A Smartphone Sensors-based Personalized Human Activity Recognition System for Sustainable Smart Cities

Abstract

According to the Sustainable Development Agenda 2030 of the World Health Organization, maintaining physical activities have multiple societal privileges for healthier cities and societies. The amalgamation of the Internet of Things (IoT) and pervasive smartphones has become of paramount importance to produce a significant breakthrough in various domains of smart cities, including healthcare, fitness, skill assessment, and personal assistants, to support independent living. The IoT-supported devices capacitate, embedded with sensors, enabled numerous context-aware applications to recognize physical activities. There are some activity recognition applications; however, they are still deficient in recognizing activities accurately. In this paper, a novel framework for Human Activity Recognition (HAR) is proposed using raw readings from a combination of fused smartphone sensors: *accelerometer, gyroscope, magnetometer*, and *Google Fit* activity tracking module. The proposed framework applies Deep Recurrent Neural Network (DRNN) to an extensive training dataset. The latter consists of five activity classes from twelve individuals using a Deep Recurrent Neural Network (DRNN). An extensive training dataset is used consisting of five activity classes from a group of twelve individuals. The designed android application (runs in the background) collects data from the smartphone's embedded sensors fused with the Google Fit API to validate the results proposed framework. The proposed framework shows promising results in recognizing human activities compared to other similar studies and achieves an accuracy of 99.43% for activity recognition using DRNN.

Keywords: Human activity recognition, Pervasive computing, Smartphones, Sensors, Deep recurrent neural network

1. Introduction

As reported by the population division of the United Nations [1], by the year 2050, 66% of the world population will move to cities. Due to this rapid increase in urban population, a smart city's concept has become essential to improve urban life by endorsing and supporting sustainability and healthier urban environments. While smart technology is already transforming a city's critical infrastructures, lifestyle, and services (e.g., education, transportation, and public safety), citizens have now recognized its potential to tackle the challenges of health and environment [2, 3, 4, 5]. Moreover, the need to reduce healthcare costs and sustain a healthier life is an important driving factor for governments to invest in smart cities [6, 7]. IoT is about anytime, anywhere, service provisioning to end-users, thanks to a plethora of static and mobile devices, such as actuators, sensors, and controllers [8]. The intense communications between all these devices, some are smart (i.e., embedded with cognitive capabilities), have allowed IoT to penetrate every domain from home automation to the industry 4.0 revolution.

Human Activity Recognition (HAR) using sensors are imperative in meeting the needs of the urban population in terms of healthcare-related services [9, 10, 11, 12, 13]. Care providers (i.e., Doctors, physicians, nurses, gym trainers, etc.) can assess an individual's health conditions based on their daily life activities [14] to maintain economic growth while creating sustainable cities and societies [15, 16]. Many studies use sensing frameworks as benchmarks to extract and analyze people's daily life routines and behavior [17, 9]. Also, citizens have become more health-conscious in the modern era, and they care more about living healthy lifestyles [18]. Many diseases can be detected by analyzing physical activities, such as Parkinson's [19, 20], and Dementia [21]. Several other pervasive techniques help analyze individuals' health and functional ability living in smart homes [21, 15]. A smart home, embedded with sensors, provides readings against a smart home resident's activities being performed. These readings can later be analyzed to detect the functional ability of an individual. Privacy is a primary concern of smart cities as most of the data in smart cities is being collected from vision-based devices [22]. Vision-based activity recognition can be used to analyze the health status, but due to privacy measures, it is not widely applied [23]. Advancement in ubiquitous computing brings Wi-Fi signal-based activity recognition, but it can only operate in a defined area [24]. Another technique used for recognizing activities activity recognition technology is based on wearable sensors. However, due to their obtrusive nature, it is not widely used [25].

Since most users carry a smartphone activity as an inveterate while performing their daily routines (i.e., browsing, banking, measuring, chatting, etc.), it has proven to be the best daily physical activity observer [26]. Many authors have surveyed HAR using smartphone sensors in recent studies, such as in [27, 28, 29, 30, 31, 32, 33]. Most of the previous work focused on walking, running, standing, sitting, moving upstairs, and moving downstairs.

Studies have used deep learning and conventional machine learning models. For example, [34] and [35] used deep learning techniques to recognize human activities. However, these methods are fully connected without capturing the local dependency of the sensor data. Similarly, conventional machine learning models have been used in different studies (e.g., in [9, 36, 17]). However, they show less proficiency in providing promising results. They also lack in recognizing the activities accurately as they focus more on routine activities. In our case, we choose complex activities, such as "In a vehicle, tilting, Still, On Foot, and Walking". Thus, a novel framework is proposed for the recognition of physical activities using a self-collected dataset. The dataset contains fused readings of multiple sensors: accelerometer, gyroscope, and magnetometer with Google Fit API readings by applying a Deep Recurrent Neural Network (DRNN). We apply machine learning algorithms and evaluate the proposed framework using F-score, Recall, Accuracy, and Area Under the Curve (AUC). We focus on tracking complex activities that are also being recognized by Google Fit [26].

This paper makes the following contributions:

1. Presents a study of hardware sensors' utilization, fused with Google Fit activity monitoring API for HAR.
2. Presents the deployment of DRNN, along with the evaluation of its advantages in comparison with state-of-the-art methods.
3. Investigates various DRNN's parameters to explore the factors that affect the performance.
4. Provides a systematic and functional framework to monitor daily life activities continuously.

The rest of the paper is organized as follows: Section 2 discusses the related work and recent developments in activity recognition. Section 3 gives a background of Google Fit API and deep recurrent neural network. Furthermore, Section 4 presents the proposed methodology for HAR, and Section 5 describes the detailed evaluation and results of our proposed framework. Lastly, Section 6 gives conclusions of our research and gives direction for future work.

2. Related Work

This section confers the most befitting state-of-the-art and discusses related research studies to sustainable health monitoring through HAR. A sustainable healthcare monitoring system is accomplished by conveying excellent public well-being without debilitating natural assets or causing extreme ecological harm [37]. It is essential for smart cities and society to guarantee people's well-being and prosperity for sustainable development and is feasible just through powerful and ceaseless health monitoring services [38]. Many smart devices are being used to monitor to recognize daily life activity and other critical diseases. Activity recognition is the subdomain of healthcare monitoring which is directly related to public well-being in a society. The worldwide smart wearable healthcare (SWH) device market is required to ascend at a Compound Annual Growth Rate (CAGR) of 5.6%, and by 2020 it is relied upon to arrive at 25 Billion [39]. The developing trends of the way of life ailments, sedentary lifestyles, busy work routines, innovative progressions in healthcare monitoring devices, and expanded utilization of smart devices seem to be a portion of the significant elements fuelling this development [39].

The authors in [9] uses a novel system of smart, collaborative healthcare to enhance an individual's lives using smartphone sensors and machine learning algorithms. The data statistics generate a report, which the professionals are examining. [36] proposed a system that uses the smartphone 2-axes accelerometer sensor to recognize user activities. [40] thoroughly reviews the current robust studies based on deep learning, which uses sensors for recognizing human activities. [41] proposed a novel deep learning architecture for multiple sensor-based HAR. Their system processes time-series data as an image and uses computer vision methods for activity classification. [8] enhanced HAR by embedding a deep learning algorithm in a sensor-based wearable. This wearable tracks the sensor data, and thus, a deep learning algorithm classifies human activities. They used Convolutional Neural Network (CNN) for classification. The authors in [42] present a comparison for data traffic scheduling techniques to increase mobile networks' quality.

The authors in [43] discuss the differences and advantages of 5G in the healthcare domain with preceding generations. The authors highlighted the impact of 5G-based systems in delivering care, diagnosis, imaging, and medication to improve living standards and health. Similarly, Latif et al. [44] feature 5G bounties and some other technologies in healthcare applications. They also discussed the use of machine learning algorithms in mitigating healthcare anomalies. Chen et al. [45] propound a 5G-Smart diabetes system for diabetic patients using sensors and analysis of the patients' vitals. They proposed a personalized data sharing and analysis system for their proposed 5G-Smart Diabetes model. [17] proposes a machine learning approach, which uses the clustering of data and uses machine learning classifiers for human activity classification. Lloret et al. [46] present a novel 5G-based smart eHealth monitor architecture for chronic patients. To carry out diagnosis and data collection, they used wearable devices. They also use smartphones at the patient's end to process and analyze the received data (i.e., sensed from wearable devices).

In recent studies, many HAR systems have been surveyed [27, 28, 29, 31, 32]. The authors worked on several methods for several activities in distinguished application domains [47]. Several activities include walking, standing, sitting, running, upstairs, downstairs, cooking, toileting, eating, sleeping, exercising, etc. Some of these can be recognized using time duration-based analysis, some activities can be recognized using binary state sensors, and some can be recognized using motion sensors. Activities can be further broken down into two classes based on the activities' duration and complexity: simple activities and complex activities. The simple activities consist of repetitive behavior activities like sitting, standing, running, walking, etc. The latter combines basic activities while interacting with other objects, such as toileting, cleaning, chores, cycling, driving, etc.

As afore-mentioned, the HAR has been studied and explored well [48]. These studies sometimes vary on the used sensors and some time on the proposed methodology. Sensors include motion, pressure, proximity, microphone, video sensors, reed switches, analog and binary sensors, etc. Besides, they may help recognize different activities, such as walking, sleeping. In the early stages of activity recognition, several researchers produce different approaches for activity recognition. The authors in [49] provide results of activity recognition back in 2004. Their results show that sensors placed at different body locations can be used to track physical activities. They show that the thigh sensor shows an efficient activity recognition rate than all other places. Later on, authors in [50] recognize activities using smartphone sensors. They collected accelerometer sensor data. They report that J48 and MLP perform well in comparison with other techniques. Shoaib et al. [51] demonstrated a combination of sensors used in recognizing the activities. Their work shows that a combination of accelerometer, gyroscope, and magnetometer could enhance the recognition rate.

There is not much research is done on deep learning-based techniques for HAR. Among the primitive works are [34, 35]. [34] uses Restricted Boltzmann Machines (RBM) for feature extraction from the time-series sensor data automatically. The disadvantage of the technique mentioned above is that these methods are fully connected, and there is no capturing of the local dependency of sensor data [52]. The authors in [53] use convolutional neural networks (convnets) with accelerometer and gyroscope sensor readings for recognizing gestures. They show that convnet provides promising results compared to previous activity recognition methods, including Dynamic Time Wrapping (DTM) and Hidden Markov Model (HMM). Table 1 provides an accuracy of existing studies focusing on HAR. The highest accuracy achieved by [36] is 95.7% for standing, 95.0% by [54] for sitting and 98.5% by [55] for walking. All of the studies in Table 1 have not focused on complex activities such as In-Vehicle and tilting, which are considered in this paper.

Table 1: Summary of the Sensor, Activities and Results (Precision) used in Existing Studies

Studies	Sensors	Standing (On Foot)	Sitting (Still)	Walking
[55]	Accelerometer	93.3	82.6	98.5
[36]	Accelerometer	95.7	94.0	96.5
[54]	Accelerometer	94.9	93.9	86.3
[50]	Accelerometer	91.9	95.0	91.7
[51]	Accelerometer, Gyroscope and Magnetometer	-	91.0	88.9

3. Preliminaries

An overview of the Google Fit and Deep Recurrent neural network (DRNN) is presented in this section.

3.1. Google Fit Overview

Google Fit application is capable of tracking human physical activity [26]. Background service is executed for data collection. The data is collected based on five daily physical activities: walking, running, still, on a bicycle, and cycling. Individuals can monitor their step counter, activity duration, and countdown, etc. An estimate of burned calories, while an activity, can also be made. Some limitations that are also addressed in [26] are activity recognition inaccuracy and inefficiency. Besides, high accuracy mode (distance) reduces battery life. These are the two main limitations that we are addressing in this work.

It will be true to assume that deep learning will dominate other techniques for activity recognition soon. This paper aims to provide promising activity recognition performance using Deep Recurrent Neural Network DRNN while using our fused methodology.

3.2. Deep Recurrent Neural Network Model

Figure 1 shows the DeepDRNN consisting of L number of layers. DRNN network is a network embedded with LSTM units on the internal layers [56]. Here, x_t represents the input vector, and y_t represents the output vector at time t. The output of the lth internal layer at time t is shown as $RNN_{l,t}$. A unit generates each element of the output vector. The amount of units in each layer depends upon the dimension of each layer. The proposed DeepDRNN model comprises an error function, gradient descent, an LSTM layer, over-fitting, and hyper-parameter tuning.

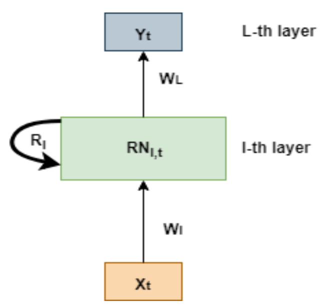

Figure 1: Representation of the Deep RNN

- **Error Function**: is an essential part of calculating each participating neuron's output. After error analysis, the weight of the neurons is updated to improve the output. While solving a multi-class problem, the output is classified into $C_1, C_2, \ldots, C_{ns}$, based on the softmax function. The n^{th} unit output may be illustrated using Equation 1.

$$y_n \equiv z_n^{(L)} = \frac{exp(u_n)}{\sum_{q=1=p(C_n|x)}^{n}} \tag{1}$$

The probability of each unit belonging to class C_n is represented by y_n.

Here x is the input that is classified into the majority instance class.

$$E(w) = -\sum_{i=1}^{I}\sum_{n=1}^{H} d_{ih} log y_i(x_n; w) \tag{2}$$

Error is calculated by Cross entropy, as shown in Equation 2. The above equation describes the error function. The d_i represents ith vector and the d_{ih} shows h^{th} elements of d_i.

- **Mini Batch Stochastic Gradient Descent:** is used for minimizing the error rate. Let e be the number of elements of w; following Equation 3 represents the gradient of the error function.

$$\nabla E = \frac{\partial E(w)}{\partial (w)} = \left[\frac{\partial E(w)}{\partial w_1} \cdots \frac{\partial E(w)}{\partial w_e}\right]^T \tag{3}$$

To reduce the error rate, the local minimum value is searched in the neighborhood *w* by the gradient descent. If gradient descent is going in a negative direction, it changes the gradient descent value by a small amount to go in the right direction.

$$E^t(w) = \frac{1}{||B^t||}\sum_{n \epsilon B^t} E_n(w) \tag{4}$$

In Equation 4, the learning rate of ϵ is updated in each repetition to find the local minimum of the error function. With quite a higher learning rate, it oscillates around the local minimum. To control this, ADaptive moment estimation (Adam) based adjustment methods are adopted.

- **Long Short Term Memory (LSTM):** is a type of Neural Network model for time sequence data. LSTM is used to solve the problem of vanishing gradient and is more efficient than RNN [57]. LSTM consists of three gates: forget gate, input gate, output gate, and one cell state. Each gate has a different set of weights.

- **Over-fitting:** Deep learning is more efficient than long-established neural networks. However, there are still some disadvantages: time consumption and over-fitting. Over-fitting shows high variance and low bias. It might happen when a model faces too many similar characteristics of the training data characteristics and does not fit on the unseen test data. There are multiple ways to prevent over-fitting. Regularization is used to prevent over-fitting. Moreover, the dropout procedure is often adopted to overcome the over-fitting by selecting the layers at a constant rate of p. There are some other scenarios by which a model gets over-fit: Over-fitting might occur if there is not enough data to train it. It can be due to training neural networks on too many epochs on iterations. It can also occur due to more minor variations in the dataset. Most of the training samples are alike, which cripples the algorithm and does not let it generalize. In the process of activity recognition, a thorough analysis of the hyper-parameters, such as activation function, optimizer, the numbers of layers, and batch size, should contribute to the recognition process using the DRNN, as explained in Table 2.

- **Hyper-Parameter Tuning:** A massive combination of hyper-parameter settings could be possible for the recurrent neural network. Therefore, this paper analyzes the effects of varying hyper-parameters on the performance of the DRNN. We incorporated the grid search approach by analyzing the previous best model. Tuning starts from the number of layers L1 (one-layer, L1; two-layer, L2; three-layer, L3; and four-layer, L4) learning rate affects the recognition rate. Other parameters, such as weight decay, momentum, and epochs size, also contributed toward improving the recognition rate.

4. Human Activity Recognition (HAR)

This section presents the proposed framework for the Human Activity Recognition framework (HAR). Fig. 2 illustrates an abstract overview of the proposed framework. With the help of an illustration.

4.1. Smartphone Sensing and Dataset Collection

We develop an android application that can be installed on the participant's smartphone. However, there are various smartphones; this application is meant for only every Android-based smartphone only. For Android, the requirement of the operating system is fixed to 4.0. The following are the characteristic of the proposed application.

- Graphical user interface.

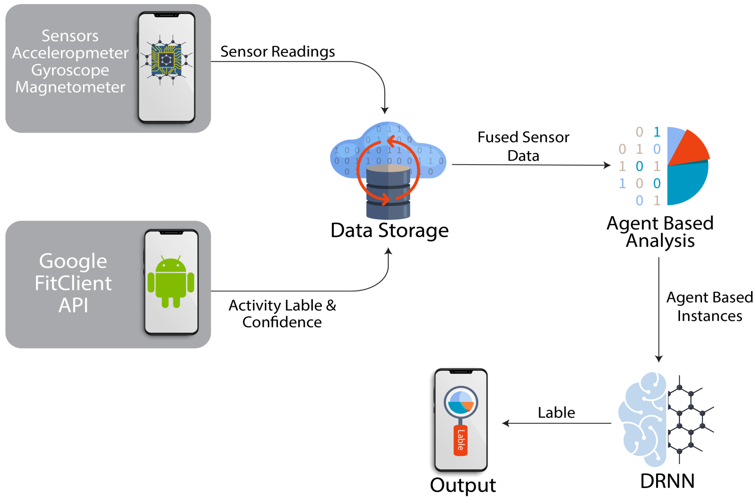

Figure 2: Abstract data flow architecture of HAR. The illustration shows smartphone sensor data begin fused with Google FIT data. The agent-based analysis is applied to the fused data and agent-based instance features passed onto the DRNN module which classifies the activity.

- A continuous service running in the background to collect data from the (three) sensors and Google Fit API at the same timestamp.
- Accelerometer, gyroscope, and magnetometer were selected and fused with Google Fit API.
- Collected data stored in a file on the smartphone.

An extensive analysis is conducted before fusing Google Fit API with a smartphone sensor. Hardware sensors work independently and do not share information. Firstly, we fuse three sensors to get the data from these sensors at the same timestamp. After this, we fused Google Fit API with these sensors to get data reading in the same timestamp, as shown in Fig. 2.

4.2. Feature Extraction

To obtain the optimum results, the raw data is transformed into a sensor event window, a window having 1000 activity samples of every participant. That is a diverse window, which captures the required readings needed for classification. Contemporary researchers have suggested this length [31, 50]. We pick a window and appended this window into a file. Later, a feature matrix of 130000 raw sensor data observations containing three axes each of three sensors, $[Ax_1, Ay_2, Az_3]$, $[Gx_1, Gy_2, Gz_3]$, $[Mx_1, My_2, Mz_3]$, is generated. The accuracy of the activity being performed is stored between $0 - 100$ as a label. An input matrix of the 11 axis feature matrix is given to the proposed machine learning model DRNN to be trained on a large dataset to predict the activity being performed as described in Section 3. The output of the model is a label of an activity. This label is then sent to the *Agent-based* Analysis (ABA) 4.3. Similarly, the labels from Google Fit API are also sent to this *ABA* to make further progress. Working of *ABA* is discussed next.

4.3. Agent-based

Analysis (ABA) This agent is responsible for analyzing the accurate label of activity being performed by an individual. In first, Google Fit API returns a list of labels containing each activity name and their confidence level in a percentage ranging from $0 - 100$. The agent mines these activities and selects only those activities with a confidence level greater than 50%. The activity confidence value processed by the agent is then appended with the reading returned by the smartphone sensors. It is passed as an input to the proposed machine learning model, as shown in Fig. 2.

4.4. Activity Recognition Algorithm

Algorithm 1 summarizes the steps of the proposed framework. Suppose, d represents the dataset containing instance $I = i_1, i_2, \ldots, i_n$. Let l represents the target class labels to be predicted by each classifier, and nl represents the total target classes. The data preprocessing phase normalizes data and removes duplicates stored as d_{new}. The preprocessed data is given as an input to our agent-based analysis (ABA) that checks the thresholds of each fused instance denoted as I_{new}. In the next phase, I_{new} is fed to the machine. And deep learning classifiers for activity recognition.

4.5. RNN Based Activity Recognition

HAR uses fused sensor reading along with fused Google Fit activity tracking module on a smartphone as the direct input after performing agent-based analysis, as explained in Section 4.3. We use DRNN such that the 10-axis data corresponds to the input layer of 10-dimensions and five activity labels correspond to the five-dimensional output layer. Neurons on the internal layer are computed using the LSTM unit. A softmax defines the output layer's activation function with a cross-entropy loss function common in multinomial NN classifiers. The network

Algorithm 1 HAR Algorithm for Activity Recognition

procedure HAR(d)
  $d_{new} \leftarrow Preprocess\ (d)$
  for $i \in d_{new}$ **do**
    $I_{new} \leftarrow ABA\ (d_{new})$
    $I_l \leftarrow classifiers\ (I_{new})$
    for each epoch in range (n): **do**
      Calculate Loss
      Calculate Accuracy
      Calculate Precision, Recall, F-Score, Roc Curve
    return Output
  return I_l

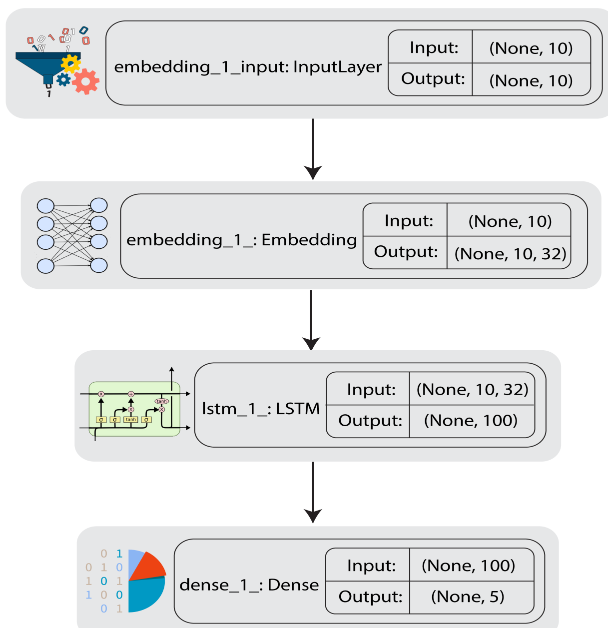

Figure 3: Proposed design of the Deep RNN network employed by HAR.

provides an activity class as an output label. Figure 3 illustrates the architecture of the proposed Deep DRNN used for activity recognition.

The details of RNN are summarized in Table 2 and the flow of epochs is also shown in Fig. 4.

Table 2: Details of Recurrent Neural Network Parameters

Parameter	Value
Initial weights	rand [-0.1, 0.1]
Initial Bias	None
Type of internal layer unit	LSTM
Input dimension	9
Output dimension	5
Activation Function	Softmax
Mini-batch size	20
Learning Optimizer	Adam
Error Function	Categorical CrossEntropy

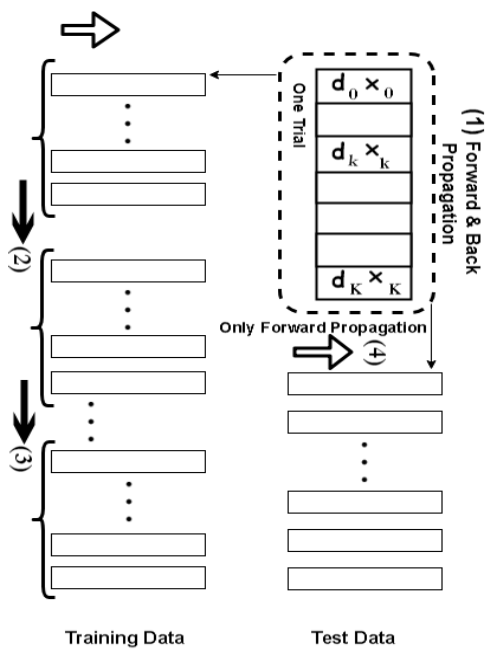

Figure 4: Flow of data during an epoch through the DRNN [56]

4.6. Comparative Methods

The results are compared using Sequential Minimal Optimization (SMO), decision tree (c4.5), Meta-Heuristic algorithm (Adaboost), Naive Bayes (NB), and Multilayer Perceptron (MLP) with state-of-the-art techniques. The SMO is used with various parameters and finally set to the Polykernal and. The value of gamma and alpha is set to 1. NB is used with a kernel estimator. The decision tree is used with a confidence factor of 0.25. Existing studies [49, 50] prefer the data sensing frequencies between $10-50$ samples per second and a window of $2-3$ seconds for the recognition task. Lastly, for the constructed feature matrix, machine learning methods SMO, c4.5, Adaboost, MLP, NB, and DRNN is applied.

5. Experimental Evaluation

This section presents different evaluation measures used in the experimentation. We then explain and discuss the activity recognition results. Conventional machine learning classifiers are applied using WEKA [58]. We try 3, 5, 7 Fold-Cross-validation and select 5 Fold-Cross-validation as the F-score is promising using five folds. For DRNN, data is split into 80 and 20 for training and testing, respectively. Furthermore, 20% of the total 80% training data is used for validation. The main idea of using RNN is finalized after performing statistical analysis, initial visualizations, and classification reports. The data is comprised of activities performed; thus, the temporal information becomes vital in this regard. In contrast, CNN does not make use of temporal information. For time-series repeated pattern-based data, the RNN not only recognizes but may also exploit the time-related context. More weights are applied to the patterns for recognition. Furthermore, the previous and following tokens are recognized instead of being evaluated in isolation. Consequently, this led us to select RNN for our use case.

5.1. Dataset

The data is collected from twelve volunteers using different smartphone models:*Oppo F3, Oppo F1, Samsung J7, Samsung Grand Prime*, and *Huawei Honor*. Table 3 demonstrate the details of the dataset. Some of the smartphones are not embedded with a magnetometer or with a gyroscope. In this case, we use these phone datasets only to build a model for accelerometer data. The application permitted us to control the frequency of collecting sensor readings. The frequency is set, keeping in sight the battery life. Invoking the hardware sensors continuously consumes more battery [59]. Therefore, a fixed sampling frequency of $30/sec$ is used to collect the data. This frequency is fine between $10 - 50$ samples per second and reduces less battery as preferred by existing studies [50, 59]. The more the sensors are invoked, the more it consumes battery. A larger dataset is required to build a most effective framework. The data collection duration is approximately $2 - 3$ minutes which is efficient enough to learn the classes' boundaries. This task is controlled and administered by our group members for data quality assurance.

Table 4 shows the features used in this study. Existing studies such as [36] used

Table 3: Dataset Characteristics

Dataset	Self Collected
Number of Participants	12
Mean Age	25
Total Activities	5
Still	When the participant is not moving
Tilting	When the participant is moving
On Foot	When the participant is on foot
Walking	When the participant is walking
In a Vehicle	When the participant is moving in a vehicle

only three axes of the Accelerometer sensor. [17, 51] used all features that are specified in Table 4. [55, 50] used only three axes of the Accelerometer sensor.

Table 4: Features of the dataset

Number	Feature Name
1	Accelerometer-Ax
2	Accelerometer-Ay
3	Accelerometer-Az
4	Gyroscope-Gx
5	Gyroscope-Gy
6	Gyroscope-Gz
7	Magnetometer-Mx
8	Magnetometer-My
9	Magnetometer-Mz

The dataset contains only numerical features. For the ground truth data, manual labels are assigned to each sensed observation for the already-known task being performed. Fig. 5 shows one vector of data collected using smartphone sensors for walking activity. Each axis of the three sensors is plotted to observe the behavior of the walking patterns. It can be observed that while walking, a lively acceleration is being produced, as shown in the blue line Ax axis. This line is showing the continuous movement of the leg in a positive x-axis direction. The Ay axis movement is continuous, and the same for the whole activity as the leg is

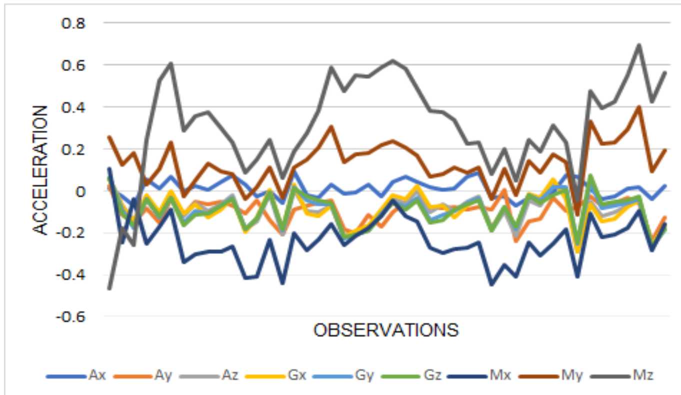

Figure 5: One Vector of Walking Activity Accelerometer, Gyroscope and Magnetometer Sensor Data

not moving towards the upward. Similarly, the large peaks of the Az axis represent the upside-down forward movement of the leg. The same is the case with the other sensors showing a periodic pattern. Gx and Gz axis present normal uninterrupted movement. It may be observed that there is a sudden acceleration peak in the Gy axis representing the change in leg transition. The magnetometer sensor is essential for identifying a smartphone's orientation relative to the Earth's magnetic north. Mx, My, and Mz axes contribute a lot for activity recognition as these show sudden peak and then low peak while walking, representing leg movement. The same characteristics are shown by a study [55, 50], which proves the proposed framework's applicability. Furthermore, it can be seen in Fig. 6 that when a participant switched his/her state from walking to sitting, all the acceleration readings become stable and flat and do not show periodic behavior.

5.2. Experimental Settings

Table 5 provides the overview of experimental setup of our computing environment for evaluating the proposed activity recognition framework.

5.3. Evaluation Metrics

Evaluation measures are vital for assessing the performance of the classifiers. Below are the applicable terms used in evaluation metrics for data analysis. The confusion matrix gives better metrics. Some measures that can be determined by the confusion matrix are defined as follows where the F-score is the harmonic mean of Precision and Recall:

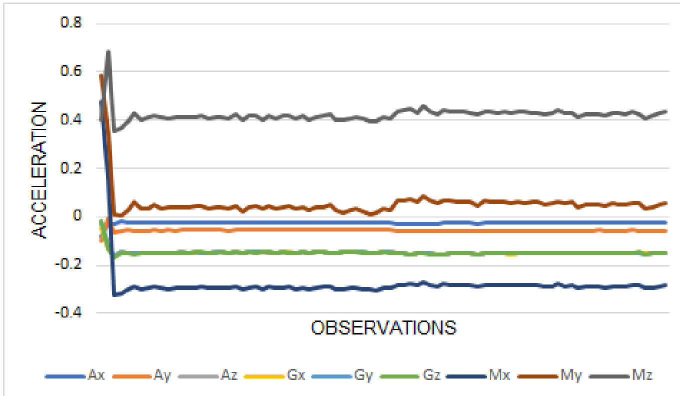

Figure 6: One Vector of Sitting Activity Accelerometer, Gyroscope and Magnetometer Sensor Data

Table 5: Computing Environment

Parameter	Specification
Operating System	Ubuntu 18.04.2 LTS
CPU	Xeon E5/Corei5
RAM	128GB
GPU	NVIDIA GeFroce 1080
CUDA Verion	9.0
Python Version	3.2

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

5.4. Optimized Model

The best model shows the best recognition results for both scenarios: using agent-based analysis (ABA) and without agent-based analysis (WABA) 4.3 of data while using the best value of hyperparameters. Table 6 shows the best parameters selected and used for the recognition task. For better illustration of the results, Fig. 9 shows the transition of improvement in accuracy as the learning rate increased in the case of *Agent-based* Analysis. In the best model of *Agent-based* Analysis, the test recognition rate is 99.43% at maximum while 94% without agent-based analysis, as shown in Fig. 10.

Table 6: Optimized Model Parameter

Parameter	Value
Input vector	32
Embedding Size	751817
Learning Rate	0.001
Weight Decay	0.005
Momentum	0.5-0.99
Batch Size	64
Maximum epochs	50

5.5. General Analysis and Results

The machine learning models perform well; however, they still need some guidance when data is proliferating. On the other hand, deep learning techniques work effectively when there is a lack of domain understanding for feature introspection, as there is less need to worry about feature engineering. It is shown in Fig. 7 that the deep learning model works efficiently when data is large in comparison with the machine learning model. Similarly, in Table 8, it is shown that the proposed deep learning model outperformed existing studies [55, 36, 54, 50] focusing on these activities: standing, still, and walking. Two activities Tilting and in a vehicle, were out of focus in existing studies. If the ”in a vehicle” activity is recognized by other machine learning approaches, it will get confused with sitting activity. For this reason, the Google Fit-based framework is preferred that best predicts this activity based on location and sensor reading. Therefore, we fuse our framework with Google fit to strengthened the recognition process.

The accuracy comparison of the existing methods is shown in Fig. 7. The bar graph represents the mean recognition rate of various methods. SMO achieves the accuracy of 90% using WABA and 96%. C4.5 achieve an accuracy of 97.2% using WABA and 97.9%, Adaboost achieves an accuracy of 92.8% using WABA and 82.7%, NB achieves the accuracy of 91.8% using WABA and 92.2%, MLP achieves the accuracy of 92.2% using WABA and 94.3%, and DRNN achieves the accuracy of 92.8% using ABA and best accuracy of 99.4% using WABA. DRNN takes 780 seconds on a single epoch while other classifiers, such as MLP, take 1029 seconds, SMO takes 203.9 seconds, NB takes 1.2 seconds, c4.5 takes 0.06 seconds, and Adaboost takes 0.42 seconds. However, inference time is below 1 second for each algorithm.

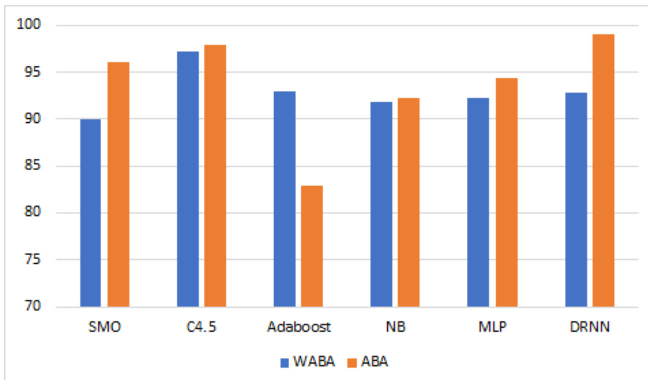

Figure 7: Performance of Best DRNN Compared to other Algorithms. ABA denotes Agent-based Analysis as stated in Section 4.3

5.6. Parameter Tuning and Comprehensive Analysis

We discuss the analysis of parameter tuning of DRNN by using a grid search. For a learning algorithm, a set of optimal hyper-parameters is provided in the grid search. The learning process is well controlled by using these parameters. After three iterations of training, the model started to converge on a particular set of parameters, which later on, we used for final training. The dataset consists of 32 temporal features. These features are ordered in sequences leading toward time series prediction. The parameter used to train our DRNN are shown in Table 2. As discussed, the input vector's size is fixed to 32, leading to a context vector of size 751817. The learning rate is set to 0.001 along with the "Adam" optimizer to train the DRNN. We chose to choose Categorical Cross-entropy as a loss function with our labels as categorical and trained the DRNN for 50 epochs with a batch size of 64. The recognition rate is 99% at the 27^{th} epoch. We use five-fold cross-validations. The transition of recognition rate accuracy in each epoch is presented below in Figure 8.

Specifically, for the *ABA*, the accuracy is about 5% higher than that of the model trained without *ABA* 4.3. Fig. 9 and 10 depicts that varying the learning rate for both cases helps improve the model's result. We got the best accuracy at

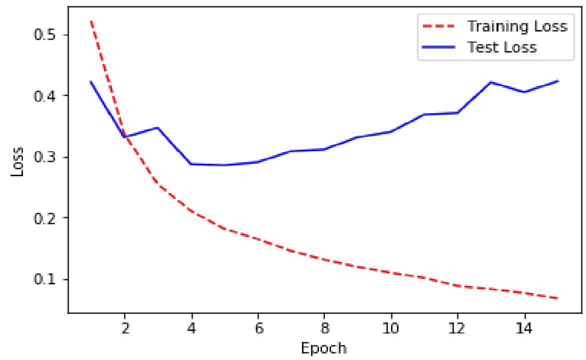

Figure 8: Training and Testing loss for HAR Dataset

learning of 0.001. Epochs counts help to improve recognition accuracy when it is fixed to 50.

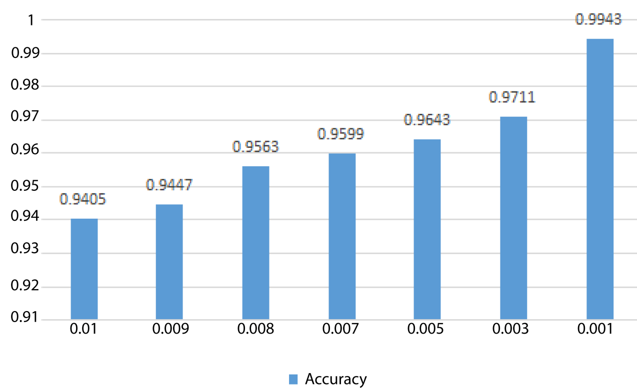

Figure 9: Evaluation with Agent-based Analysis (A.B.A) Feature

Table 7: Evaluation of Daily Life Activity Recognition Using DRNN

Activities	**Precision**	**Recall**	**F-score**	**AUC**
In a vehicle	99.9	97.7	98.9	99.6
On Foot	96.7	78.8	86.8	97.3
Still	97.0	99.8	98.4	99.5
Tilting	99.8	100	99.9	100
Walking	99.7	99.4	99.7	99.8
Average	98.0	98.0	98.0	99.4

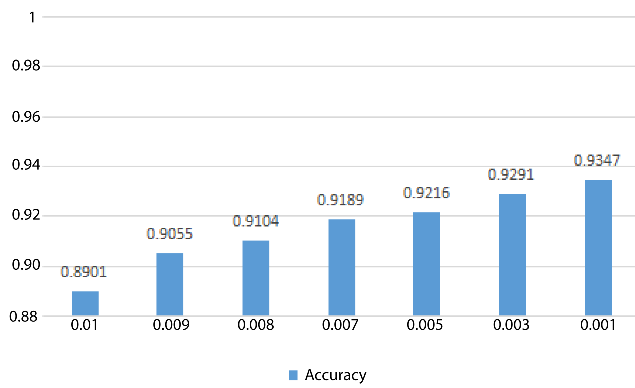

Figure 10: Evaluation without Agent Based Analysis (A.B.A) Feature

In Table 7, the evaluation of the proposed framework is presented based on daily life activity recognition in terms of performance metrics. The system recognizes activities that include activities in a vehicle, on foot, while being still, tilting, and walking. The evaluation criteria include Precision, Recall, F-score, and Area Under the Curve (AUC). As seen in Table 7, the model shows a graceful performance in recognizing the activities with an average Precision, Recall, and F-score of 98% and an average AUC of 99.4%. Fig. 11 illustrates the confusion matrix, demonstrating the percentage of correctly classified activities concerning misclassified activities. There is too slight confusion between tilting and still activities. Moreover, other activities are classified efficiently.

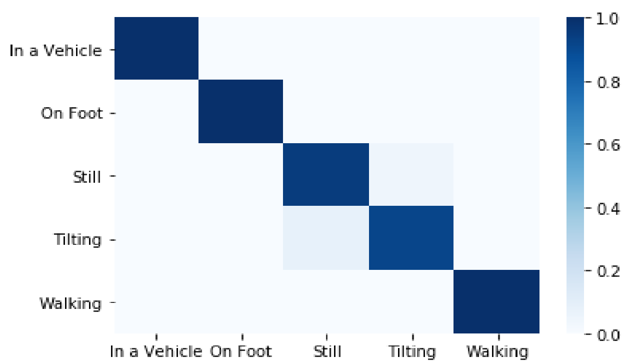

Figure 11: Confusion Matrix of the Proposed Activity Recognition Task

Fig. 12 and 13 illustrates weights and biases of a given layer. Fig. 12 shows the model's maximum and mean bias on the training and validation set of layer

1. Fig. 13 illustrates the maximum, minimum, mean, and standard deviation bias of the model on the training and validation set of layer 2. Fig. 14 and 15 show the gradient value on which the model is performing best on the number of iterations. The gradient norm is an important parameter to assess the weights of the renal network. It indicates that the weight is being updated positively. A too low gradient value can take toward vanishing gradient, or a too-high gradient can take towards exploding gradient. The gradient updates are also diminishing over time and approaching zero on both layers. Fig. 15 depicts that the best learning is near 0, as our best learning parameter 6 shows the same best parameter, which is 0.001.

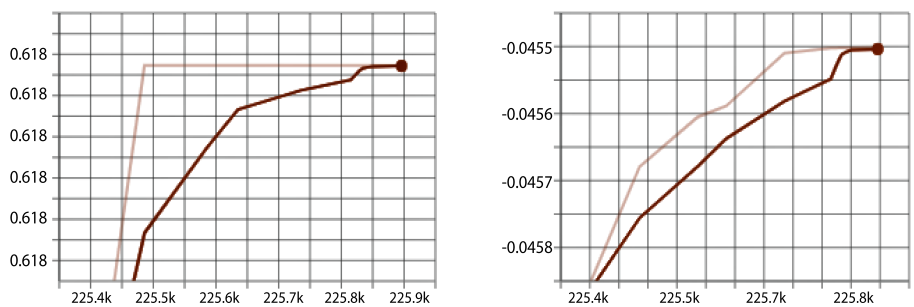

Figure 12: Maximum and Minimum Bias of Layer 1 on Training and Validation Set using Tensor Board

Table 8 presents a comparison of the accuracy metric with existing state-of-the-art studies while recognizing the activities. The proposed framework achieves a detection accuracy of 96.7% while standing on foot, 97.0% while Sitting still, and 99.7% while walking. Our model shows a significant gain of 1% while recognizing Standing (On foot), 2% while recognizing Sitting (Still) activity, and 1.2% while recognizing walking activity. Thus, a significant gain in accuracy over previous studies shows the proficiency of the proposed framework. Furthermore, our framework recognizes two novel activities: ”In a Vehicle” and ”Tilting,” which focus on existing studies. We achieve an accuracy of 99.9% for ”in a vehicle” activity and 99.8% for ”tilting” activity.

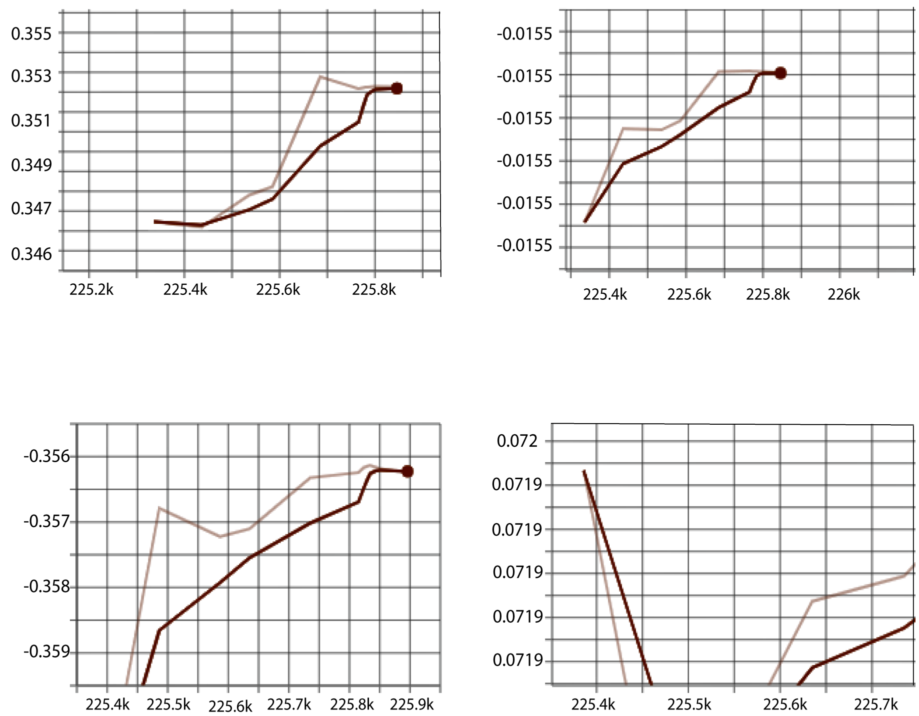

Figure 13: Maximum, Minimum, Mean and Standard Deviation Bias at Layer 2 on Training and Validation Set using Tensor Board

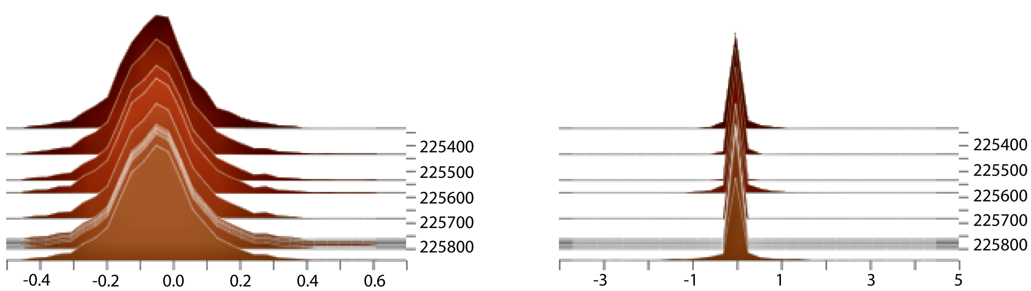

Figure 14: Gradient Norm of Layer 1 on Training and Validation Set using Tensor Board

Table 9 presents a comparison of the Precision, Recall, and F-score metrics with existing state-of-the-art studies for recognizing the activities. The proposed framework achieved the highest Precision of 96.7 for standing, F-score of 86.8, and Recall of 78.8 while achieved the highest Precision of 97.0 for standing, F-score of 98.4, and Recall of 99.8 and achieved the highest Precision of 99.7 for

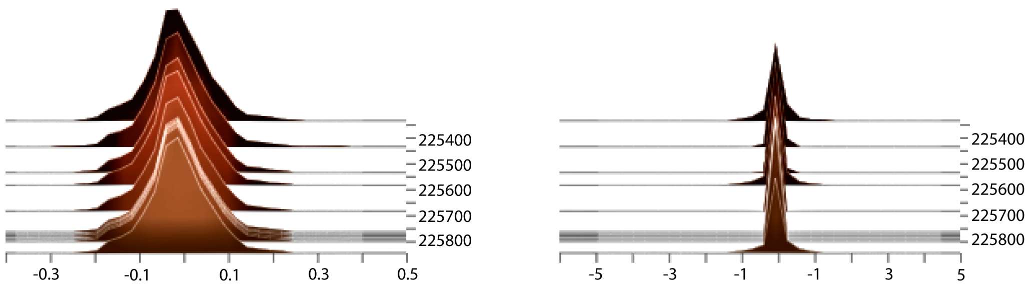

Figure 15: Gradient Norm of Layer 2 on Training and Validation Set using Tensor Board

Table 8: Comparison of the Proposed framework with Existing Studies (Accuracy)

Activities Studies	Standing (On Foot)	Sitting (Still)	Walking
[55]	93.3	82.6	98.5
[36]	95.7	94.0	96.5
[54]	94.9	93.9	86.3
[50]	91.9	95.0	91.7
Proposed framework	96.7	97.0	99.7

standing, F-score of 99.7 and Recall of 99.4. The results show that the proposed framework efficiently improves the recognition rate in comparison with state-of-the-art studies.

Table 9: Comparison of the Proposed framework with Existing Studies using Other Performance Measures

Studies	**Precision**	**Recall**	**F-score**
Standing (On Foot) [55]	-	-	90.7
Sitting (Still) [55]	-	-	93.6
Walking [55]	-	-	99.6
Standing (On Foot) UCI Dataset [54]	75.1	65.4	-
Sitting (Still) UCI Dataset [54]	46.8	56.7	-
Walking [54]	88.9	95.9	-
Standing (On Foot) WISDM Dataset [54]	84.0	88.1	-
Sitting (Still) WISDM Dataset [54]	89.0	90.9	-
Walking WISDM Dataset [54]	91.7	87.9	-
Standing (On Foot) **Proposed framework**	96.7	78.8	86.8
Sitting (Still) **Proposed framework**	97.0	99.8	98.4
Walking **Proposed framework**	99.7	99.4	99.7

6. Conclusion and Future Work

This paper investigated that HAR using an accelerometer, gyroscope, and magnetometer fused with Google Fit API improve activity recognition. We evalaute our framework for both scenarios: using ABA and WABA. Our evaluation shows that our framework works better when used with fused and with Google Fit API. We applied a DRNN for HAR using raw hardware sensors and Google Fit readings. The obtained recognition accuracy is 99% against the test dataset when trained on the dataset provided by Agent-based Analysis and 95% when trained without Agent-based Analysis. We provided a thorough analysis of the combination of hyper-parameters of the DRNN. We provided the best combination of various parameters to classify HAR correctly. We believe that our framework is significant for activity recognition. In the future, we intend to analyze more routine life activities and apply our proposed framework to data gathered from a more significant number of participants belonging to different age groups. Also, we plan to compare the proposed framework with some meta-heuristic algorithms to evaluate feature selection strategies.

[1] U. Nations, Worlds population increasingly urban with more than half living in urban areas, United Nations.

[2] E. Ugljanin, E. Kajan, Z. Maamar, M. Asim, V. Burégio, Immersing citizens

and things into smart cities: a social machine-based and data artifact-driven approach, Computing (2020) 1–20.

[3] B. N. Silva, M. Khan, K. Han, Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities, Sustainable Cities and Society 38 (2018) 697–713.

[4] G. Lilis, G. Conus, N. Asadi, M. Kayal, Towards the next generation of intelligent building: An assessment study of current automation and future iot based systems with a proposal for transitional design, Sustainable cities and society 28 (2017) 473–481.

[5] H. Zahmatkesh, F. Al-Turjman, Fog computing for sustainable smart cities in the iot era: Caching techniques and enabling technologies-an overview, Sustainable Cities and Society 59 (2020) 102–139.

[6] M. M. Aborokbah, S. Al-Mutairi, A. K. Sangaiah, O. W. Samuel, Adaptive context aware decision computing paradigm for intensive health care delivery in smart citiesa case analysis, Sustainable cities and society 41 (2018) 919–924.

[7] M. Al-Khafajiy, L. Webster, T. Baker, A. Waraich, Towards fog driven iot healthcare: challenges and framework of fog computing in healthcare, in: Proceedings of the 2nd International Conference on Future Networks and Distributed Systems, ACM, 2018, pp. 9:1–9:7.

[8] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, I. De Munari, Iot wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment, IEEE Internet of Things Journal 6 (5) (2019) 8553–8562.

[9] M. U. Sarwar, A. R. Javed, Collaborative health care plan through crowdsource data using ambient application, in: 2019 22nd International Multitopic Conference (INMIC), IEEE, 2019, pp. 1–6.

[10] O. Habimana, Y. Li, R. Li, X. Gu, W. Yan, Attentive convolutional gated recurrent network: a contextual model to sentiment analysis, International Journal of Machine Learning and Cybernetics.

[11] S. Liu, T. Li, H. Ding, B. Tang, X. Wang, Q. Chen, J. Yan, Y. Zhou, A hybrid method of recurrent neural network and graph neural network for next-period prescription prediction, International Journal of Machine Learning and Cybernetics.

[12] H. Naveed, G. Khan, A. U. Khan, A. Siddiqi, M. U. G. Khan, Human activity recognition using mixture of heterogeneous features and sequential minimal optimization, International Journal of Machine Learning and Cybernetics 10 (9) (2019) 2329–2340.

[13] A. R. Javed, M. U. Sarwar, M. O. Beg, M. Asim, T. Baker, H. Tawfik, A collaborative healthcare framework for shared healthcare plan with ambient intelligence, Human-centric Computing and Information Sciences 10 (1) (2020) 1–21.

[14] H. Ghasemzadeh, R. Jafari, Physical movement monitoring using body sensor networks: A phonological approach to construct spatial decision trees, IEEE Transactions on Industrial Informatics 7 (1) (2010) 66–77.

[15] A. R. Javed, L. G. Fahad, A. A. Farhan, S. Abbas, G. Srivastava, R. M. Parizi, M. S. Khan, Automated cognitive health assessment in smart homes using machine learning, Sustainable Cities and Society (2020) 102572.

[16] S. U. Rehman, A. R. Javed, M. U. Khan, M. Nazar Awan, A. Farukh, A. Hussien, Personalisedcomfort: a personalised thermal comfort model to predict thermal sensation votes for smart building residents, Enterprise Information Systems (2020) 1–23.

[17] M. Usman Sarwar, A. Rehman Javed, F. Kulsoom, S. Khan, U. Tariq, A. Kashif Bashir, Parciv: Recognizing physical activities having complex interclass variations using semantic data of smartphone, Software: Practice and Experience.

[18] N. Alshurafa, W. Xu, J. J. Liu, M.-C. Huang, B. Mortazavi, C. K. Roberts, M. Sarrafzadeh, Designing a robust activity recognition framework for health and exergaming using wearable sensors, IEEE Journal of Biomedical and Health Informatics 18 (5) (2013) 1636–1646.

[19] T. Wang, Z. Wang, D. Zhang, T. Gu, H. Ni, J. Jia, X. Zhou, J. Lv, Recognizing parkinsonian gait pattern by exploiting fine-grained movement function

features, ACM Transactions on Intelligent Systems and Technology (TIST) 8 (1) (2016) 1–22.

[20] H. Arshad, M. A. Khan, M. Sharif, M. Yasmin, M. Y. Javed, Multi-level features fusion and selection for human gait recognition: an optimized framework of bayesian model and binomial distribution, International Journal of Machine Learning and Cybernetics 10 (12) (2019) 3601–3618.

[21] P. N. Dawadi, D. J. Cook, M. Schmitter-Edgecombe, Automated cognitive health assessment using smart home monitoring of complex tasks, IEEE transactions on systems, man, and cybernetics: systems 43 (6) (2013) 1302–1313.

[22] M. Anisetti, C. Ardagna, V. Bellandi, M. Cremonini, F. Frati, E. Damiani, Privacy-aware big data analytics as a service for public health policies in smart cities, Sustainable cities and society 39 (2018) 68–77.

[23] C. Tran, M. M. Trivedi, 3-d posture and gesture recognition for interactivity in smart spaces, IEEE Transactions on Industrial Informatics 8 (1) (2011) 178–187.

[24] Y. Gu, F. Ren, J. Li, Paws: Passive human activity recognition based on wifi ambient signals, IEEE Internet of Things Journal 3 (5) (2015) 796–805.

[25] D. Tao, Y. Wen, R. Hong, Multicolumn bidirectional long short-term memory for mobile devices-based human activity recognition, IEEE Internet of Things Journal 3 (6) (2016) 1124–1134.

[26] P. Menaspà, Effortless activity tracking with google fit, Br J Sports Med 49 (24) (2015) 1598–1598.

[27] O. D. Lara, M. A. Labrador, A survey on human activity recognition using wearable sensors, IEEE communications surveys & tutorials 15 (3) (2012) 1192–1209.

[28] A. Mannini, A. M. Sabatini, Machine learning methods for classifying human physical activity from on-body accelerometers, Sensors 10 (2) (2010) 1154–1175.

[29] R. Poppe, A survey on vision-based human action recognition, Image and vision computing 28 (6) (2010) 976–990.

[30] A. R. Javed, M. U. Sarwar, S. ur Rehman, H. U. Khan, Y. D. Al-Otaibi, W. S. Alnumay, Pp-spa: Privacy preserved smartphone-based personal assistant to improve routine life functioning of cognitive impaired individuals, Neural Processing Letters (2021) 1–18.

[31] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, P. J. Havinga, A survey of online activity recognition using mobile phones, Sensors 15 (1) (2015) 2059–2085.

[32] M. Cornacchia, K. Ozcan, Y. Zheng, S. Velipasalar, A survey on activity detection and classification using wearable sensors, IEEE Sensors Journal 17 (2) (2016) 386–403.

[33] A. R. Javed, M. O. Beg, M. Asim, T. Baker, A. H. Al-Bayatti, Alphalogger: detecting motion-based side-channel attack using smartphone keystrokes, Journal of Ambient Intelligence and Humanized Computing.

[34] T. Plötz, N. Y. Hammerla, P. L. Olivier, Feature learning for activity recognition in ubiquitous computing, in: Twenty-second international joint conference on artificial intelligence, 2011.

[35] S. Bhattacharya, P. Nurmi, N. Hammerla, T. Plötz, Using unlabeled data in a sparse-coding framework for human activity recognition, Pervasive and Mobile Computing 15 (2014) 242–262.

[36] A. R. Javed, M. U. Sarwar, S. Khan, C. Iwendi, M. Mittal, N. Kumar, Analyzing the effectiveness and contribution of each axis of tri-axial accelerometer sensor for accurate activity recognition, Sensors 20 (8) (2020) 2216.

[37] M. Al-Khafajiy, T. Baker, A. Waraich, D. Al-Jumeily, A. Hussain, Iot-fog optimal workload via fog offloading, in: 2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion, UCC Companion 2018, Zurich, Switzerland, IEEE, 2018, pp. 359–364.

[38] M. Mapar, M. J. Jafari, N. Mansouri, R. Arjmandi, R. Azizinezhad, T. B. Ramos, A composite index for sustainability assessment of health, safety and environmental performance in municipalities of megacities, Sustainable Cities and Society (2020) 102164.

[39] A. Papa, M. Mital, P. Pisano, M. Del Giudice, E-health and wellbeing monitoring using smart healthcare devices: An empirical investigation, Technological Forecasting and Social Change 153 (2020) 119226.

[40] J. Wang, Y. Chen, S. Hao, X. Peng, L. Hu, Deep learning for sensor-based activity recognition: A survey, Pattern Recognition Letters 119 (2019) 3–11.

[41] Z. Qin, Y. Zhang, S. Meng, Z. Qin, K.-K. R. Choo, Imaging and fusing time series for wearable sensor-based human activity recognition, Information Fusion 53 (2020) 80–87.

[42] M. Dighriri, A. S. D. Alfoudi, G. M. Lee, T. Baker, R. Pereira, Comparison data traffic scheduling techniques for classifying qos over 5g mobile networks, in: 2017 31st international conference on advanced information networking and applications workshops (WAINA), IEEE, 2017, pp. 492–497.

[43] D. M. West, How 5g technology enables the health internet of things, Brookings Center for Technology Innovation 3 (2016) 1–20.

[44] S. Latif, J. Qadir, S. Farooq, M. A. Imran, How 5g wireless (and concomitant technologies) will revolutionize healthcare?, Future Internet 9 (4) (2017) 93.

[45] M. Chen, J. Yang, J. Zhou, Y. Hao, J. Zhang, C.-H. Youn, 5g-smart diabetes: Toward personalized diabetes diagnosis with healthcare big data clouds, IEEE Communications Magazine 56 (4) (2018) 16–23.

[46] J. Lloret, L. Parra, M. Taha, J. Tomás, An architecture and protocol for smart continuous ehealth monitoring using 5g, Computer Networks 129 (2017) 340–351.

[47] E. M. Tapia, S. S. Intille, K. Larson, Activity recognition in the home using simple and ubiquitous sensors, in: International conference on pervasive computing, Springer, 2004, pp. 158–175.

[48] D. J. Cook, Learning setting-generalized activity models for smart spaces, IEEE intelligent systems 2010 (99) (2010) 1.

[49] L. Bao, S. S. Intille, Activity recognition from user-annotated acceleration data, in: International conference on pervasive computing, Springer, 2004, pp. 1–17.

[50] J. R. Kwapisz, G. M. Weiss, S. A. Moore, Activity recognition using cell phone accelerometers, ACM SigKDD Explorations Newsletter 12 (2) (2011) 74–82.

[51] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, P. J. Havinga, Fusion of smartphone motion sensors for physical activity recognition, Sensors 14 (6) (2014) 10146–10176.

[52] Y. LeCun, Y. Bengio, et al., Convolutional networks for images, speech, and time series, The handbook of brain theory and neural networks 3361 (10) (1995) 1995.

[53] S. Duffner, S. Berlemont, G. Lefebvre, C. Garcia, 3d gesture classification with convolutional neural networks, in: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2014, pp. 5432–5436.

[54] B. Kolosnjaji, C. Eckert, Neural network-based user-independent physical activity recognition for mobile devices, in: International Conference on Intelligent Data Engineering and Automated Learning, Springer, 2015, pp. 378–386.

[55] A. Ignatov, Real-time human activity recognition from accelerometer data using convolutional neural networks, Applied Soft Computing 62 (2018) 915–922.

[56] M. Inoue, S. Inoue, T. Nishida, Deep recurrent neural network for mobile human activity recognition with high throughput, Artificial Life and Robotics 23 (2) (2018) 173–185.

[57] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.

[58] G. H. B. P. P. Reutemann, I. H. W. M. Hall, E. Frank, I. H. Witten, The weka data mining software: An update, SIGKDD Explorations 11 (1) (2009) 10–18.

[59] A. Krause, M. Ihmig, E. Rankin, D. Leong, S. Gupta, D. Siewiorek, A. Smailagic, M. Deisher, U. Sengupta, Trading off prediction accuracy and power consumption for context-aware wearable computing, in: Ninth IEEE International Symposium on Wearable Computers (ISWC'05), IEEE, 2005, pp. 20–26.