20] X. Zhang, L. Yao, and F. Yuan, (2019). Adversarial variational embedding for robust semi-supervised learning. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 139–147, doi: 10.1145/3292500.3330966.
[21] I. E. Livieris, K. Drakopoulou, V. T. Tampakas, T. A. Mikropoulos, and P. Pintelas, (2019). Predicting Secondary School Students' Performance Utilizing a Semi-supervised Learning Approach. Journal of Educational Computing Research, 57(2), 448–470, 2019, doi: 10.1177/0735633117752614.
[22] H. He, D. Han, and J. Dezert, (2019). Disagreement based semi-supervised learning approaches with belief functions. Knowledge-Based Systems, 193, 105426, doi: 10.1016/j.knosys.2019.105426.
[23] S. Ji, Y. Peng, H. Zhang, and S. Wu, (2021). An Online Semisupervised Learning Model for Pedestrians' Crossing Intention Recognition of Connected Autonomous Vehicle Based on Mobile Edge Computing Applications. Wireless Communications and Mobile Computing, 2021(February 2021), 1–14, 2021, doi:10.1155/2021/6621451.
[24] A. Zahin, L. T. Tan, and R. Q. Hu, (2019). Sensor-based human activity recognition for smart healthcare: A semi-supervised machine learning. International conference on artificial intelligence for communications and networks., 450–472, doi: 10.1007/978-3-030-22971-9_39.
[25] D. Wu et al., (2018). Self-training semi-supervised classification based on density peaks of data. Neurocomputing, 275, 180–191, doi: 10.1016/j.neucom.2017.05.072.
[26] Z. Hua and Y. Yang, (2022). Robust and sparse label propagation for graph-based semi-supervised classification. Applied Intelligence, 52(3), 3337–3351, 2022, doi: 10.1007/s10489-021-02360-z.
[27] Y. Chong, Y. Ding, Q. Yan, and S. Pan, (2020). Graph-based semi-supervised learning: A review. Neurocomputing, 408, 216–230, 2020, doi: 10.1016/j.neucom.2019.12.130.
[28] V. Cheplygina, M. de Bruijne, and J. P. W. Pluim, (2019). Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Medical Image Analysis, 54(May 2019), 280–296, 2019, doi: 10.1016/j.media.2019.03.009.
[29] S. Ud Din, J. Shao, J. Kumar, W. Ali, J. Liu, and Y. Ye, (2020). Online reliable semi-supervised learning on evolving data streams. Information Sciences, 525, 153–171, doi: 10.1016/j.ins.2020.03.052.
[30] Y. Hu, B. Wang, Y. Sun, J. An, and Z. Wang, (2020). Graph-Based Semi-Supervised Learning for Activity Labeling in Health Smart Home. IEEE Access, 8, 193655–193664, 2020, doi: 10.1109/ACCESS.2020.3033589.