

Article

# Semi-Supervised Adversarial Auto-Encoder to Expedite Human Activity Recognition

Keshav Thapa <sup>1</sup> , Yousung Seo <sup>2</sup>, Sung-Hyun Yang <sup>3</sup> and Kyong Kim <sup>1,\*</sup>

<sup>1</sup> Department of Rehabilitation Medical Engineering, Daegu Haany University, Gyeongsan-si 38610, Republic of Korea

<sup>2</sup> Department of Geriatric Rehabilitation, Daegu Haany University, Gyeongsan-si 38610, Republic of Korea

<sup>3</sup> Department of Electronic Engineering, Kwangwoon University, Seoul 01897, Republic of Korea

\* Correspondence: kkruddy@dhu.ac.kr

**Abstract:** The study of human activity recognition concentrates on classifying human activities and the inference of human behavior using modern sensing technology. However, the issue of domain adaptation for inertial sensing-based human activity recognition (HAR) is still burdensome. The existing requirement of labeled training data for adapting such classifiers to every new person, device, or on-body location is a significant barrier to the widespread adoption of HAR-based applications, making this a challenge of high practical importance. We propose the semi-supervised HAR method to improve reconstruction and generation. It executes proper adaptation with unlabeled data without changes to a pre-trained HAR classifier. Our approach decouples VAE with adversarial learning to ensure robust classifier operation, without newly labeled training data, under changes to the individual activity and the on-body sensor position. Our proposed framework shows the empirical results using the publicly available benchmark dataset compared to state-of-art baselines, achieving competitive improvement for handling new and unlabeled activity. The result demonstrates SAA has achieved a 5% improvement in classification score compared to the existing HAR platform.

**Keywords:** semi-supervised; auto-encoder; human activity recognition; adversarial learning



**Citation:** Thapa, K.; Seo, Y.; Yang, S.-H.; Kim, K. Semi-Supervised Adversarial Auto-Encoder to Expedite Human Activity Recognition. *Sensors* **2023**, *23*, 683. <https://doi.org/10.3390/s23020683>

Academic Editor: Andrea Cataldo

Received: 21 November 2022

Revised: 26 December 2022

Accepted: 29 December 2022

Published: 6 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Embedded wearable inertial sensors permit unobtrusive and regular monitoring, making human activity recognition an ideal platform for health assessment [1], predicting depression, cognitive and mental states [2], and monitoring sleep and fitness [3]. Applications for HAR systems in the actual world include smart homes [4], defenses [5], astronauts [6], senior care [7], and defense applications. However, as they must now take into account all unexpected changes in the real-time scenario, the current approaches face significant difficulties in accurately recognizing activities. Modern gesture recognition systems have great accuracy and are based on shallow or deep neural network (DNN) models [8]; however, they still have a significant problem. The pre-trained ML models are strong enough to handle instance-specific differences in the sensor data due to user diversity and their changing activity schedules. It is impractical for actual societal-scale deployment to use the standard method for addressing such heterogeneity, which involves using instance-specific labeled data to develop individual classifiers. Instead, substantial research has concentrated on automated domain adaptation methods, requiring no labeled training data. The majority of HAR machine learning methods, including k-nearest neighbor [9], decision trees [10], and support vector machines [11] in the literature, rely on heuristic feature extraction to build their models. For 3D sensors, this comprises correlation (Pearson correlation) between axes, mean and standard deviation for each sensor signal, and time-domain calculations. The use of transfer learning to alter a training domain model with only modest amounts of labeled data from the test domain has been proposed in recent years [12]; the mapping of domain-dependent sensor values to a domain-independent,

common low-dimensional latent space [13]; and the use of adversarial learning. At the moment, efforts at HAR are mainly directed toward learning actively—acquiring user comments about new activities, and detecting changes—finding new activities [14–16]. The model must be rebuilt and retrained when adding a new activity class. Few studies have looked into the potential for an activity model to emerge automatically with various activities [17]. However, this capability has the advantage of maintaining the knowledge in the time-tested business model while reducing the need for manual feature engineering, manual configuration, and training expenses.

Several supervised [18] and semi-supervised [19] approaches to activity recognition exist. These models offer good accuracy when given enough training data. However, their performance suffers when applied to novel and unexplored distributions. Therefore, it is still difficult for the model to identify a new user's activities. Most machine learning [14] and deep learning are not conceptually aware of all activities, but with the proper learning and models, they may effectively recognize human behavior. Many artificial intelligence models and cutting-edge techniques [20] are based on deep neural networks.

On the other hand, deep learning needs a lot of data to serve as a label for learning. The main disadvantage of HAR is that each sensor must be installed and controlled separately. Additionally, the domain's experts should only understand and label unlabeled data, increasing the labeling task. As a result, our strategy ensures that the activity detection needs are primarily met by better performance than prior methods. This research focuses on semi-supervised adversarial learning, which combines adversarial learning, deep learning, and semi-supervised learning (VAE) to ensure that no labels based on previously learned data can be fully expected. Additionally, there is a chance that this method could enhance speed by employing fewer tagging classes. The advantages of our approach are as follows:

1. We proposed a semi-supervised model that can adapt without labeled data or changes in the pre-trained classifier to identify human activity.
2. Demonstration of adversarial autoencoder (AAE) efficacy and robustness, so that the model will be able to comprehend fresh modifications, which are all inescapable in real-world scenarios.
3. The suggested joint model can directly and automatically structure and learn spatiotemporal characteristics from the unprocessed sensor data without requiring manual feature extraction.
4. This technique may be the most effective state-of-the-art and can probably be used across various platforms and domains.

The remaining paper sections are arranged as follows. Section 2 contains related work. The materials and procedures used in the proposed strategy are illustrated in Section 3. Our experimental setup for the activity recognition method is covered in Section 4 of this article. In Section 5, the activity recognition performance analysis is explained. Finally, a conclusive summary is provided in Section 6.

## 2. Related Work

HAR refers to a set of techniques used to automatically identify the task humans are executing by examining the video, readings from wearable sensors, or wireless signals reflected by the human body [21]. Shallow learning and deep learning techniques can categorize the HAR algorithms. SVM [22], k-nearest neighbors (kNN) [23], linear discriminant analysis (LDA) [24], and random forest [25] are examples of popular shallow HAR approaches. By learning to extract features from raw sensor data automatically, deep learning approaches, such as LSTM [26], CNN [27], convLSTM [28], and CNN-LSTM [29] have demonstrated impressive improvements in performance compared to their shallow counterparts. These eliminate the need for human experts to provide hand-engineered features.

The decline of cross-subject performance and change in activity schedule [30] is a significant obstacle when using deep learning for HAR. When testing the trained deep learning models on individuals not included in the training set, the difference in data distribution between the training and testing sets frequently results in considerable performance

degradation since different subjects carry out the same tasks in different ways [31]. An ideal training set would consist of data collected from tens or even hundreds of additional participants in order to address this problem. However, gathering and classifying data is a tedious and time-consuming operation.

Domain adaption techniques are rooted in natural language processing and computer vision [32] and have recently drawn increased attention for HAR applications [33]. These strategies can be classified as shallow or deep models depending on the feature extraction technique. Transfer learning strategies aim to align statistics of particular features between the source and destination domains for shallow models. Although DNN-based techniques use intermediate representations that are automatically learned by DNNs rather than manually created features, they nevertheless aim to achieve feature alignment. Several adversarial learning-based methods have recently tried to implement the new activity detecting process [34] by automatically identifying characteristics unaffected by the domain mismatch and (ii) useful for categorizing a particular activity. Model ensemble and feature concatenation have been proposed in earlier works [35]. for multi-sensor fusion [36]. However, these works combine a predetermined set of sensors without considering scenarios in which the fusing of several sensor configurations is required. Missing data imputation is a common issue focusing on completing the input space's unobserved activity data. Wearable sensor data typically involves multiple body locations over an extended time period and is highly dimensional. Furthermore, the majority of deep-learning based categorization models for this type of data offer little to no interpretability for the expected result. Some progress has been made in the challenge of video-based action recognition [37]. The auto-encoder is one class of neural networks capable of learning a condensed representation of the input signals. Ref. [38] suggested encoding high-dimensional continuous data as low-dimensional data by using auto-encoders with several hidden layers, so the features are retrieved. Ref. [39] features LightGBM as the classifier with stacking denoising auto-encoder for feature extraction. For instance, the learned features can be stacked using a stacked auto-encoder, which can then be used to create a classification model [40]. In their continuous auto-encoder proposal, unsupervised outlier detection can also be carried out using ensemble learning. To save on computing costs, ref. [41] presented an ensemble auto-encoder randomly connected with various architectures and connection densities.

### 3. Materials and Proposed Method

Systems for recognizing human activity go through data collection, pre-processing, feature extraction, training, and recognition. A similar procedure is also used in our approach, but the motivating aspect is new to HAR as of yet. Figure 1 depicts the process flow for our suggested technique. First, segmentation and filtration are used to pre-process the sensor data. Then, the feature is automatically extracted as we employ the deep learning model. The activity is then trained, classified, and recognized. Finally, we reprocessed and categorized activity that was not annotated. Semi-supervised learning is the term for this method. The auto-encoder model is what we are employing for this method. As a mechanism for adversarial learning, we add some perpetuation to the network to help it create its immune system. So, using the auto-encoder model and semi-supervised adversarial learning, we describe a method to identify human activities.

#### 3.1. Semi-Supervised Learning

Learning data and labels under supervision is a technique used in many situations or domains. When solving complex problems, supervised learning employs labeled data to learn and gain knowledge [42]. The supervised learning process has been implemented using a variety of deep learning and machine learning techniques. To train, however, hundreds to millions of learning data points may be given, and categorizing each point is crucial. Due to these problems, supervised learning cannot be applied in the absence of enough learning data. This problem can be solved by semi-supervised learning. It is a method for identifying unlisted data with crucial criteria, such as thresholds, and

re-learning models using learning data that is already accessible to improve performance based on the anticipated values of the learned sequences. The semi-supervised approach lessens manual annotation while creating a self-learning model that eventually builds up a solid body of knowledge and improves the recognition model's efficiency or accuracy. The collection of  $n$  labeled data points.

$$d_N = (x_i, y_i)_{i=1, \dots, n}^l \quad (1)$$

$x_i, y_i$  consists of data  $x_i \in \mathfrak{R}$  from a given sensor space  $\mathfrak{R}$ . We also have access to the extraction of  $m$  data whose labels are unknown.

$$d_M = (x_i)_{i=l+1, \dots, l+m}^{l+u} \quad (2)$$

### 3.2. Auto-Encoder

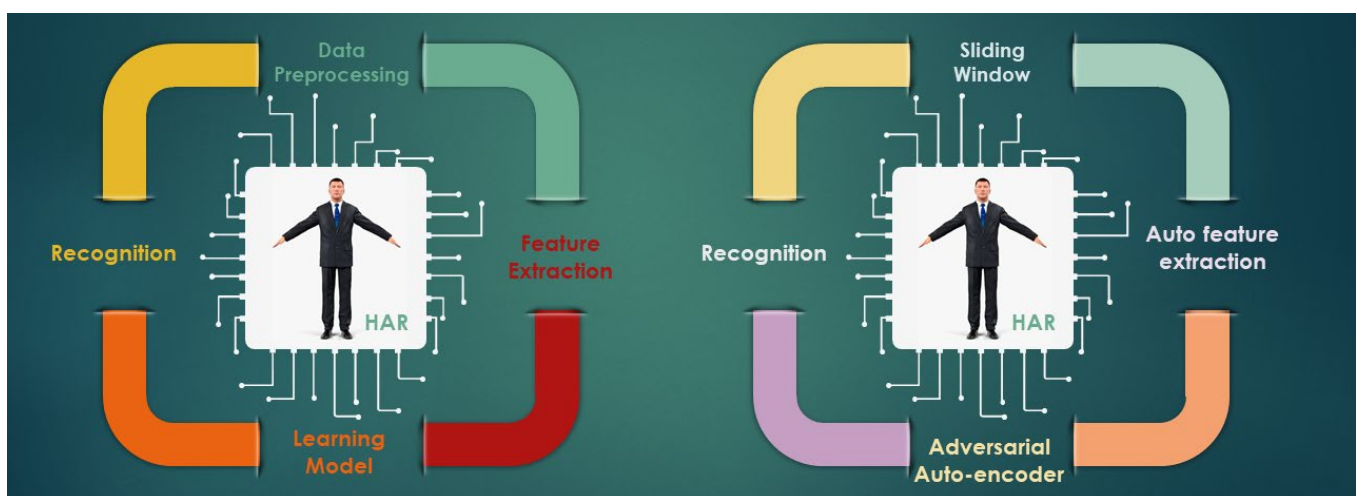
A fundamental AE is a neural network model in which the output replicates the input. The encoder and decoder are the two components that make up an AE. The encoder develops the ability to condense the inputs into a smaller subset of encoded features, or the bottleneck. The decoder learns how to recreate the original input given the encoded features. Consequently, an AE's output is a rough reconstruction of its input. Formulating the encoder phase is Equation (3), where  $W$  is the weight matrix and  $b$  is the bias vector for the encoder phase. The decoder phase is expressed in Equation (4), for the labeled data. Furthermore, Equations (5) and (6) represent the unlabeled data.  $W$  is the weight matrix and  $b$  is the bias vector. In this work, the sigmoid function is used as the nonlinear activation function denoted by the letters  $f$  in Equations (1) and (2). In the paragraphs that follow, we define  $x, h$  and  $\bar{x}$  as input layer, hidden layer and output (is approximately the same as the output of  $x$ ), respectively.

$$h_{d_N} = f(W_i d_N) + b_i \quad (3)$$

$$\bar{x}_{d_N} = f(W_i h_{d_N}) + b_i \quad (4)$$

$$h_{d_m} = f(W_i d_m) + b_i \quad (5)$$

$$\bar{x}_{d_m} = f(W_i h_{d_m}) + b_i \quad (6)$$



**Figure 1.** System workflow of our proposed method for HAR.

### 3.3. Adversarial Learning

With the addition of minute disturbances or noises to the training data, adversarial learning is a method to regularize neural networks that enhances their ability to predict the

future or approaches to deep learning by increasing the loss of a more profound learning model. However, according to [43], even minor changes to the deep learning input could produce very confident wrong decisions. Therefore, the following terms are added by adversarial learning to its cost function during the training phase of a predictive model where  $x$  and  $y$  are the input and two distinct parameters.

$$\log p(Y_t^s | X_t^s + r_t^s; \theta) = \text{where } r_t^s = \text{argmin} \log p(Y_t^s | X_t^s + r_t^s; \hat{\theta}) \quad (7)$$

According to Equation (7),  $r$  in the input data is hostile. A set of the recognition model's constant parameters  $\hat{\theta}$  is inherited from the  $\theta$ . The suggested algorithm recognizes the worst-case perturbations  $r_t^s$  at each training. In opposition to the present trained model, adversarial training generates disturbances or random noise that are easily misclassified in the learning model by changing the input instances, in contrast to other regularization strategies such as dropout.

The Algorithm 1 illustrate the pseudo-code for the overall process of our proposed method. Based on this algorithm we performed the experiment using the python coding.

---

**Algorithm 1:** Semi-supervised auto-encoder model with adversarial training

---

Step 1. Initialize the network

Step 2. Reset: inputs = 0, activations = 0

Step 3. Initialize the inputs

Step 4. Create encoder and decoder

$$h_{d_N} = f(W_i d_N) + b_i$$

$$\bar{x}_{d_N} = f(W_i h_{d_N}) + b_i$$

Step 5. Predict and calculate the loss function

a. Calculate seq2seq loss

b. Calculate class loss using cross-entropy

Step 6. Add random perturbations,

$$\log p(Y_t^s | X_t^s + r_t^s; \theta) = \text{where } r_t^s = \text{argmin} \log p(Y_t^s | X_t^s + r_t^s; \hat{\theta})$$

Step 7. Calculate loss function by adding adversarial loss

Step 8. Optimize the model based on AdamOptimizer

Step 9. Recognize unlabeled data based on Algorithm 1

$$h_{d_m} = f(W_i d_N) + b_i$$

$$\bar{x}_{d_m} = f(W_i h_{d_m}) + b_i$$

Step 10. Add recognized dataset to original training dataset

Step 11. Retrain the model

---

#### 4. Experimental Configuration

This section presents the complete results for both training and recognition. The assignment and processing of numerous design hypotheses come first. The proposed model is then trained using labeled and unlabeled data; the outcomes are compared to the outputs of the currently existing models. Finally, the experimental examination of the suggested approach is carried out using the CASAS dataset. This publicly available dataset can be downloaded free from the UCI Machine learning Repository.

To obtain the HAR dataset, a series of tests were run. For this work, 30 individuals with ages ranging from 19 to 48 were chosen. Each participant was given instructions on how to conduct themselves while sporting a Samsung Galaxy S II smartphone on their waist. The six chosen ADLs were *walking*, *walking upstairs and downstairs*, *sitting*, *lying down*, and *standing*. Each participant went through the process twice: on the first trial, the smartphone was fixed to the left side of the belt; on the second, the user chose where to put it. Additionally, there is a 5 s break between each task where people are instructed to rest. This promotes repeatability (every activity is attempted at least twice) and relaxation.

#### 4.1. Parameter Setting

The suggested technique was trained and tested using scikit-learn and the TensorFlow GPU1.13.1 library. The resulting data was pre-processed and sampled in sliding windows that overlapped and had a fixed width of 200 ms and a window length that ranged from 0.25 s to 7 s. Our technique was tested using an i7 CPU with 16 GB of RAM, a GTX Titan GPU running on CUDA 9.0, and the cuDDN 7.0 library. To use as little memory as possible, the CPU and GPU were utilized. A training set, a validation set, and a testing set comprised the three components of the dataset. The remaining 30% of data was used for testing, with the remaining 70% going toward training. The k-fold CV was used to validate the data (cross-validation). To verify, we employed 10-fold cross-validation (K = 10).

$$V = \frac{1}{p} \sum_{p=1}^{10} E(\text{error}) \quad (8)$$

In order to reduce overfitting, the dropout rate was adjusted during training to 0.5, removing unneeded neurons from each hidden layer. Training loss can also be decreased by using random initialization and optimizing training parameters. Cross-entropy and L2 normalization were incorporated to prevent overfitting and make the model stable.

$$L = -\frac{1}{k} \sum_{k=1}^n y_t^m \cdot \log y_t^{s'} + \Gamma \cdot \|W\|, \quad (9)$$

where  $W$  stands for the weighting parameter and  $k$  for the batch size, the label is  $y_t^m$ ; and the recognized output is  $y_t^{s'}$ . By reducing the amount of the weighting parameters, L2 normalization avoids overfitting. Adding minute disturbances or noises to the network with training data increases the loss of a more profound learning model for regularization that improves the recognition ability. Adversarial training is a technique for regularizing neural networks that enhances the neural network's prediction performance and may even approach deep learning. If the adversarial input is given by  $r_t^s$ , then the perturbations are given by, which is written as

$$r_t^s = \operatorname{argmin} \log p(Y_t^s | X_t^s + r_t^s; \hat{\theta}) \quad (10)$$

In order to achieve the most significant performance, we aimed to select the optimal hyperparameters, such that the learning rate, L2 weight, and difference all decreased. We utilized a learning rate of 0.005 with a batch value of 100 for each epoch to train the model. Learning begins at 0.001. The training is completed when the outputs are stable, which takes about 12,000 epochs. The Adam optimizer is a parameter-free adaptive moment estimator that produces adaptive learning rates. There are two dimensions: 128 for the input and 256 for the output, and 8 hidden layers. Gradient clipping was changed to 5 to lower the gradient crossing threshold. An Adam optimizer was used.

#### 4.2. Evaluation Parameter Setting

The model's performance was assessed using accuracy, *F1-score*, and training duration. The confusion matrix, where the row denotes the anticipated class and the column indicates the actual class, can be used to calculate these. The computational recognition accuracy of human activity recognition was assessed using the precision and recall parameters. The percentage of correctly identified instances from perceived activity occurrences is known as precision. A recall is the percentage of cases that were identified adequately out of all the instances. The weighted average of precision and recall between [0, 1] is known as an *F-score*, and a number closer to 1 indicates the more incredible performance

$$\text{Precision} : \frac{1}{N} \left( \sum_{m=1}^N \frac{TP_m}{TP_m + FP_m} \times 100 \right) \quad (11)$$



$$Recall : \frac{1}{N} \left( \sum_{m=1}^N \frac{TP_m}{TP_m + FN_m} \times 100 \right) \quad (12)$$

$$F - score : \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (13)$$

$$Accuracy = \frac{1}{N} \left( \sum_{m=1}^N \frac{TP_m + TN_m}{TP_m + TN_m + FP_m + FN_m} \times 100 \right) \quad (14)$$

Through the confusion matrix, these terminologies were evaluated for true positive ( $TP$ ), false positive ( $FP$ ), and false-negative ( $FN$ ) results. Each dataset was divided into three groups: a training set, a validation set for parameter optimization, and a test set for final assessment.

### 5. Activity Recognition Results, Analysis and Evaluation

The experimental findings are discussed and examined in this section. The dataset was used to locate all actions. In the context of a smart home, the contemporaneous and interspersed activities that happen the most frequently are referred to as presiding activities. Concurrent activity recognition, interleaved activity recognition, and the recognition average are the three sections that make up the analysis section. The general recognition accuracy is then contrasted with the other state-of-the-art approaches, including the CNN [44], the long short-term memory (LSTM) [45], and synchronized long short-term memory (Syn-LSTM) [46].

The recognition confusion matrix of the proposed method is shown in Figure 2. According to the confusion matrix, the average  $F$ -score for recognition is over 0.98, indicating that the average accuracy is high and desirable.

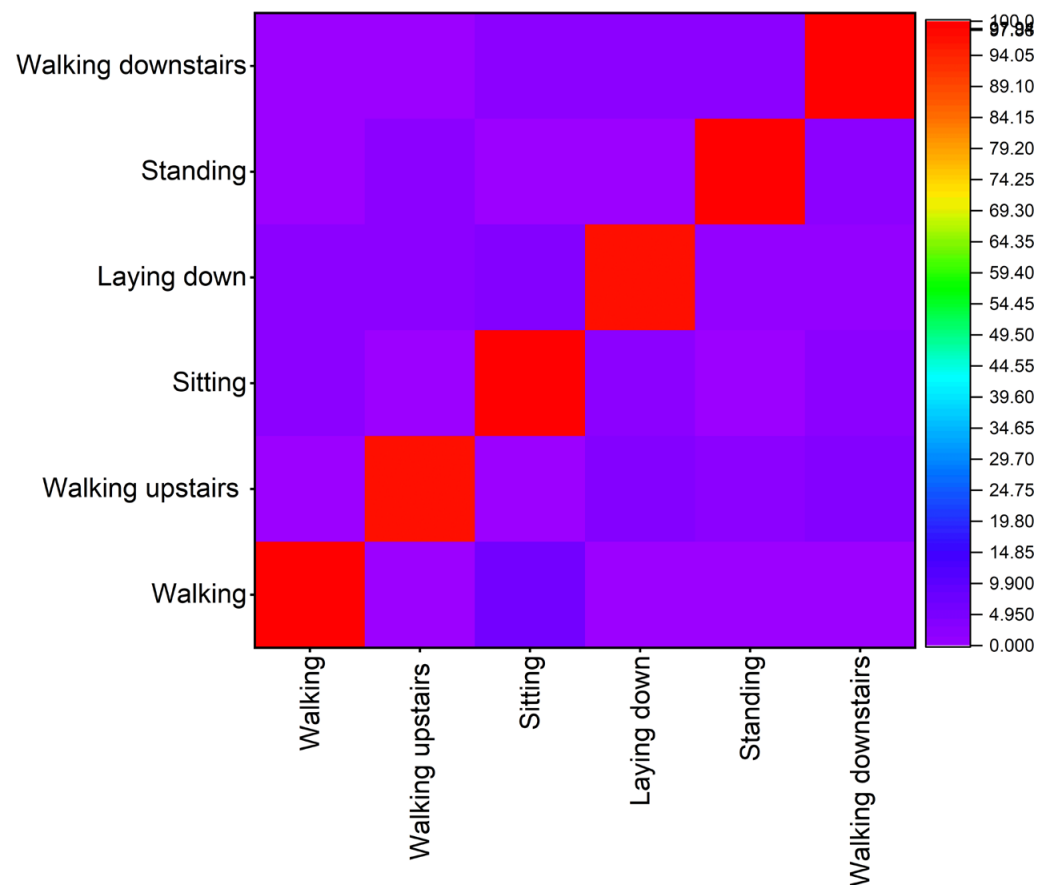
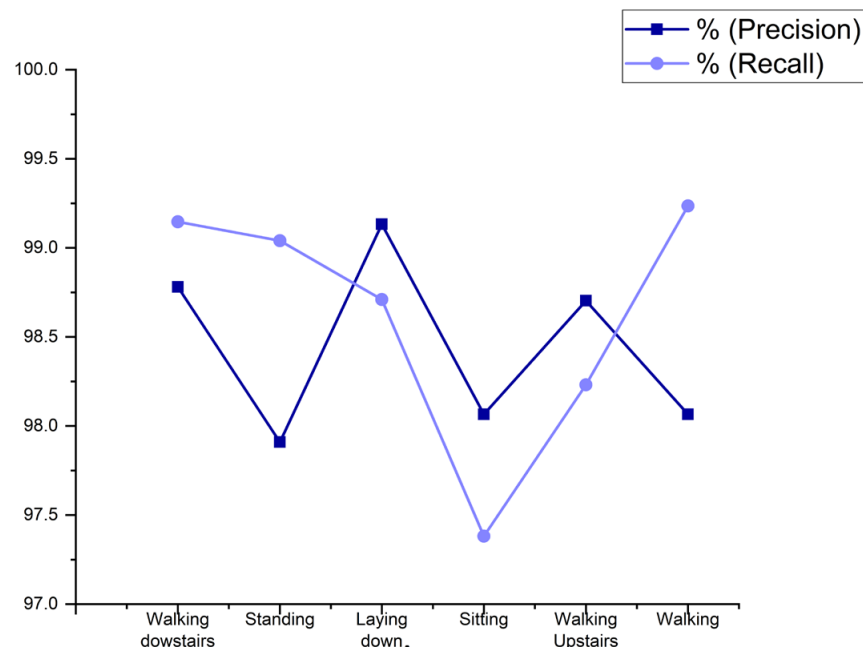


Figure 2. Confusion matrix accuracy.

As previously said, the accuracy is generally decent despite the sparse data. The walking activity to be true is 98% walking upstairs and walking downstairs is 97% and 98%, respectively. The sitting and the lying down possess similar accuracy with 97% and 96%. The sitting and lying down signal changes are very similar until and unless going into depth-hidden layers. Standing activity recognition accuracy is almost 99%. Walking and walking upstairs/downstairs are inherently more accessible to recognize than other activities as their signal changes are more or equal to the walking threshold. This dataset only has limited data and instances, so it is a bit easy to recognize, and accuracy is high enough, i.e., a large number of datasets could obtain the actual and accurate recognition distribution. The main aim of the proposed method is to recognize human activity and find and prove the algorithm.

Regarding analysis, the proposed algorithm is more reliable and competent than the existing method. The Figure 3 represents the percentage of cases that were identified correctly as activity instances. The precision of 98.4420% and recall of 98.6231% is received from the proposed model to determine the activity correctly.



**Figure 3.** The precision and the recall on the given set of data.

We have performed the validation to minimize the error and remove the unwanted activity that leads to increase accuracy. The 10-fold cross-validation is used for the model validation, whose deviation is measured referring epochs and batch sizes. Table 1 displays the mean and standard deviation after 10,000 iterations of changing the batch size hyperparameter. Similar results are shown in Table 2 for the epochs parameter when the batch size is 100, and the mean and standard deviation are computed. The choice of window size is also important for system accuracy; technically, a window size range of 500 ms to 5000 ms will be useful. Table 3 displays the mean and standard deviation of accuracy and error.

**Table 1.** Mean and standard deviation of variable range batch size on 10,000 epochs.

Epochs	1000	5000	8000	10,000
	Mean ( $\mu$ ) $\pm$ SD ( $\sigma$ )	Mean ( $\mu$ ) $\pm$ SD ( $\sigma$ )	Mean ( $\mu$ ) $\pm$ SD ( $\sigma$ )	Mean ( $\mu$ ) $\pm$ SD ( $\sigma$ )
	0.9300 $\pm$ 0.0125	0.9245 $\pm$ 0.0321	0.9386 $\pm$ 0.0621	0.9421 $\pm$ 0.0221



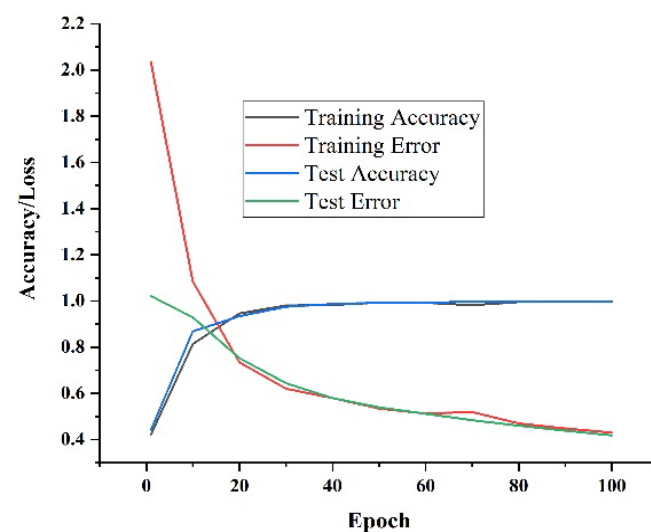
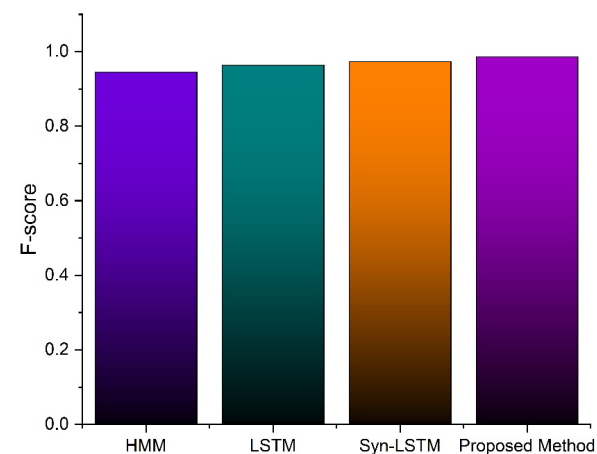
**Table 2.** Mean and standard deviation of different epochs on 100 batch size.

Batch Size	10	30	60	100
	Mean ( $\mu$ ) $\pm$ SD ( $\sigma$ )	Mean ( $\mu$ ) $\pm$ SD ( $\sigma$ )	Mean ( $\mu$ ) $\pm$ SD ( $\sigma$ )	Mean ( $\mu$ ) $\pm$ SD ( $\sigma$ )
	0.9561 $\pm$ 0.065	0.9602 $\pm$ 0.0423	0.9631 $\pm$ 0.04125	0.9531 $\pm$ 0.0431

**Table 3.** 10-fold cross-validation result.

	Mean ( $\mu$ ) $\pm$ SD ( $\sigma$ ) Accuracy	Mean ( $\mu$ ) $\pm$ SD ( $\sigma$ ) Error
UCI	0.9410 $\pm$ 0.0522	0.3215 $\pm$ 0.0121

Figure 4 depicts the accuracy and loss curve. The graphs' relatively small difference between training and testing accuracy demonstrates the model's efficacy. The dropout approaches, adversarial training, and semi-supervised learning are advantageous since the difference between training and test loss is also relatively small: Method 98.34, as proposed with an average inaccuracy of 0.1571, had an average accuracy of 98.154%, illustrating the effectiveness of the suggested strategy in comparison to the current framework, including the HMM [33], LSTM [34], and sync-LSTM [35] techniques (algorithms). The *F1-score* is higher than 0.98, as shown in Figure 5. Although the sync-LSTM is equally accurate to our approach, it is unable to handle fresh or unannotated data.

**Figure 4.** Accuracy and loss curve of training and testing.**Figure 5.** *F1-Score* comparison with an existing method.

## 6. Conclusions

By thoroughly comparing a semi-supervised adversarial auto-encoder with recently introduced activity recognition techniques such as deep learning and its variants, the work presented in this paper demonstrates a workable solution for detecting human activities. However, these strategies do not address the novel and new data in the sequence. On annotated and routine activity detection, many methods have been studied. Few of them, nevertheless, have made an effort to find intricate and unannotated activity. The proposed method recognized unannotated human behavior from the data gathered from the sensors by semi-supervised learning capability. The adversarial learning technique improves learning capacity by introducing slight disturbances or noises to the network. The challenges in recognizing human activity still include accuracy, processing complexity, complex activity, and unannotated activity. Nevertheless, the accuracy is 98%, and the precision and recall are also high, yielding an f1 score of greater than 0.98.

However, due to sensor timing, noise interference, and limited data, the accuracy is not equal or tends to be 100%. The current best-performing model encounters many real-time difficulties while interacting with various datasets. Essential factors affecting model performance include the number of activities carried out, sensor kinds, sensor deployment, population size, and time periods. Since tiny windows might not contain all the information and wide windows might result in overfitting and overload, window size also significantly impacts model performance. Identifying and processing the unannotated data is advantageous for extremely unbalanced datasets.

The suggested method uses reduced pre-processing time and manual feature extraction to automatically extract spatio-temporal information and identify unannotated activity. The proposed method can be improved and upgraded in the future to distinguish more complicated, multiuser, and multivariate actions. Additionally, we may benefit from edge computing, cloud computing, and IoT services to process a lot of data efficiently. Finally, different settings and domains, such as sign language identification, cognitive capacities, etc., can be employed in our approach. As a result, the strategy we recommend is a better, state-of-the-art one for HAR.

**Author Contributions:** Conceptualization, K.T.; methodology, K.T. and K.K.; software, K.T., K.K. and Y.S.; validation, K.T., S.-H.Y., Y.S. and K.K.; formal analysis, K.T. and K.K.; investigation, K.K.; resources, K.T., Y.S. and S.-H.Y.; data curation K.T. and K.K.; writing—original draft preparation, K.T.; writing—review and editing, K.K., Y.S. and S.-H.Y.; visualization, K.T.; supervision, K.K. and S.-H.Y.; project administration, K.K.; funding acquisition, K.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This manuscript is based upon work supported by Basic Science Research Program through the National Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2020R1C1008728).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This research is also supported by the excellent researcher support project of Daegu Haany University.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ware, S.; Yue, C.; Morillo, R.; Lu, J.; Shang, C.; Bi, J.; Kamath, J.; Russell, A.; Bamis, A.; Wang, B. Predicting Depressive Symptoms Using Smartphone Data. *Smart Health* **2020**, *15*, 100093. [[CrossRef](#)]
2. Rastegari, E.; Ali, H. A Bag-of-Words Feature Engineering Approach for Assessing Health Conditions Using Accelerometer Data. *Smart Health* **2020**, *16*, 100116. [[CrossRef](#)]
3. Milanko, S.; Jain, S. Liftright: Quantifying Strength Training Performance Using a Wearable Sensor. *Smart Health* **2020**, *16*, 100115. [[CrossRef](#)]

4. Shi, X.; Li, Y.; Zhou, F.; Liu, L. Human activity recognition based on deep learning method. In Proceedings of the 2018 International Conference on Radar (RADAR), Brisbane, Australia, 13–31 August 2018; pp. 1–5. [\[CrossRef\]](#)
5. Das, A.; Jens, K.; Kjærgaard, M.B. Space utilization and activity recognition using 3D stereo vision camera inside an educational building. In Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and 2020 ACM International Symposium on Wearable Computers (UbiComp/ISWC '20 Adjunct), Virtual Event, 12–16 September 2020; ACM: New York, NY, USA, 2020.
6. Thapa, K.; Abdullah Al, Z.M.; Lamichhane, B.; Yang, S.-H. A Deep Machine Learning Method for Concurrent and Interleaved Human Activity Recognition. *Sensors* **2020**, *20*, 5770. [\[CrossRef\]](#)
7. Van Kasteren, T.L.; Englebienne, G.; Kröse, B.J. An Activity Monitoring System for Elderly Care Using Generative and Discriminative Models. *Pers. Ubiquitous Comput.* **2010**, *14*, 489–498. [\[CrossRef\]](#)
8. Nweke, H.F.; Teh, Y.W.; Al-Garadi, M.A.; Alo, U.R. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Syst. Appl.* **2018**, *105*, 233–261. [\[CrossRef\]](#)
9. Yao, S.; Hu, S.; Zhao, Y.; Zhang, A.; Abdelzaher, T. Deepsense. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017.
10. Lara, O.D.; Labrador, M.A. A Survey on Human Activity Recognition Using Wearable Sensors. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 1192–1209. [\[CrossRef\]](#)
11. Mannini, A.; Intille, S.; Rosenberger, M.; Sabatini, A.; Haskell, W. Activity Recognition Using a Single Accelerometer Placed at the Wrist or Ankle. *Med. Sci. Sport. Exerc.* **2013**, *45*, 2193–2203. [\[CrossRef\]](#)
12. Qin, X.; Chen, Y.; Wang, J.; Yu, C. Cross-Dataset Activity Recognition via Adaptive Spatial-Temporal Transfer Learning. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 1–25. [\[CrossRef\]](#)
13. Khan, M.A.; Roy, N.; Misra, A. Scaling human activity recognition via deep learning-based domain adaptation. In Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), Athens, Greece, 19–23 March 2018.
14. Fang, L.; Ye, J.; Dobson, S. Discovery and Recognition of Emerging Human Activities Using a Hierarchical Mixture of Directional Statistical Models. *IEEE Trans. Knowl. Data Eng.* **2020**, *32*, 1304–1316. [\[CrossRef\]](#)
15. French, R. Catastrophic Forgetting in Connectionist Networks. *Trends Cogn. Sci.* **1999**, *3*, 128–135. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Hossain, H.M.S.; Roy, N.; Al Hafiz Khan, M.A. Active learning enabled activity recognition. In Proceedings of the 2016 IEEE International Conference on Pervasive Computing and Communications (PerCom), Sydney, Australia, 14–18 March 2016.
17. Ye, J.; Dobson, S.; Zambonelli, F. Lifelong Learning in Sensor-Based Human Activity Recognition. *IEEE Pervasive Comput.* **2019**, *18*, 49–58. [\[CrossRef\]](#)
18. Kabir, M.H.; Hoque, M.R.; Thapa, K.; Yang, S.-H. Two-Layer Hidden Markov Model for Human Activity Recognition in Home Environments. *Int. J. Distrib. Sens. Netw.* **2016**, *12*, 4560365. [\[CrossRef\]](#)
19. Oh, S.; Ashiquzzaman, A.; Lee, D.; Kim, Y.; Kim, J. Study on Human Activity Recognition Using Semi-Supervised Active Transfer Learning. *Sensors* **2021**, *21*, 2760. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Ponce, H.; Martínez-Villaseñor, L.; Miralles-Pechúan, L. A Novel wearable sensor-based human activity recognition approach using artificial hydrocarbon networks. *Sensors* **2016**, *16*, 1033. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Wang, W.; Liu, A.X.; Shahzad, M.; Ling, K.; Lu, S. Device-Free Human Activity Recognition Using Commercial WIFI Devices. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1118–1131. [\[CrossRef\]](#)
22. Moin, A.; Zhou, A.; Rahimi, A.; Benatti, S.; Menon, A.; Tamakloe, S.; Ting, J.; Yamamoto, N.; Khan, Y.; Burghardt, F.; et al. An EMG gesture recognition system with flexible high-density sensors and brain-inspired high-dimensional classifier. In Proceedings of the 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 27–30 May 2018.
23. Normani, N.; Urru, A.; Abraham, L.; Walsh, M.; Tedesco, S.; Cenedese, A.; Susto, G.A.; O'Flynn, B. A machine learning approach for gesture recognition with a lensless smart sensor system. In Proceedings of the 2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Las Vegas, NV, USA, 4–7 March 2018.
24. Jiang, S.; Lv, B.; Guo, W.; Zhang, C.; Wang, H.; Sheng, X.; Shull, P.B. Feasibility of Wrist-Worn, Real-Time Hand, and Surface Gesture Recognition via SEMG and Imu Sensing. *IEEE Trans. Ind. Inform.* **2018**, *14*, 3376–3385. [\[CrossRef\]](#)
25. Mummadi, C.K.; Leo, F.P.; Verma, K.D.; Kasireddy, S.; Scholl, P.M.; Van Laerhoven, K. Real-time embedded recognition of sign language alphabet fingerspelling in an IMU-based glove. In Proceedings of the 4th international Workshop on Sensor-based Activity Recognition and Interaction, Rostock, Germany, 21–22 September 2017.
26. Bao, L.; Intille, S.S. Activity recognition from user-annotated acceleration data most. In *Pervasive Computing*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 1–17.
27. Atarashi, K.; Oyama, S.; Kurihara, M. Semi-Supervised Learning from Crowds Using Deep Generative Models. *Proc. AAAI Conf. Artif. Intell.* **2018**, *32*, 11513. [\[CrossRef\]](#)
28. Kwon, H.; Abowd, G.D.; Plötz, T. Complex Deep Neural Networks from Large Scale Virtual IMU Data for Effective Human Activity Recognition Using Wearables. *Sensors* **2021**, *21*, 8337. [\[CrossRef\]](#)
29. Nair, R.; Ragab, M.; Mujallid, O.A.; Mohammad, K.A.; Mansour, R.F.; Vijju, G.K. Impact of wireless sensor data mining with hybrid deep learning for human activity recognition. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 9457536. [\[CrossRef\]](#)
30. Wan, S.; Qi, L.; Xu, X.; Tong, C.; Gu, Z. Deep Learning Models for Real-Time Human Activity Recognition with Smartphones. *Mob. Netw. Appl.* **2019**, *25*, 743–755. [\[CrossRef\]](#)

31. Thapa, K.; Mi, Z.M.A.; Sung-Hyun, Y. Adapted Long Short-Term Memory (LSTM) for Concurrent Human Activity Recognition. *Comput. Mater. Contin.* **2021**, *69*, 1653–1670. [[CrossRef](#)]
32. Duan, L.; Tsang, I.W.; Xu, D. Domain Transfer Multiple Kernel Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 465–479. [[CrossRef](#)] [[PubMed](#)]
33. Cook, D.; Feuz, K.D.; Krishnan, N.C. Transfer Learning for Activity Recognition: A Survey. *Knowl. Inf. Syst.* **2013**, *36*, 537–556. [[CrossRef](#)] [[PubMed](#)]
34. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *Domain Adapt. Comput. Vis. Appl.* **2017**, *17*, 189–209.
35. Bulling, A.; Ward, J.A.; Gellersen, H. Multimodal Recognition of Reading Activity in Transit Using Body-Worn Sensors. *ACM Trans. Appl. Percept.* **2012**, *9*, 1–21. [[CrossRef](#)]
36. Zappi, P.; Stiefmeier, T.; Farella, E.; Roggen, D.; Benini, L.; Troster, G. Activity recognition from on-body sensors by classifier fusion: Sensor scalability and robustness. In Proceedings of the 2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information, Melbourne, Australia, 3–6 December 2007.
37. Monfort, M.; Pan, B.; Ramakrishnan, K.; Andonian, A.; McNamara, B.A.; Lascelles, A.; Fan, Q.; Gutfreund, D.; Feris, R.S.; Oliva, A. Multi-Moments in Time: Learning and Interpreting Models for Multi-Action Video Understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 9434–9445. [[CrossRef](#)]
38. Wang, L. Recognition of Human Activities Using Continuous Autoencoders with Wearable Sensors. *Sensors* **2016**, *16*, 189. [[CrossRef](#)]
39. Gao, X.; Luo, H.; Wang, Q.; Zhao, F.; Ye, L.; Zhang, Y. A Human Activity Recognition Algorithm Based on Stacking Denoising Autoencoder and Lightgbm. *Sensors* **2019**, *19*, 947. [[CrossRef](#)] [[PubMed](#)]
40. Garcia, K.D.; Carvalho, T.; Mendes-Moreira, J.; Cardoso, J.M.; de Carvalho, A.C. A Study on Hyperparameter Configuration for Human Activity Recognition. *Adv. Intell. Syst. Comput.* **2019**, *950*, 47–56.
41. Chaurasia, S.; Goyal, S.; Rajput, M. Outlier detection using autoencoder ensembles: A robust unsupervised approach. In Proceedings of the 2020 International Conference on Contemporary Computing and Applications (IC3A), Lucknow, India, 5–7 February 2020.
42. Cozman, F.G.; Ira, C. Risks of Semi-Supervised Learning: How Unlabeled Data Can Degrade Performance of Generative Classifiers. In *Semi-Supervised Learning*; MIT Press: Cambridge, MA, USA, 2006; pp. 56–72.
43. Balabka, D. Semi-supervised learning for human activity recognition using adversarial autoencoders. In Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and 2019 ACM International Symposium on Wearable Computers, London, UK, 9–13 September 2019.
44. Ignatov, A. Real-Time Human Activity Recognition from Accelerometer Data Using Convolutional Neural Networks. *Appl. Soft Comput.* **2018**, *62*, 915–922. [[CrossRef](#)]
45. Hammerla, N.Y.; Halloran, S.; Ploetz, T. Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv* **2016**, arXiv:1604.08880.
46. Yang, S.-H.; Baek, D.-G.; Thapa, K. Semi-Supervised Adversarial Learning Using LSTM for Human Activity Recognition. *Sensors* **2022**, *22*, 4755. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.