

Received 20 August 2022, accepted 8 September 2022, date of publication 15 September 2022,
date of current version 27 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3206952

RESEARCH ARTICLE

Fully Connected Generative Adversarial Network for Human Activity Recognition

ALI OLOW JIMALE^{1,2}, (Member, IEEE), AND MOHD HALIM MOHD NOOR^{ID}¹

¹School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang 11800, Malaysia

²Faculty of Computing, SIMAD University, Mogadishu 252, Somalia

Corresponding author: Mohd Halim Mohd Noor (halimnoor@usm.my)

This work was supported in part by the Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme under Project FRGS/1/2019/ICT02/USM/02/1.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Human Research Ethics Committee of Universiti Sains Malaysia under Application No. USM/JEPeM/18040205.

ABSTRACT Conditional Generative Adversarial Networks (CGAN) have shown great promise in generating synthetic data for sensor-based activity recognition. However, one key issue concerning existing CGAN is the design of the network architecture that affects sample quality. This study proposes an effective CGAN architecture that synthesizes higher quality samples than state-of-the-art CGAN architectures. This is achieved by combining convolutional layers with multiple fully connected networks in the generator's input and discriminator's output of the CGAN. We show the effectiveness of the proposed approach using elderly data for sensor-based activity recognition. Visual evaluation, similarity measure, and usability evaluation are used to assess the quality of generated samples by the proposed approach and validate its performance in activity recognition. In comparison to the state-of-the-art CGAN, the visual evaluation and similarity measure demonstrate that the proposed models' synthetic data more accurately represents actual data and creates more variations in each synthetic data than the state-of-the-art approach respectively. The experimental stages of the usability evaluation, on the other hand, show a performance gain of 2.5%, 2.5%, 3.1%, and 4.4% over the state-of-the-art CGAN when using synthetic samples by the proposed architecture.

INDEX TERMS Activity recognition, deep learning, generative adversarial network.

I. INTRODUCTION

Pervasive computing and sensing technologies have advanced dramatically, enabling automatic analysis and recognition of human behavior and activities [1]. One of the most important applications of this topic is Sensor-based Activity Recognition, a research field that recognizes human activities by analyzing motion data collected via fixed or wearable sensors [2]. In the first, sensing technologies are either tagged to a certain location and human activity inference is based on the user's interaction with the tagged object, or they are deployed in an environment where no tag or device is required. Passive infrared sensors, pressure sensors, and contact switches are examples of fixed sensors.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang ^{ID}.

In the second, the sensing technologies are worn by users or attached to portable devices such as mobile phones and smartwatches. Accelerometers and gyroscopes are examples of wearable sensors. Wearable sensors are ubiquitous, unobtrusive, cheaper, less harmful, easier to deploy and use, and capable to support real-time activity recognition compared to other sensing modalities.

Because of these advantages, several machine learning and deep learning methods have been explored to classify and recognize human activities using wearable sensors such as accelerometers, and gyroscopes. Machine learning methods for sensor-based activity recognition use hand-crafted features that are manually extracted with the help of human domain experts [3], [4]. However, expert-driven feature extraction methods have issues [5]. First of all, domain experts can only learn very limited features [6] related to

some statistical information including mean, frequency, variance, and amplitude which cannot fully support the dynamic nature of today's ubiquitous and seamless collection of wearable and mobile sensor streams [7]. These shallow features also fail to support modeling complex activities [8] and involve very time-consuming feature selections [9]. Second, manually engineered features are error-prone which may result in the loss of important information for activity recognition [10]. This affects the performance and accuracy of the human activity recognition system [5]. Third, the current manual feature extraction is application-dependent or problem-specific that cannot be transferred to another activity with similar patterns. Finally, there is no universal rule for selecting appropriate human activity features.

To overcome the limitation of feature engineering, deep learning methods have been widely adopted for sensor-based activity recognition to automate feature extraction and extract higher-level representation to recognize human activities [4], [11]. Although deep learning methods have been shown to be effective, the methods require a huge amount of training data which is challenging to collect in sensor-based activity recognition due to the cost and the time required to collect the training data as well as battery and storage capacity constraints of sensing technologies. For this reason, researchers address the scarcity of training data in sensor-based activity recognition through data augmentation. Several studies used data augmentation to increase the size of sensory data [12], [13], [14], [15], [16]. There are two types of sensor data augmentation for activity recognition: basic data augmentation and deep learning data augmentation methods. Basic data augmentation techniques use conventional algorithms to perform different data augmentation techniques that add noise to the sensor readings, increase/decrease the magnitude of the sensor readings, or flip the sign of the original sensor data [17]. One drawback of the basic data augmentation method is generating only limited samples that are not suitable for training deep learning models [18].

Recently, deep generative models have shown impressive results in generating massive samples for sensor-based human activity recognition [19]. This is to say that applying data augmentation on encoded inputs rather than raw sensory data generates more plausible synthetic data due to the manifold unfolding in feature space [20]. Two techniques can be used to apply data augmentation using deep generative models. These two techniques are Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs).

VAEs are generative models that learn the low-dimensional representation of sensor data points. It consists of two networks (encoder, decoder) and a loss function. The encoder converts input data to small latent (hidden) space while the decoder maps latent space input into original input. The loss function is the negative log-likelihood with a regularizer to penalize the decoder mistakes [21]. VAEs data augmentation techniques are applied in several areas including speech recognition [21], traffic estimation [22], text generation [23], adverse drug reactions detection [24], and others. However,

VAE is not well known in synthetic sensor data generation for human activity recognition. This is because synthetic data generated by VAEs tend to be more blurred.

GANs generate more realistic synthetic data than VAEs in the sensor-based activity recognition field [25]. Due to its capacity to create verisimilar synthetic examples, GAN has become the most prominent data generative model for overcoming the lack of data challenges for sensor-based activity recognition. One of the recent achievements of GAN network architectures for sensor-based activity recognition is developing the Unified Conditional GAN (CGAN) for synthesizing more than one human activity data in a single training process [26], [27]. However, state-of-the-art CGAN eliminates the use of fully connected layers in their model's generator and discriminator. This architectural choice suggested by [14] is also adopted by the majority of other existing GANs without investigating its effect on sample quality.

The main objective of this study is to develop an improved CGAN architecture that combines convolutional layers with multiple fully connected networks in the input and output layers of the generator and discriminator to generate more realistic synthetic activity signals.

This study's contributions are summarized as follows:

- a. To our best knowledge, we are the first to propose an enhanced CGAN architecture that combines convolutional layers with multiple fully connected networks in the input and output layers of the generator and discriminator respectively to generate more quality synthetic samples for sensor-based activity recognition.
- b. We conduct comprehensive experiments to compare the quality of the generated samples by the proposed approach with the state-of-the-art approach using visual and similarity measure evaluation techniques. The visual evaluation and similarity measure techniques demonstrate that the proposed models' synthetic data more accurately captures the real data and creates more variations than the state-of-the-art approach.
- c. The performance of the proposed approach is trained on elderly datasets for sensor-based activity recognition. Using synthetic samples, the proposed architecture outperforms the state-of-the-art CGAN by 2.5%, 2.5%, 3.1%, and 4.4%.

The rest of this paper is organized as follows. Section 2 reviews the related work, Section 3 details the proposed architecture, Section 4 explains the experimental setup, the performance of the proposed architecture is evaluated in section 5, and Section 6 concludes the study.

II. RELATED WORK

Activity Recognition refers to the process of identifying human movements and actions based on motion data collected through digital cameras and sensor devices [28], [29], [30], [31]. Human Activities recognized by activity recognition systems can be classified into two main categories: transitional and basic human activities [32]. Transitional physical

activities are simple events with a small duration in the order of seconds. These activities are further divided into two subcategories: gesture and transition. Gesture refers to the visual movements of a part of the human body such as the arm, hand, head, and finger to communicate nonverbally [33], while transitions are the activities that connect two different human activities such as lie-to-sit, lie-to-stand, stand-to-sit, stand-to-lie, sit-to-lie, and sit-to-stand [32]. On the other hand, basic activities are human activities with a long duration in the order of minutes. These activities can be characterized as either dynamic activity or static activity. Dynamic activities are continuous activities with periodicity (i.e., walking, running), while static activities are activities with static postures (i.e., sitting, standing). Existing activity recognition approaches can be divided into two categories: vision-based activity recognition and sensor-based activity recognition [34], [35]. The vision-based activity recognition approach analyzes digital images and/or video sequences with human motions from cameras to recognize human activities [19]. Initially, this approach has been a hot scientific topic. Many researchers investigated human activity recognition from images and videos [29]. This is due to its wide applications in sports, surveillance systems, health care, smart rooms, video retrieval, and human-computer interfaces [36]. Later, the sensor-based activity recognition approach became a popular and fast-growing topic [37]. This is due to technological advancements and low prices of sensor devices as well as the issues of the vision-based approach including privacy, space, cost, angle, obstruction, and light dependency issues [12], [38].

Sensor-based activity recognition has achieved good progress by utilizing various deep learning classifiers such as Convolutional Neural Network, Restricted Boltzmann Machine Restricted, Deep Autoencoder, Sparse coding, and Recurrent neural network [4]. This is due to the automatic feature extraction that deep learning methods employ to automatically generate human activity features and select the best ones [39]. The deep learning models are trained with huge amounts of labeled data which is challenging in sensor-based activity recognition for various reasons. First, data collection and dataset creation with hundreds of subjects with different age groups are expensive and time-consuming with great effort requirements [40]. Collecting high-quality data requires a huge number of participants, logistics, costs, experimenters, equipment, sensor modalities, and long-term recordings. Sensors also have storage and battery limitations. Data annotation is another issue. Manual annotation of the collected data with activity labels is expensive, physically laborious, and error-prone. Annotating these activities may also dictate requiring domain experts for annotation purposes. One of the best methods for overcoming limited datasets in activity recognition using deep learning is data augmentation [41]. GAN, originally proposed by [42], has dominated data augmentation methods for sensor-based activity recognition using deep learning as it generates more realistic synthetic data than other data augmentation methods [25].

It contains two components built by multilayer perceptron: generator and discriminator. The generator takes a noise vector, which is randomly generated via a-priori distribution, as an input to generate fake samples. It maximizes the probability of the fake samples being classified as real. The real and fake samples are then fed into the discriminator to estimate the probability that the fake data is drawn using the real data [43]. Equation (1) contains the overall objective function of GAN where G is the generator, D is the discriminator, z is an a-priori distribution noise, $p_z(z)$ is fake data distribution, and $p_{data}(x)$ is real data distribution.

$$\min_G \max_D = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} + [\log(1 - D(G(z)))] \quad (1)$$

GANs for sensor-based activity recognition fall into two categories: semi-supervised and supervised GANs. Semi-supervised GANs were mainly used to overcome left-out user's activity recognition, which may have declined the recognition performance of the learning model [44]. Any human subject whose data is not fed to the deep learning model for training is a left-out user. A limitation of semi-supervised GANs is that they require the use of test data during the training phase. In addition to that, a different model must be trained for a single or group of target subjects [43]. However, the focus of this study is to improve sample quality of state-of-the-art GANs that generate synthetic data for sensor-based activity recognition in a single model training process without using any data from the test set during the training. This could be achieved using supervised GANs. The first supervised GAN for sensor-based activity recognition using deep learning was proposed by Wang *et al.* [45]. They developed a GAN framework called SensoryGAN that contains a generator and a discriminator. First, their method takes random noise and real sensor data as input. Then, the generator and the discriminator play a mini-max game to generate synthetic sensor data for three human activities: stay, walk, and jog. They also applied three visualization techniques: local, global, and memory independent to satisfy the GANs community from computer vision while evaluating the quality of generated data.

Nevertheless, the limitation of this study is that it can generate data for a single class of activity but not accommodate various classes of different human activities in a single training process. This is time-consuming and makes the learning process long. Shi *et al.* [46] also implemented a DCGAN-based data augmentation method called HARAUGAN to enlarge the activity scope of the SensoryGAN method. Although they consider more than three activities in their method, their method is not unified as they generate the sensory data for each activity separately. Hong *et al.* [26] solved this issue by developing a unified model for generating synthetic data of 5 human activities (standing, laying down, walking, cycling, and jogging) in a single training process. This is achieved by adapting CGAN architecture that adds a conditional factor with class label information to the

TABLE 1. Architecture of the unified CGAN's generator [26].

Layer (type)	Configuration Details	Output Shape
1D Convolution	Filters = 16, Kernel Size = 7, Activation = ReLU, Input Shape = 112, 4	100, 16
Dropout	Rate = 0.4	100, 16
1D Convolution	Filter = 32, Kernel Size = 5, Activation = ReLU	100, 32
Dropout	Rate = 0.4	100, 32
1D Convolution	Filter = 64, Kernel Size = 3, Activation = ReLU	100, 64
Dropout	Rate = 0.4	100, 64
1D Convolution	Filter = 128, Kernel Size = 1, Activation = ReLU	100, 128
Dropout	Rate = 0.4	100, 128
LSTM	Units = 200, Activation = Tanh	100, 200
LSTM	Units = 200, Activation = Tanh	100, 200
Dense	Units = 150, Activation = ReLU	100, 150
Dense	Units = 3, Activation = Linear	100, 6

TABLE 2. Architecture of unified CGAN'S discriminator [26].

Layer (type)	Configuration	Output Shape
1D Convolution	Filters = 64, Kernel Size = 7, Activation = ReLU, Input Shape = 100,6	100, 64
Dropout	Rate = 0.4	100, 64
1D Convolution	Filters = 128, Kernel Size = 3, Activation = Leaky ReLU	100, 128
Dropout	Rate = 0.4	100, 128
Max Pooling	Pool Size=2	50, 128
Flatten	-	6400
Dense	Units = 100, Activation = ReLU	100
Dense	Units = 1, Activation = Sigmoid	1

generator and discriminator models. The conditional factor c here is the label of the human activities and real data input is sensor-based human activity signals. The network architecture of this model restricts the use of fully connected networks in the input layer of its generator, $G(z|c)$, and the output layer of its discriminator, $D(x|c)$. Four 1-D CNN layers with rectified linear unit (ReLU) activation function followed by dropout, two LSTM layers with Tanh activation and two fully connected layers with rectified linear unit (ReLU) activation function make up its generator while two 1-D CNN layers with dropout each and two fully connected layers with a rectified linear unit (ReLU) and sigmoid activation functions make up its discriminator. Table 1 and Table 2 show the architecture of the Generator and Discriminator respectively.

Li *et al.* [27] have also proposed a unified GAN model named ActivityGan for generating sensor-based activity

recognition. However, their model eliminates the use of fully connected from the GAN network. Its network architecture consists of a 1D-convolution chain and 1D- transposed convolution chain while CNN architecture is adopted to build their discriminator.

Literature confirms that state-of-the-art Unified GAN networks for generating sensor-based activity recognition data have adopted models that eliminate or restrict the use of fully connected layers from the GAN network architecture. This hurts the output quality of the GAN network [47]. This study proposes an enhanced CGAN architecture that combines convolutional layers with multiple fully connected networks in the input and output layers of the generator and discriminator respectively to generate better synthetic activity signals as explained in the next section.

III. PROPOSED MODEL

The proposed method in this study, the Fully Connected CGAN (FCGAN) Model, enhances the Unified CGAN model architecture by Hong *et al.* [26] to improve sample generation quality. Unlike the state-of-the-art Unified GANs, the Fully Connected CGAN converts the low-dimensional fake input of the generator to a high-dimensional space of the activity signals. This is achieved by employing three fully connected layers as embedding layers for the first task of the generator, followed by convolutional and LSTM layers for the second and third task respectively. The fully connected layers employed as the first task of the generator learn the relationship between noise vectors and human activity features by mapping noise vectors to activity features of the human activities. This allows the model to generate subtle variations in different spatial zones, which helps the generator to synthesizing more realistic samples. Table 3 shows the generator's architecture of the fully connected CGAN.

This study also enhances the discriminator's network architecture of the base model by adding three fully connected networks to the discriminator's network. This provides the functionality to convert the discriminator's input to a lower-dimensional space before classification, preventing the discriminator's loss from becoming too low. In this regard, the generator learns faster, leading to a faster model convergence. Table 4 shows the discriminator's architecture of the fully connected CGAN.

The used fully connected networks in the FCGAN generator's and discriminator's architecture comprise three fully connected layers with almost similar configurations to the base model as a recent study show impressive results in having three fully connected layers with convolution layers in GANs generator and discriminator architecture [47].

IV. EXPERIMENTAL SETUP

In this research, a sensory dataset collected from elderly subjects using an accelerometer and gyroscope is adopted to train the generative and classification models in this study. The subjects of the dataset were asked to perform walking, standing, sitting, lying down, sit-to-stand, lie-to-sit, and

TABLE 3. Architecture of the FCGAN'S generator.

Layer (type)	Configuration Details	Output Shape
Input	Latent Dim = 100, Number of Classes = 8	108
Dense	Units = 64, Activation = ReLU	64
Dense	Units = 512, Activation = ReLU	512
Dense	Units = 3600, Activation = ReLU	3600
Reshape		100, 36
1D Convolution	Filters = 16, Kernel Size = 7, Activation = ReLU	100, 16
Dropout	Rate = 0.4	100, 16
1D Convolution	Filter = 32, Kernel Size = 5, Activation = ReLU	100, 32
Dropout	Rate = 0.4	100, 32
1D Convolution	Filter = 64, Kernel Size = 3, Activation = ReLU	100, 64
Dropout	Rate = 0.4	100, 64
1D Convolution	Filter = 128, Kernel Size = 1, Activation = ReLU	100, 128
Dropout	Rate = 0.4	100, 128
LSTM	Units = 200, Activation = Tanh	100, 200
LSTM	Units = 200, Activation = Tanh	100, 200
Dense	Units = 150, Activation = ReLU	100, 150
Dense	Units = 3, Activation = Linear	100, 6

TABLE 4. Architecture of the FCGAN'S discriminator.

Layer (type)	Configuration	Output Shape
1D Convolution	Filters = 64, Kernel Size=7, Activation = ReLU, Input Shape = 100,6	100, 64
Dropout	Rate=0.4	100, 64
1D Convolution	Filters = 128, Kernel Size = 3, Activation = Leaky ReLU	100, 128
Dropout	Rate=0.4	100, 128
Flatten	-	12800
Dense	Units = 512, Activation = ReLU	512
Dense	Units = 64, Activation = ReLU	64
Dense	Units = 16, Activation = ReLU	16
Dense	Units = 1, Activation = Sigmoid	1

stand-to-sit activities in their preferred style and pace. The elderly dataset was sampled in fixed-width sliding windows

of 2 seconds and 50% overlap (100 readings/window). A full description of the used dataset can be found in section 4.1.1 of [48].

Preprocessed data is used to train the FCGAN and compare it with the state-of-the-art Unified CGAN by Hong *et al.* [26]. Both models learn the latent patterns of human activities in a single training process. The models' training was continued for 1000 epochs with a batch size of 60. The experiments were implemented in Google Colaboratory (also known as Colab)'s Jupyter notebook environment using Keras python library with Tensorflow as the backend. Google Colab is chosen due to its runtime configurations for deep learning applications and free access to a robust GPU.

V. MODEL EVALUATION

This study uses four techniques to evaluate the performance of the proposed approach: training evaluation, visual evaluation, similarity measure evaluation, and usability evaluation.

A. TRAINING EVALUATION

GAN fully converges when the discriminator cannot differentiate real examples from fake ones and is considered stable when the generator's active models did not have a high loss value after model convergence. The stability and the speed of GAN full convergence are highly associated with the loss pattern of the generator and discriminator as it dictates the number of epochs required by the models to fully converge. The objective of the first evaluation technique, training evaluation, is to provide more insights into how far the proposed architecture is more stable and achieves faster model convergence. It visualizes the proposed architecture, FCGAN, model convergence, and stability and compares it with the state-of-the-art CGAN unified CGAN architecture. This is performed by extracting the discriminator's loss, and generator's loss information during the models' training process.

Figure 1 compares the discriminator's loss of the Unified CGAN with the discriminator's loss of the FCGAN. It shows that the FCGAN discriminator has higher loss values than the Unified CGAN discriminator across all epochs.

Discriminator's higher loss has the potential to provide larger gradients to help the generator learn the data distribution faster as shown in Figure 2 which visualizes the generator's loss of the compared models. As can be observed, the FCGAN generator scores lower loss values compared to the Unified CGAN. This leads to a faster model convergence and achieves better learning stability.

The experiments show that the generator of the Unified CGAN starts to converge at epoch 874 while the generator of the Fully Connected CGAN starts to converge at epoch 235. This is considered the first stage of full convergence of both models. This study uses the generated data in this period for investigation and analysis. The experiments also confirm that the FCGAN, unlike the Unified CGAN, remains more stable after its first stage till its convergence at the end of training epochs. On the other hand, the Unified CGAN fails

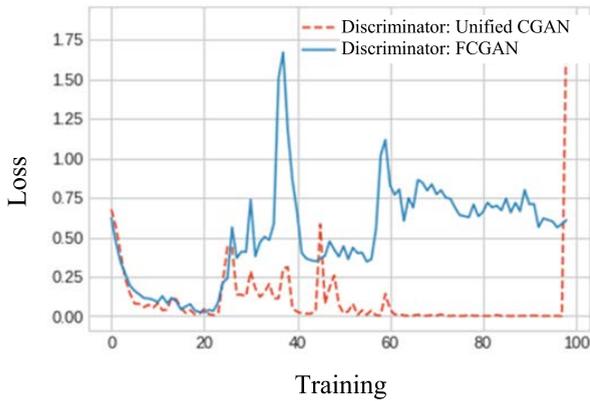


FIGURE 1. Comparison of discriminator's loss of Unified CGAN and FCGAN.

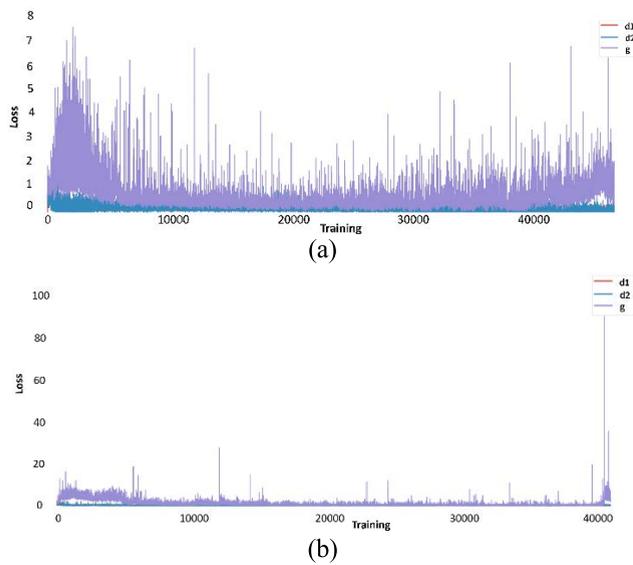


FIGURE 2. Comparison of training process for (a) Unified CGAN and (b) FCGAN.

to converge at 1000 epochs, confirming this model is unstable during the training process.

These confirm that the proposed architecture achieves faster model convergence than the state-of-the-art GAN architecture.

B. VISUAL EVALUATION

The objective of this evaluation techniques is to assess the quality of the generated synthetic data. It compares the acceleration of real data samples with the acceleration of synthetic data samples to see how the generative models can fabricate the patterns of the real data. The samples of the real data and the synthetic data are selected randomly from the three types of the activities in this study: dynamic (walking), static (standing), and transition (stand-to-sit). The visual evaluation of the real sample data and fake sample data of FCGAN and Unified CGAN for walking, standing, and stand-to-sit activities are shown in Figure 3.

In general, considering standing activity, the patterns of the synthetic signals of both FCGAN and Unified CGAN are

verisimilar to the pattern of the real signals. In the case of walking and stand-to-sit activities, the patterns of the synthetic signals of the generative models are not similar to the pattern of the real sample data. However, the Unified CGAN manages to reconstruct the signal pattern of the different axes of real data for each activity.

Although Unified CGAN able to generate synthetic data that captures the underlying pattern of the real data, the pattern is quite different from the real data. The Gaussian input noise feed into the unified CGAN model only has a minor effect in increasing synthetic data variations.

Nevertheless, the local evaluation also shows that FCGAN able to generate signals that are more similar to the real signals than the Unified CGAN in all types of activities, and more precisely in dynamic and transition activities. This confirms that the synthetic data produced by the FCGAN more accurately represents the real data than Unified CGAN.

C. SIMILARITY MEASURE EVALUATION

Euclidean Distance Measure is a well-known method used by both past [49] and state-of-the-art studies [26] for resource-limited devices like sensors to evaluate the similarity between time series data for sensor-based activity recognition. This is due to its low computational complexity and data storage requirement while demonstrating the variation within each class of generated samples.

The objective of this evaluation is to evaluate the variability of the generated data by the proposed model. We, therefore, use Euclidean Distance Measure to evaluate the performance of the synthetic data generated by the proposed model. This is conducted by comparing the similarity of the synthetic data generated by the proposed model in this study with the similarity of the synthetic data generated by the state-of-the-art CGAN. We also compare the e similarity of synthetic data of the comparative models with the real data.

This determines the similarity between the real and synthetic data (RTS) as well as the similarity between the real to real data (RTR) and synthetic to synthetic data (STS). For each experiment, the degree of similarity between two activity signals (2 seconds) is measured and recorded. The signals are randomly selected, and each experiment was run at least 10 times to avoid bias in the analysis of the evaluation. The distance function to measure the similarity between *a* (one data point of synthetic sample *a*) and *b* (one data point of synthetic sample *b*) is given in (2).

$$D_E(a, b) = \left[\sum_{i=1}^n (a - b) \right]^{1/2} \tag{2}$$

Table 5 shows the Euclidean Distance of the synthetic data generated by FCGAN. The results show that the Euclidean Distance between real to real (RTR) and real to synthetic (RTS) are very similar in all the dynamic and transitional activities. This is good indication as the gap between most of the generated data and real data is small. The dissimilarity in RTR and RTS for stationary activities is high compared to other activities. This problem is also observed in

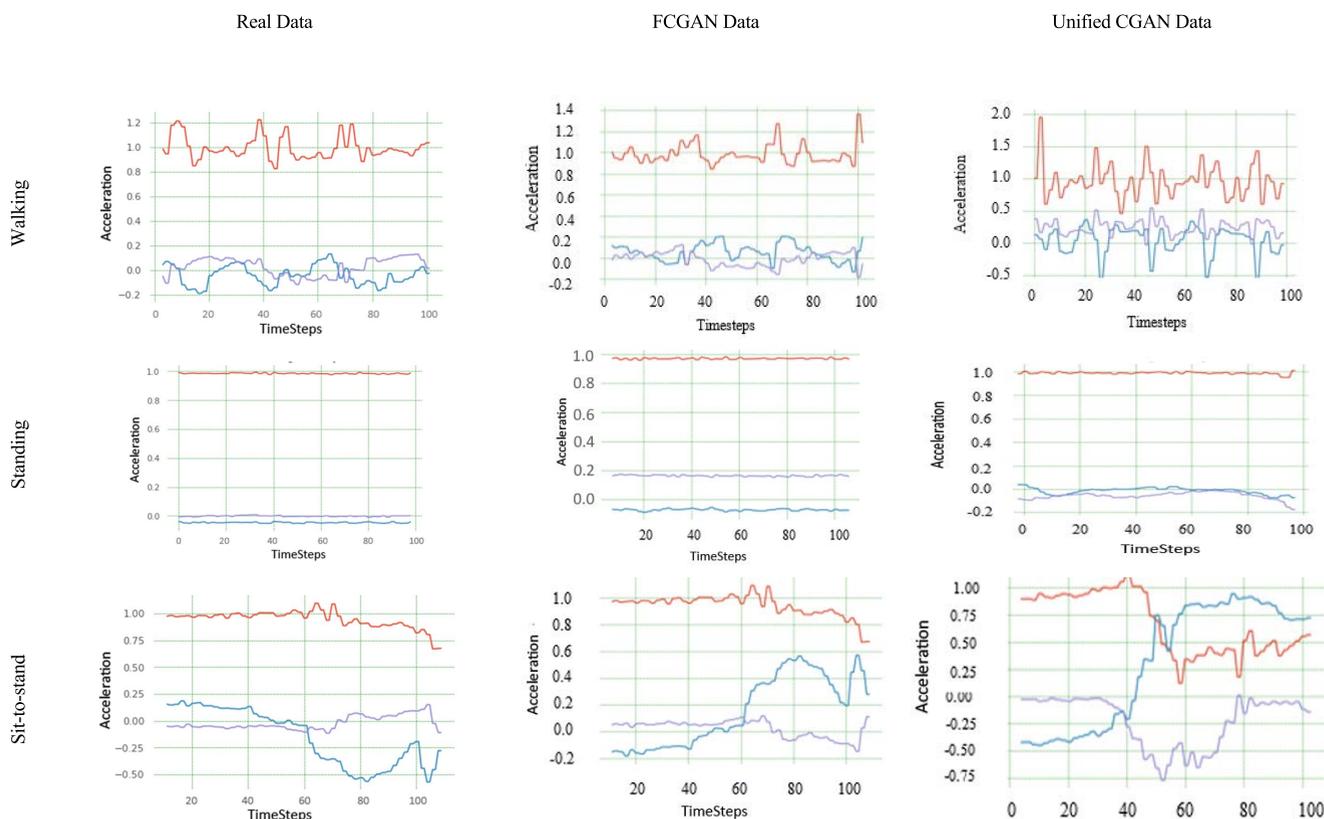


FIGURE 3. Visual evaluation of walking, standing and stand-to-sit activities.

TABLE 5. FCGAN similarity measure using euclidean distance.

Activity	RTR	RTS	STS
Walking	0.54	0.60	0.31
Standing	0.39	0.61	0.48
Stand-to-sit	0.68	0.71	0.42
Sitting	0.37	0.70	0.47
Sit-to-stand	0.66	0.73	0.63
Sit-to-lie	0.69	0.82	0.56
Lying down	0.48	0.71	0.34
Lie-to-sit	0.77	0.80	0.55

the state-of-the-art CGAN method for sensor-based activity recognition. The results also show that dissimilarity in STS for all activities are lower than the those in RTR and RTS. This confirms that there is a little variation in the resulting pattern among the synthetic data.

Table 6 shows the Euclidean Distance of the synthetic data generated by Unified CGAN. Like the FCGAN, the Euclidean Distance between real to real (RTR) and real to synthetic (RTS) are very similar in all the dynamic and transitional activities and the dissimilarity in RTR and RTS for stationary activities is high compared to other activities. The results also show that dissimilarity in STS for all activities are lower than the those in RTR and RTS. They are even lower than those in FCGAN. This proves that the Unified CGAN model suffer from a greater mode collapse than FCGAN.

D. USABILITY EVALUATION

The third evaluation technique, usability evaluation, aims to investigate the quality of the synthetic data generated by the proposed architecture in improving the performance of activity recognition classification models. First, the synthetic data generated by both FCGAN architecture and Unified GAN is preprocessed to perform four experiments on it together with the real data using the best-performing deep learning classifiers by Jimale and Noor[48]. These experiments are experiments on 70% real data and 30% synthetic data, experiments on 50% real data and 50% synthetic data, experiments on 30% real data and 70% synthetic data, and experiments on 100% synthetic data. The classification accuracy, measured using (3), is shown in a form of average accuracy. To record the average accuracy, each experiment was run at least 10 times.

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative respectively.

Table 7 shows the overall classification performance of the comparative models. As the table demonstrates, there is about 2.5%, 2.5%, 3.1%, and 4.4% recognition performance gain by the proposed architecture in the overall experimental stages. This confirms that Fully Connected CGAN generates better synthetic samples than the Unified CGAN. The results also show that the average accuracy of the classification models

TABLE 6. Unified CGAN similarity measure using euclidean distance.

Activity	RTR	RTS	STS
Walking	0.54	0.59	0.29
Standing	0.39	0.62	0.47
Stand-to-sit	0.68	0.71	0.41
Sitting	0.37	0.58	0.43
Sit-to-stand	0.66	0.77	0.61
Sit-to-lie	0.69	0.84	0.55
Lying down	0.48	0.73	0.32
Lie-to-sit	0.77	0.83	0.52

TABLE 7. Average classification performance using deep learning.

Experimental Stages	Stage I	Stage II	Stage III	Stage IV
Ratio of Training Data (Real %: Synthetic%)	70:30	50:50	30:70	0:100
Unified CGAN average accuracy (%)	80.1	78.9	74.2	71.9
FCGAN average accuracy (%)	82.6	81.4	77.3	76.3

TABLE 8. Confusion matrix of unified CGAN for stage I.

	0	1	2	3	4	5	6	7
0	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.01
1	0.00	0.83	0.12	0.05	0.00	0.00	0.00	0.00
2	0.07	0.09	0.76	0.00	0.04	0.02	0.02	0.00
3	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.01
4	0.17	0.04	0.00	0.00	0.68	0.04	0.07	0.00
5	0.10	0.00	0.04	0.00	0.07	0.77	0.01	0.01
6	0.00	0.00	0.00	0.20	0.12	0.00	0.68	0.00
7	0.12	0.00	0.00	0.03	0.00	0.13	0.01	0.71

0: walking, 1: standing, 2: stand-to-sit, 3: sitting, 4: sit-to-stand, 5: sit-to-lie, 6: lying down, 7: lie-to-sit

drops whenever real data is hybrid with synthetic data from the generative models in both models. However, the accuracy drop of the Unified CGAN is 0.6% and 1.3% higher than the FCGAN when the ratio of the real and synthetic data is 50%:50%, and 30%:70% respectively. This confirms that the proposed architecture generates more realistic synthetic data than the state-of-the-art architecture.

We also show the confusion matrix of the four experimental stages to better understand the performance of each class of human activities. The best classification performance of the overall experimental stages is achieved in stage I as shown in Table 8 and Table 9. All classes of activities perform well with quite low classification performance for class 7 (Lie-to-sit), class 6 (Lying down), and class 4 (Sit-to-stand). The results of the stage I also reveal that FCGAN outperforms Unified CGAN in all activities except for class 2 (Stand-to-sit) and class 4 (Sit-to-stand).

As shown in Table 10 and Table 11, the recognition performance of all activity classes drops in stage II except for Unified CGAN’s class 3 (sitting) and class 7(lie-to-sit). Class 0(walking) of the comparative models suffers the highest drop while class 7(lie-to-sit) of Unified CGAN enjoys the highest improvement. Nevertheless, the majority of FCGAN activity classes maintain to have higher True Positives than Unified CGAN.

TABLE 9. Confusion matrix of FCGAN for stage I.

	0	1	2	3	4	5	6	7
0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0.00	0.86	0.10	0.04	0.00	0.00	0.00	0.00
2	0.07	0.09	0.76	0.00	0.04	0.02	0.02	0.00
3	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
4	0.17	0.04	0.00	0.00	0.68	0.04	0.07	0.00
5	0.10	0.00	0.04	0.00	0.07	0.79	0.00	0.00
6	0.00	0.00	0.00	0.13	0.10	0.00	0.77	0.00
7	0.12	0.00	0.00	0.00	0.00	0.13	0.00	0.75

0: walking, 1: standing, 2: stand-to-sit, 3: sitting, 4: sit-to-stand, 5: sit-to-lie, 6: lying down, 7: lie-to-sit

TABLE 10. Confusion matrix of unified CGAN for stage II.

	0	1	2	3	4	5	6	7
0	0.92	0.00	0.04	0.00	0.02	0.00	0.00	0.02
1	0.00	0.86	0.10	0.04	0.00	0.00	0.00	0.00
2	0.07	0.10	0.71	0.02	0.04	0.03	0.03	0.00
3	0.00	0.00	0.00	0.98	0.00	0.01	0.00	0.01
4	0.17	0.04	0.00	0.01	0.65	0.05	0.08	0.00
5	0.10	0.00	0.05	0.00	0.07	0.74	0.04	0.00
6	0.00	0.00	0.01	0.13	0.11	0.00	0.75	0.00
7	0.13	0.00	0.00	0.01	0.00	0.13	0.03	0.70

0: walking, 1: standing, 2: stand-to-sit, 3: sitting, 4: sit-to-stand, 5: sit-to-lie, 6: lying down, 7: lie-to-sit

TABLE 11. Confusion matrix of FCGAN for stage II.

	0	1	2	3	4	5	6	7
0	0.94	0.00	0.04	0.00	0.01	0.00	0.00	0.01
1	0.00	0.86	0.10	0.04	0.00	0.00	0.00	0.00
2	0.07	0.09	0.73	0.01	0.04	0.03	0.03	0.00
3	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
4	0.17	0.04	0.00	0.00	0.68	0.04	0.07	0.00
5	0.10	0.00	0.04	0.00	0.07	0.78	0.01	0.00
6	0.00	0.00	0.00	0.13	0.10	0.00	0.77	0.00
7	0.12	0.00	0.00	0.00	0.00	0.13	0.00	0.75

0: walking, 1: standing, 2: stand-to-sit, 3: sitting, 4: sit-to-stand, 5: sit-to-lie, 6: lying down, 7: lie-to-sit

TABLE 12. Confusion matrix of unified CGAN for stage III.

	0	1	2	3	4	5	6	7
0	0.87	0.00	0.04	0.00	0.04	0.00	0.02	0.03
1	0.00	0.84	0.11	0.03	0.02	0.00	0.00	0.00
2	0.07	0.10	0.68	0.03	0.06	0.03	0.03	0.00
3	0.00	0.01	0.00	0.95	0.00	0.01	0.01	0.02
4	0.17	0.05	0.01	0.02	0.62	0.05	0.08	0.00
5	0.10	0.00	0.06	0.00	0.08	0.68	0.06	0.02
6	0.00	0.02	0.03	0.21	0.1	0.00	0.64	0.00
7	0.13	0.00	0.00	0.02	0.00	0.14	0.05	0.66

0: walking, 1: standing, 2: stand-to-sit, 3: sitting, 4: sit-to-stand, 5: sit-to-lie, 6: lying down, 7: lie-to-sit

Likewise, the recognition performance of all activity classes drops in stage III but without any improvements seen in stage III (see Table 12 and Table 13). Class 5 (sit-to-lie), class 7 (lie-to-sit), and class 0 (walking) suffers the most drops of FCGAN respectively while class 6 (lying down), class 5 (sit-to-lie), and class 0 (walking) scores the highest drop of the unified CGAN respectively.

As shown in Table 14 and Table 15, the recognition performance of all activity classes drops in the final stage of experiments for the Unified CGAN without any performance improvement of a single class. class 0 (walking), class 2 (stand-to-sit), class 3 (sitting) scores the worst drop in this

TABLE 13. Confusion matrix of FCGAN for stage III.

	0	1	2	3	4	5	6	7
0	0.89	0.00	0.04	0.00	0.02	0.00	0.02	0.03
1	0.00	0.86	0.10	0.04	0.00	0.00	0.00	0.00
2	0.07	0.10	0.69	0.03	0.05	0.03	0.03	0.00
3	0.00	0.01	0.00	0.97	0.00	0.01	0.00	0.01
4	0.17	0.05	0.00	0.02	0.63	0.05	0.08	0.00
5	0.10	0.00	0.06	0.00	0.08	0.71	0.05	0.00
6	0.00	0.00	0.01	0.13	0.11	0.00	0.75	0.00
7	0.13	0.00	0.00	0.01	0.00	0.13	0.04	0.69

0: walking, 1: standing, 2: stand-to-sit, 3: sitting, 4: sit-to-stand, 5: sit-to-lie, 6: lying down, 7: lie-to-sit

TABLE 14. Confusion matrix of unified CGAN for stage IV.

	0	1	2	3	4	5	6	7
0	0.81	0.00	0.06	0.01	0.02	0.00	0.05	0.05
1	0.00	0.79	0.11	0.10	0.00	0.00	0.00	0.00
2	0.07	0.10	0.62	0.06	0.05	0.05	0.05	0.00
3	0.00	0.01	0.00	0.89	0.00	0.02	0.00	0.02
4	0.17	0.05	0.00	0.03	0.61	0.06	0.08	0.00
5	0.10	0.01	0.07	0.01	0.08	0.68	0.05	0.00
6	0.00	0.00	0.01	0.22	0.12	0.00	0.65	0.00
7	0.13	0.00	0.01	0.02	0.00	0.14	0.04	0.66

0: walking, 1: standing, 2: stand-to-sit, 3: sitting, 4: sit-to-stand, 5: sit-to-lie, 6: lying down, 7: lie-to-sit

TABLE 15. Confusion matrix of FCGAN for stage IV.

	0	1	2	3	4	5	6	7
0	0.85	0.00	0.06	0.00	0.02	0.00	0.02	0.05
1	0.00	0.84	0.11	0.05	0.00	0.00	0.00	0.00
2	0.07	0.10	0.69	0.03	0.05	0.03	0.03	0.00
3	0.00	0.01	0.00	0.95	0.00	0.02	0.00	0.02
4	0.17	0.05	0.00	0.02	0.63	0.05	0.08	0.00
5	0.10	0.00	0.07	0.00	0.08	0.70	0.05	0.00
6	0.00	0.00	0.01	0.13	0.11	0.00	0.75	0.00
7	0.13	0.00	0.00	0.01	0	0.13	0.04	0.69

0: walking, 1: standing, 2: stand-to-sit, 3: sitting, 4: sit-to-stand, 5: sit-to-lie, 6: lying down, 7: lie-to-sit

case. In FCGAN stage IV experiments, on the other hand, half of activity classes maintained the same performance as in stage III. A minor performance drop is seen in the other activity classes (class 0 (walking), class 1 (standing), class 3 (sitting), and class 5 (sit-to-lie) respectively).

VI. CONCLUSION

This paper has proposed an enhanced GGAN architecture for sensor-based activity recognition that synthesizes a more natural transformation of human activity signals and achieves faster model learning convergence and training stability. In the proposed architecture, the generator and discriminator networks encompass deep fully connected and convolution layers, in contrast to state-of-the-art network architecture. We have conducted several experiments on sensory data collected from elderly data and showed that our proposed architecture generates better samples and converges faster through visual and usability evaluation techniques. All our experimental stages were limited to supervised CGANs. However, the proposed method is significant enough to be combined with any other GAN configuration. Therefore, a possible extension of the proposed work is to study its effectiveness in unsupervised and semi-supervised GAN setups. In the future,

the enhanced architecture will also be improved further to produce more quality samples. In addition, other datasets for sensory-based activity recognition will be experimented with to show the robustness of the newly proposed network.

REFERENCES

- [1] A. Wang, S. Zhao, C. Zheng, H. Chen, L. Liu, and G. Chen, "Hier-HAR: Sensor-based data-driven hierarchical human activity recognition," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3353–3365, Feb. 2021, doi: [10.1109/JSEN.2020.3023860](https://doi.org/10.1109/JSEN.2020.3023860).
- [2] M. H. M. Noor, Z. Salcic, and K. I.-K. Wang, "Ontology-based sensor fusion activity recognition," *J. Ambient Intell. Hum. Comput.*, vol. 11, no. 8, pp. 3073–3087, Aug. 2020, doi: [10.1007/s12652-017-0668-0](https://doi.org/10.1007/s12652-017-0668-0).
- [3] G. Farias, S. Dormido-Canto, J. Vega, G. Rattá, H. Vargas, G. Hermosilla, L. Alfaro, and A. Valencia, "Automatic feature extraction in large fusion databases by using deep learning approach," *Fusion Eng. Des.*, vol. 112, pp. 979–983, Nov. 2016, doi: [10.1016/j.fusengdes.2016.06.016](https://doi.org/10.1016/j.fusengdes.2016.06.016).
- [4] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, Mar. 2019, doi: [10.1016/j.patrec.2018.02.010](https://doi.org/10.1016/j.patrec.2018.02.010).
- [5] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl.*, vol. 105, pp. 233–261, Sep. 2018, doi: [10.1016/j.eswa.2018.03.056](https://doi.org/10.1016/j.eswa.2018.03.056).
- [6] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Jan. 2015, pp. 3995–4001.
- [7] M. Hasan and K. A. Roy-Chowdhury, "A continuous learning framework for activity recognition using deep hybrid feature models," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 1909–1922, Nov. 2015, doi: [10.1109/TMM.2015.2477242](https://doi.org/10.1109/TMM.2015.2477242).
- [8] Q. Yang, "Activity recognition: Linking low-level sensors to high-level intelligence," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2009, pp. 20–25.
- [9] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, Oct. 2016, doi: [10.1016/j.eswa.2016.04.032](https://doi.org/10.1016/j.eswa.2016.04.032).
- [10] D. Shi, Y. Li, and B. Ding, "Unsupervised feature learning for human activity recognition," *Guofang Keji Daxue Xuebao/J. Nat. Univ. Defence Technol.*, vol. 37, no. 5, pp. 128–134, 2015, doi: [10.11887/j.cn.201505020](https://doi.org/10.11887/j.cn.201505020).
- [11] F. Shaheen, B. Verma, and M. Asafuddoula, "Impact of automatic feature extraction in deep learning architecture," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov. 2016, pp. 1–8, doi: [10.1109/DICTA.2016.7797053](https://doi.org/10.1109/DICTA.2016.7797053).
- [12] L. Alawneh, T. Alsarhan, M. Al-Zinati, M. Al-Ayyoub, Y. Jararweh, and H. Lu, "Enhancing human activity recognition using deep learning and time series augmented data," *J. Ambient Intell. Hum. Comput.*, vol. 12, no. 12, pp. 10565–10580, Dec. 2021, doi: [10.1007/s12652-020-02865-4](https://doi.org/10.1007/s12652-020-02865-4).
- [13] K. M. Rashid and J. Louis, "Times-series data augmentation and deep learning for construction equipment activity recognition," *Adv. Eng. Informat.*, vol. 42, Oct. 2019, Art. no. 100944, doi: [10.1016/j.aei.2019.100944](https://doi.org/10.1016/j.aei.2019.100944).
- [14] S. Zhang and N. Alshurafa, "Deep generative cross-modal on-body accelerometer data synthesis from videos," in *Proc. Adjunct Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput.*, Sep. 2020, pp. 223–227, doi: [10.1145/3410530.3414329](https://doi.org/10.1145/3410530.3414329).
- [15] M. Kim and C. Y. Jeong, "Label-preserving data augmentation for mobile sensor data," *Multidimensional Syst. Signal Process.*, vol. 32, no. 1, pp. 115–129, Jan. 2021, doi: [10.1007/s11045-020-00731-2](https://doi.org/10.1007/s11045-020-00731-2).
- [16] S. N. Tran, T. D. Nguyen, T.-S. Ngo, X.-S. Vu, L. Hoang, Q. Zhang, and M. Karunanithi, "Deep learning for multi-resident activity recognition in ambient sensing smart homes," *Artif. Intell. Rev.*, vol. 53, no. 3, pp. 340–341, 2019, doi: [10.1007/s10462-019-09783-8](https://doi.org/10.1007/s10462-019-09783-8).
- [17] C. F. S. Leite and Y. Xiao, "Improving cross-subject activity recognition via adversarial learning," *IEEE Access*, vol. 8, pp. 90542–90554, 2020, doi: [10.1109/ACCESS.2020.2993818](https://doi.org/10.1109/ACCESS.2020.2993818).
- [18] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," Nov. 2017, *arXiv:1711.04340*. Accessed: Jun. 18, 2020.

- [19] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–40, May 2022, doi: [10.1145/3447744](https://doi.org/10.1145/3447744).
- [20] T. DeVries and G. W. Taylor, "Dataset augmentation in feature space," in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–12.
- [21] W.-N. Hsu, Y. Zhang, and J. Glass, "Unsupervised domain adaptation for robust speech recognition via variational autoencoder-based data augmentation," in *Proc. IEEE Autom. Speech Recognit. Understand. Workshop (ASRU)*, Dec. 2017, pp. 16–23, doi: [10.1109/ASRU.2017.8268911](https://doi.org/10.1109/ASRU.2017.8268911).
- [22] G. Boquet, J. Vicario, A. Morell, and J. Serrano, "Missing data in traffic estimation: A variational autoencoder imputation method," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, U.K., May 2019, pp. 2882–2886.
- [23] H. Ko, J. Lee, J. Kim, J. Lee, and H. Shim, "Diversity regularized autoencoders for text generation," in *Proc. 35th Annu. ACM Symp. Appl. Comput.*, Mar. 2020, pp. 883–891, doi: [10.1145/3341105.3373998](https://doi.org/10.1145/3341105.3373998).
- [24] S. Mesbah, J. Yang, R.-J. Sips, M. V. Torre, C. Lofi, A. Bozzon, and G.-J. Houben, "Training data augmentation for detecting adverse drug reactions in user-generated content," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2020, pp. 2349–2359.
- [25] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0).
- [26] M. H. Chan and M. H. M. Noor, "A unified generative model using generative adversarial network for activity recognition," *J. Ambient Intell. Hum. Comput.*, vol. 12, no. 7, pp. 8119–8128, Jul. 2021, doi: [10.1007/s12652-020-02548-0](https://doi.org/10.1007/s12652-020-02548-0).
- [27] X. Li, J. Luo, and R. Younes, "ActivityGAN: Generative adversarial networks for data augmentation in sensor-based human activity recognition," in *Proc. Adjunct ACM Int. Joint Conf. Pervasive Ubiquitous Comput., ACM Int. Symp. Wearable Comput.*, Sep. 2020, pp. 249–254, doi: [10.1145/3410530.3414367](https://doi.org/10.1145/3410530.3414367).
- [28] M. Abdu-Aguye and W. Goma, "Robust human activity recognition based on deep metric learning," in *Proc. 16th Int. Conf. Informat. Control, Autom. Robot.*, 2019, pp. 656–663, doi: [10.5220/0007916806560663](https://doi.org/10.5220/0007916806560663).
- [29] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, pp. 1–33, Jan. 2014, doi: [10.1145/2499621](https://doi.org/10.1145/2499621).
- [30] T.-C. Chiang, B. Bruno, R. Menicatti, C. T. Recchiuto, and A. Sgorbissa, "Culture as a sensor? A novel perspective on human activity recognition," *Int. J. Social Robot.*, vol. 11, no. 5, pp. 797–814, Dec. 2019, doi: [10.1007/s12369-019-00590-3](https://doi.org/10.1007/s12369-019-00590-3).
- [31] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, R. A. Peterson, H. Lu, X. Zheng, M. Musolesi, K. Fodor, and G.-S. Ahn, "The rise of people-centric sensing," *IEEE Internet Comput.*, vol. 12, no. 4, pp. 12–21, Jul./Aug. 2008. [Online]. Available: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4557974>
- [32] J.-H. Li, L. Tian, H. Wang, Y. An, K. Wang, and L. Yu, "Segmentation and recognition of basic and transitional activities for continuous physical human activity," *IEEE Access*, vol. 7, pp. 42565–42576, 2019, doi: [10.1109/ACCESS.2019.2905575](https://doi.org/10.1109/ACCESS.2019.2905575).
- [33] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1629–1645, 3rd Quart., 2020, doi: [10.1109/COMST.2019.2934489](https://doi.org/10.1109/COMST.2019.2934489).
- [34] D. Cook, K. D. F. Feuz, and N. C. Krishnan, "Transfer learning for activity recognition: A survey diane," *Bone*, vol. 23, no. 1, pp. 1–7, 2008, doi: [10.1038/jid.2014.371](https://doi.org/10.1038/jid.2014.371).
- [35] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 790–808, Nov. 2012, doi: [10.1109/TSMCC.2012.2198883](https://doi.org/10.1109/TSMCC.2012.2198883).
- [36] S. Ali and M. Shah, "Human action recognition in videos using kinematic features and multiple instance learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 288–303, Feb. 2010, doi: [10.1109/TPAMI.2008.284](https://doi.org/10.1109/TPAMI.2008.284).
- [37] Z. Hussain, M. Sheng, and W. E. Zhang, "Different approaches for human activity recognition: A survey," 2019, *arXiv:1906.05074*.
- [38] S. R. Ramamurthy and N. Roy, "Recent trends in machine learning for human activity recognition—A survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 4, pp. 1–11, Jul. 2018, doi: [10.1002/widm.1254](https://doi.org/10.1002/widm.1254).
- [39] E. Zdravetski, P. Lameski, V. Trajkovik, A. Kulakov, I. Chorbev, R. Goleva, N. Pombo, and N. Garcia, "Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering," *IEEE Access*, vol. 5, pp. 5262–5280, 2017, doi: [10.1109/ACCESS.2017.2684913](https://doi.org/10.1109/ACCESS.2017.2684913).
- [40] H. Ono and S. Suzuki, "Data augmentation for gross motor-activity recognition using DCGAN," in *Proc. IEEE/SICE Int. Symp. Syst. Integr. (SII)*, Jan. 2020, pp. 440–443, doi: [10.1109/SII46433.2020.9026252](https://doi.org/10.1109/SII46433.2020.9026252).
- [41] B. Almaslakh, J. Al Muhtadi, and A. M. Artoli, "A robust convolutional neural network for online smartphone-based human activity recognition," *J. Intell. Fuzzy Syst.*, vol. 35, no. 2, pp. 1609–1620, Aug. 2018, doi: [10.3233/JIFS-169699](https://doi.org/10.3233/JIFS-169699).
- [42] S. Mahdizadehghadam, A. Panahi, and H. Krim, "Generative adversarial nets," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2014, pp. 3063–3071, doi: [10.1109/ICCVW.2019.00369](https://doi.org/10.1109/ICCVW.2019.00369).
- [43] C. Xiao, D. Han, Y. Ma, and Z. Qin, "CsiGAN: Robust channel state information-based activity recognition with GANs," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10191–10204, Dec. 2019, doi: [10.1109/JIOT.2019.2936580](https://doi.org/10.1109/JIOT.2019.2936580).
- [44] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–25, Dec. 2018, doi: [10.1145/3161183](https://doi.org/10.1145/3161183).
- [45] J. Wang, Y. Chen, Y. Gu, Y. Xiao, and H. Pan, "SensoryGANs: An effective generative adversarial framework for sensor-based human activity recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8, doi: [10.1109/IJCNN.2018.8489106](https://doi.org/10.1109/IJCNN.2018.8489106).
- [46] J. Shi, D. Zuo, and Z. Zhang, "A GAN-based data augmentation method for human activity recognition via the caching ability," *Internet Technol. Lett.*, vol. 4, no. 5, pp. 4–9, Sep. 2021, doi: [10.1002/itl2.257](https://doi.org/10.1002/itl2.257).
- [47] S. Barua, S. M. Erfani, and J. Bailey, "FCC-GAN: A fully connected and convolutional net architecture for GANs," 2019, *arXiv:1905.02417*.
- [48] A. O. Jimale and M. H. M. Noor, "Subject variability in sensor-based activity recognition," *J. Ambient Intell. Hum. Comput.*, pp. 1–14, Sep. 2021, doi: [10.1007/s12652-021-03465-6](https://doi.org/10.1007/s12652-021-03465-6).
- [49] L. Liu, Y. Peng, M. Liu, and Z. Huang, "Sensor-based human activity recognition system with a multilayered model using time series shapelets," *Knowl.-Based Syst.*, vol. 90, pp. 138–152, Dec. 2015, doi: [10.1016/j.knosys.2015.09.024](https://doi.org/10.1016/j.knosys.2015.09.024).



ALI OLOW JIMALE (Member, IEEE) is currently pursuing the Ph.D. degree with the School of Computer Sciences, USM, Malaysia. He is currently the Director of the Center for Research and Development and a Lecturer with the Faculty of Computing, SIMAD University. He is also a Teaching Assistant and a Graduate Research Assistant with the School of Computer Sciences, USM. His research interests include machine learning, activity recognition, and parallel computing.



MOHD HALIM MOHD NOOR is an Academician with the School of Computer Sciences, Universiti Sains Malaysia (USM). He was a Senior Lecturer in computer engineering at the Universiti Teknologi MARA, Pulau Pinang. Currently, he is focusing on problems in human motion analysis. His research interests include machine learning and deep learning for computer vision and pervasive computing.

...