Exploring the Benefits of Time Series Data Augmentation for Wearable Human Activity Recognition.

Md Abid Hasan
md.hasan@student.uni-luebeck.de
Institute for Medical Informatics,
University of Lübeck
Lübeck, Schleswig-Holstein
Germany

Frédéric Li
Institute for Medical Informatics,
University of Lübeck
Lübeck, Schleswig-Holstein
Germany

Artur Piet
Institute for Medical Informatics,
University of Lübeck
Lübeck, Schleswig-Holstein
Germany

Philip Gouverneur
Institute for Medical Informatics,
University of Lübeck
Lübeck, Schleswig-Holstein
Germany

Muhammad Tausif Irshad
Institute for Medical Informatics,
University of Lübeck
Lübeck, Schleswig-Holstein
Germany

Marcin Grzegorzek
Institute of Medical Informatics,
University of Lübeck
Lübeck, Germany
Fraunhofer IMTE
Lübeck, Germany

ABSTRACT

Wearable Human Activity Recognition (HAR) is an important field of research in smart assistive technologies. Collecting the data needed to train reliable HAR classifiers is complex and expensive. As a way to mitigate data scarcity, Time Series Data Augmentation (TSDA) techniques have emerged as a promising approach for generating synthetic HAR data. TSDA is not as trivial as image augmentation and has been relatively less investigated. In this paper, a comparative study of various state-of-the-art TSDA techniques is applied in the context of wearable HAR. More specifically, we investigate the classification of human activities on the *OPPORTUNITY* dataset [26] using a deep CNN-LSTM architecture trained on raw and synthetic data. Our study highlights the importance of TSDA on performance enhancement for multivariate multi-class datasets. Interestingly very simple time domain-based TSDA techniques notably outperform complex ones based on Generative Adversarial Networks. We provide practical advice on how to apply TSDA for imbalanced datasets in practice for generating the ideal amount of synthetic data to achieve optimal classification accuracy. Our TSDA-based approach outperforms the previous state-of-the-art [24] on the *OPPORTUNITY* dataset by 4.66% and 1.66% in average and weighted F1-scores, respectively.

KEYWORDS

Deep learning, Human Activity Recognition, Time Series Data Augmentation, Conditional GAN

iWOAR 2023, September 21–22, 2023, Lübeck, Germany

ACM ISBN 979-8-4007-0816-9/23/09…$15.00
https://doi.org/10.1145/3615834.3615842

ACM Reference Format:
Md Abid Hasan, Frédéric Li, Artur Piet, Philip Gouverneur, Muhammad Tausif Irshad, and Marcin Grzegorzek. 2023. Exploring the Benefits of Time Series Data Augmentation for Wearable Human Activity Recognition.. In *8th international Workshop on Sensor-Based Activity Recognition and Artificial Intelligence (iWOAR 2023), September 21–22, 2023, Lübeck, Germany.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3615834.3615842

1 INTRODUCTION

The recognition of human activities has become an important area of research in various scientific fields. By analyzing patterns in human behavior, researchers can gain valuable insights into social interactions, lifestyle habits, and other important aspects of human behavior. Activity recognition is also being used to monitor and track the physical activity levels, sleep patterns, and other behaviors of patients or study participants, providing valuable data that can be used to detect health issues and improve overall health and wellbeing[15, 16].

Recording data and labels for HAR is relatively straightforward compared to other applications of Ubicomp, but it still poses challenges when it comes to large-scale implementation. Numerous small datasets exist, but combining them proves difficult due to variations in activities, devices, sensor placement, and other factors [19, 30, 34]. Sensor noise, inter-subject variability, data imbalance, and complex activities are the primary obstacles in sensor-based human activity recognition (HAR) techniques.

Time series data augmentation has become increasingly popular in recent years as a means to improve the quantity of data and therefore the quality of the training [8]. The high dimensionality and sequential nature of the data are also minor challenges when working with time series data.[4]. To address them, researchers have proposed a variety of TSDA techniques, which aim to increase the size and diversity of the training data by generating synthetic data that is as similar as possible to the original data. To the best of our knowledge, the existing comparative studies of TSDA are performed on the UCR dataset repository only [7].

The major objective of this study was to investigate the impact of TSDA techniques on the classification of human activities using a widely known HAR benchmark OPPORTUNITY dataset [26]. To evaluate the quality of the synthetic data generated by different DA techniques, we adopted a Deep Convolution and LSTM recurrent neural network (CNN-LSTM) model proposed by Ordonez et al.[24] that has shown to achieve state-of-the-art performances for the classification of the 17 distinct human activities on OPPORTUNITY. We experimented with a variety of DA techniques (proposed and classified by Iwana et al. [17]), including Random Transformation, Pattern Mixing, and Generative Models, and generated a synthetic dataset. The findings from our study indicate that the performances of the model are enhanced after the inclusion of the augmented data in the training set. Furthermore, we proposed an effective strategy for generating a certain volume of synthetic data from the DA techniques to enhance the data quality and model performance. The contributions are as follows:

- A practical comparison of various state-of-the-art TSDA techniques is carried out on the OPPORTUNITY dataset to determine the best-performing TSDA method
- We provide some practical advice regarding how to apply TSDA
- our augmented approach outperforms the previous state-of-the-art by Ordonez et al.[24].

The remaining part of the paper proceeds as follows. A short description of the recent state of the art of TSDA techniques is presented in section 2. A brief description of selected methods and models is discussed in section 3. Section 4 is dedicated to the presentation of the obtained results.

2 RELATED WORK

Time series recognition tasks face the challenge of acquiring ample amounts of data. For instance, the UCR time series Archive of 2018 is one of the most well-known repositories of publicly available time-series datasets, yet only 12 out of 128 datasets contain more than a thousand training examples [7]. This highlights the difficulty of collecting large amounts of time-series data in practice. Therefore, TSDA techniques were introduced as a solution to create synthetic data. While data augmentation has become a standard practice in the image processing research community, in contrast, little attention has been paid to time series recognition [35]. In [17] Iwana et al. classified the TSDA methods into three main groups as previously mentioned, i.e. Random Transformations (RT), Pattern Mixing (PM), and Generative Models (GM). Each group was further divided according to magnitude, time, and frequency domains. The majority of TSDA techniques, including for example, jittering [2, 32], slicing or cropping [20], time warping [9], and convolve or frequency warping [29] is part of the RT family since they involve applying random transformations to the training data like for image data augmentation. Random transformation-based data augmentation is limited by the fact that not all transformations are compatible with every time series dataset(e.g. application of jittering is binary dataset).

An alternative to RT, such as synthesizing time series using inherent data information that can be achieved through methods such as pattern mixing, generative models, and pattern decomposition. The Synthetic Minority Oversampling Technique (SMOTE) is the most prominent method under the magnitude domain pattern mixing approach, specially designed to tackle imbalanced datasets by interpolating patterns from minority classes[6]. SMOTE has proven to be effective in numerous time series applications, including classification [37] and prediction [1]. Nevertheless, the oversampling technique can alter the distribution of the original data, leading to over-fitting. On the other hand, Dynamic Time Warping (DTW) is a method to determine the similarity between two temporal sequences [27]. In [18] Iwana et al. proposed DTW that combines a reference pattern with the selected (Teacher) pattern by DTW. Based on the selection of the teacher pattern, DTW is categorized into two variants namely Random Guided Warping (RGW) and Discriminative Guided Warping (GTW).

Unlike RT and PM, GMs enable the sampling of time series data from feature distributions and considered as most promising by the scientific community in recent days. There are two types of generative models: statistical models and neural network-based models. time series data augmentation and generation employ a diverse range of statistical, mathematical, and stochastic models. These approaches involve creating a statistical model of the data and are often applied in classification and forecasting. A theoretical study of the Markov chain Monte Carlo method for constructing efficient data-augmentation schemes is presented in [22]. The paper presents the conditional augmentation approach and the marginal augmentation approach and uses posterior sampling on a univariate dataset only. In [5], Cao et al. propose a new oversampling method for imbalanced time series classification using a mixture of Gaussian tree model. The proposed method is effective in correcting class imbalance and achieves competitive results compared to non-oversampling-based classifiers. The limitation of the proposed method is inapplicability for very large data sets due to its computational complexity. Generative Adversarial Networks (GANs) are a type of generative network that utilizes adversarial training to simultaneously optimize two neural networks: a generator and a discriminator [12]. While there have been many time series GANs suggested in the past, most are only designed for generating univariate time series data [28, 36]. Several attempts have been made to augment wearable HAR sensor signals [14, 21, 33]. However, the models proposed in the above studies are not unified conditional GANs but rather individual generators for each activity class. No previous study has investigated a unified conditional HAR GANs framework to capture the implicit distributions of diverse human activities. Part of the aim of this paper is to investigate the feasibility of a unified conditional HAR-GAN structure that is compatible to generate sensory signals for different human-like activities (i.e. walking, sitting, standing, the opening fridge, and door, etc).

3 METHODOLOGY

Several TSDA methods used in the past have been adapted from techniques originally used in image data augmentation such as Jittering, Slicing, Time-warping, Convolving, and Generative Adversarial Networks. When using data augmentation in the context of time series, and more generally in the field of signal processing, it is crucial to consider the potential impact of manipulating the data on the information originally contained in the signal. Excessive modification could result in a distorted signal, which would ultimately compromise the effectiveness of the training process. It is worth pointing out that the application of data augmentation to the time series domain, and more specifically signal processing, can cause signal distortion and result in negative training outcomes. Therefore, the type and extent of data augmentation used should be carefully considered to avoid compromising the quality and usefulness of the training data. Under the RT family magnitude domain transformations modify time series values by keeping time steps constant while in time domain transformations, elements of the time series are displaced to different time steps than the original sequence. Pattern mixing generates new examples by attempting to align one sample instance to another referred to as the reference, where alignment is usually performed by weighted averaging between the two samples. One of the most well-known magnitude pattern mixing approaches is SMOTE which interpolates a sample selected from the minority class and uses a reference sample chosen among its k nearest neighbors. However, its impact on high-dimensional time series data has not been extensively explored [3]. On the contrary, Guided Dynamic Time Warping (GDTW) combines reference instances with sample ones directly in the time domain[18]. In this study, we decided to focus on the time-domain-based PM methods. Under GM families, statistical generative models are frequently utilized for forecasting [17]. We, therefore, opted to employ a Generative Adversarial Network (GAN) for augmenting time series data in a classification task.

3.1 Random Transformation

The application of TSDA based on RT to time series data can be categorized into three primary domains: the magnitude domain (Jittering, Flipping), the time domain (Time-Warping, Slicing), and the frequency domain (Convolution). DA based on magnitude domain transformations involves modifying the values of time series data without altering the time steps. Jittering has been widely utilized with time-series sensor data and consists in adding Gaussian noise at each time step [5]. Flipping mimics the analog transform for image data. For univariate time series, the data is flipped over time. For multivariate time series, an element-wise random rotation of a given angle is performed. Slicing in TSDA is equivalent to cropping for images with symmetric padding applied to fill in the values that were removed. Time Warping, another time domain RT, is also referred to as window-warping (WW) and is used to stretch or compress the time axis of the original time-series data to create new time-series data with different temporal resolutions [25].

Transformations in the frequency domain pertain specifically to periodic signals. Frequency warping is a commonly used DA technique in the fields of audio and speech recognition. This involves introducing Gaussian noise to the amplitude and phase spectra obtained through a discrete Fourier transform (Convolve) [11]. All the five RT augmentation methods previously presented were tested in our experiments.

3.2 Pattern Mixing

Guided dynamic time warping is a novel time series data augmentation method that merges the concept of time warping with pattern mixing. This method builds upon Dynamic time warping (DTW) and involves the following steps:

- Choosing a reference instance and a sample instance from the dataset.
- Calculating the DTW distance between the reference and sample instances.
- Utilizing the DTW distance to guide the warping of the sample instance, aligning it with the time steps of the reference instance.
- Iterating this process for all instances in the dataset.

The reference instance is usually defined according to the two following methods. In [10, 31] authors proposed to select referenced patterns randomly and referred to it as Randomly Guided Warping (RGW). Instead of random selection, Iwana et al. proposed a guided reference pattern selection technique named Discriminative Guided Warping (DGW) that proposes to use class labels to select the reference instance as the instance that shares the same class as the sample instance and that is the further away from other class instances. Here we will apply both RGW and DGW to augment HAR data and present a comparative analysis [18].

3.3 Conditional time series HAR GAN

To perform DA, generative models offer an alternative to random transformations or pattern mixing where time series are sampled from feature distributions using either statistical models or neural network-based models. The aim of our study is classification rather than forecasting. Therefore, our focus will be on discussing generative models based on neural networks exclusively.
GAN was first proposed by Ian Goodfellow in [12] where he used two neural network models named Generator (G) and Discriminator(D). The primary objective of G is to produce synthetic signals, while the role of D is to continually verify whether the generated signal is real or fake. D and G are jointly trained in a competitive way, with the objective of leading G to create data realistic enough to fool D. The performance of both G and D is monitored via GAN losses (generator and discriminator loss). Ideally for perfect GAN, it is expected that after training the model both losses are 0.5. The GAN training process can be translated into the following optimization problem:

$$\begin{aligned} Min_G Max_D V(D,G) = \mathbb{E}_{x\sim p_{data(x)}}[logD(x)] \\ + \mathbb{E}_{z\sim p_{z(z)}}[log(1-D(G(z)))] \end{aligned} \tag{1}$$

Where V is an objective value function, p_{data} is the probability distribution of the original data. While $P_z(Z)$ represents a prior distribution of input noise. z is a random noise vector sent as input of G, and $D(x)$ is the probability that the sample x came from the original data rather than being generated by G. Later conditional information y was introduced in the standard GAN framework by Mirza et al. in [23] to generate an output conditioned on class labels. The modified value equation of Conditional GAN is

$$\begin{aligned} Min_G Max_D V(D,G) = \mathbb{E}_{x\sim p_{data(x)}}[logD(x|y)] \\ + \mathbb{E}_{z\sim p_{z(z)}}[log(1-D(G(z|y)))] \end{aligned} \tag{2}$$

By imposing the condition on labels (y), the output sequence of the imbalanced time series dataset can be balanced. To generate the HAR signal the objective function of GAN from wearable signals studies from [21, 33] is also followed in this study,

$$V_G = \frac{1}{m}\sum_{i=1}^{m} log(1 - D(G(z_i))) \quad (3)$$

$$V_D = \frac{1}{m}\sum_{i=1}^{m} logD(x_i) + log(1 - D(G(z_i))) \quad (4)$$

The objective of the generator is to reduce the value of V_G described in Equation 3, while the goal of the discriminator is to increase the value of V_D mentioned in Equation 4. Here, m represents the number of actual input samples.

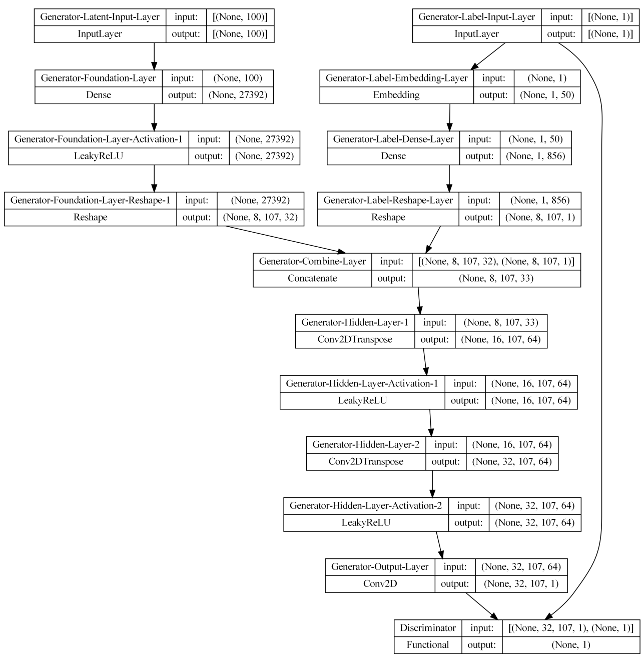

Figure 1: ctHAR-GAN Generator Architecture

Because past HAR studies have shown that regular GANs with a single generator could not generate convincing synthetic data for multiple classes (i.e stay, walk, jog) [21, 33] we construct unified conditional time series HAR GAN models (ctHAR-GAN) that consists a discriminator (figure 2) and unique generator (figure 1) structure to generate all 17 activities. In OPPORTUNITY dataset some minority classes (i.e. open and close drawers 1 and 2) have very few samples (figure 3) that inspire us to design a unique generator. The networks are mainly based on CONV2D, CONV2D transpose, and fully connected networks.

4 EXPERIMENTS AND RESULTS

This section outlines the assessments conducted to compare and evaluate the different data augmentation techniques. Our code was implemented in Python by utilizing the Keras (Version==0.11) framework with a TensorFlow (version==2.11.0) back-end. We used the well-known HAR benchmark dataset *OPPORTUNITY* [26] whose main characteristics are shown in table 1.

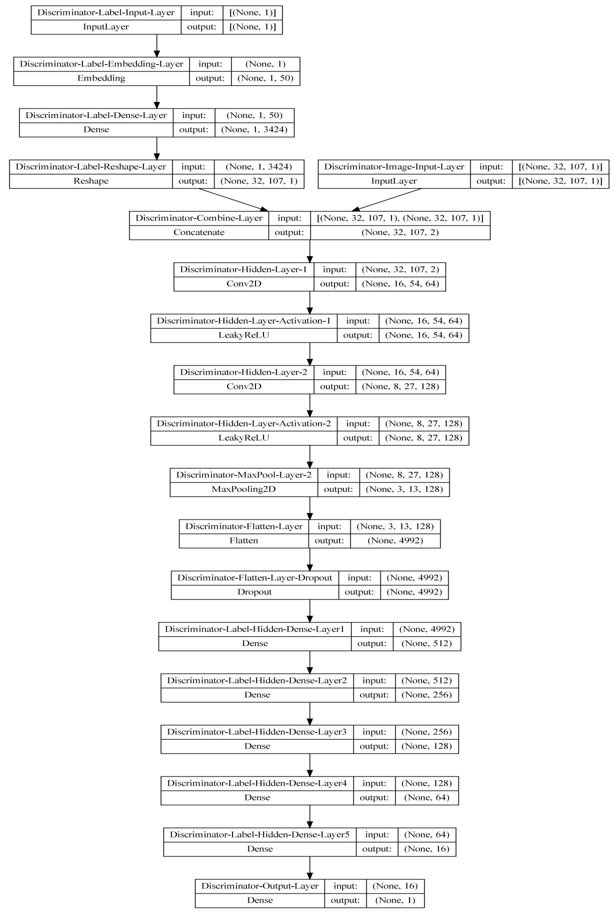

Figure 2: ctHAR-GAN Discriminator Architecture

To evaluate the effectiveness of the data augmentation methods we reproduced the experimental setup of Ordonez et al. [24] by using the same CNN-LSTM architecture they proposed and the same train and test data split. In particular, we also trained our models on segmented data frames of 1 second with a 50% overlap. We evaluated our model using both the weighted F_1 score and average F_1 score (also referred to as macro F_1 score). The OPPORTUNITY dataset is highly imbalanced with the Null class accounting for 72.28% of the entire dataset. We, therefore, use the average F_1 score macro as the main evaluation metric as it can provide an evaluation of the classifier performances unbiased by class imbalance. In this study, all selected DA techniques underwent three repetitions, and the final outcome is presented as the mean with standard deviation. Finding the proper amount of augmented data to be added to the dataset can be difficult, as it is required to maintain a balance between increasing the dataset size and not introducing too much bias (referred to as the synthetic-to-real gap) in the dataset. We, therefore, augmented the training dataset with multiple augmentation factors per class (minority) and present a comparative analysis in this paper.

Regarding the RT DA techniques, we tested each of the approaches described in Section 3 separately, as well as combinations

Table 1: Data Set Information

Data Set Characteristics	Multivariate, time series
Nr. of subjects	4
Nr. of runs per Subject	6
Annotations/Classes	17 mid-level gesture classes
Nr. of Sensor Channels (after denoising)	107
Sampling frequency	30Hz

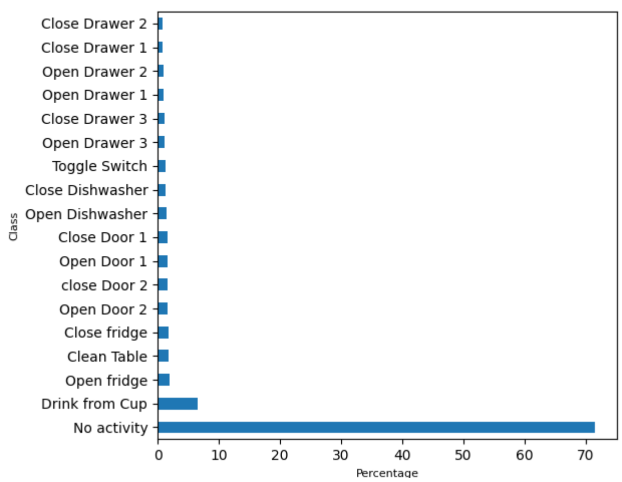

Figure 3: Class Distribution of Opportunity dataset.

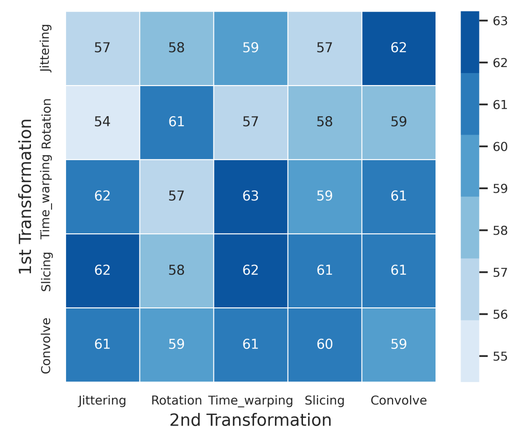

Figure 4: Linear evaluation (F_1 score macro) under individual and composition of RT data augmentation.

of two of them. Tunable parameters, the standard deviation(σ) and mean (μ) of Gaussian Noise were chosen based on the findings of Iwana et al.[17]. The classification results obtained on the testing set after augmenting the training set with the strategy previously described are provided in Figure 4. It can be seen that the time-warping DA leads to the best performances, with an average F_1 score macro of 63.22% and weighted F_1 score of 92%. For presenting comparative analysis with other TSDA we will consider time warping as RT DA.

Under the PM family, we evaluated time domain-based RGW and DGW, and the evaluation result is presented in Table 2. The best performance was obtained from DGW.

Our GAN's Loss function is shown in figure 5. Based on the figure, we can observe that even though the GAN is stable after 500 epochs, both the discriminator loss (≈ 0.57) and generator loss (≈ 0.75) deviate from the desired value of 0.5[12]. Consequently, it is not surprising at all that synthetic data from GAN is actually not improving the model's performance (Table 2).

Table 3 represents the performance measures for the TSDA approach on the Opportunity dataset. From the table 3, it can be seen that the Time Warping method outperforms all other approaches, including the baseline model (baseline is a reproduction of Ordonez et al. without augmentation[24]). Specifically, when compared to the baseline model, the F_1 score macro performance is boosted by over 4.66%.

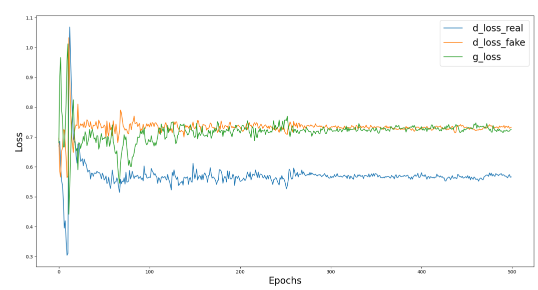

Figure 5: Training Loss of D and G

5 DISCUSSION

From the observation and analysis of the previous section, we can stipulate that TSDA improves the overall performance of the model (including both F_1 scores). Among the TSDA approaches, the simple random transformations (and more specifically time-warping) actually performed the best to generate more realistic HAR data. In this paper, the quality of the generated activity data was further evaluated using the Frechet Inception Distance (FID) score [13]. The FID score is commonly used to measure the difference between the feature distribution of the generated data and that of the real

Table 2: Average macro F_1 score(%) with standard deviation across selected DA algorithms by varying augmentation factors.

Augmentation factor	Time warping	RGW	DGW	ctHAR-GAN
1	59.30±1.96	59.52±2.30	59.21±2.84	55.18±3.17
2	59.23±1.02	56.44±2.94	59.39±0.84	55.69±0.11
3	61.98±1.99	60.5±0.281	61.21±1.00	55.69±5.79
4	60.06±0.07	60.47±0.70	60.10±0.12	56.43±0.36
5	62.74±1.62	59.69±2.39	57.81±0.08	54.04±1.78
6	61.25±1.11	60.27±1.79	58.16±1.63	55.086±1.85
7	60.39±0.80	58.10±0.02	59.75±2.52	56.23±5.55
8	59.15±0.99	59.47±1.1	58.62±2.75	55.27±0.46
9	**63.22±0.22**	60.12±0.05	58.08±1.50	54.04±0.63
10	60.60±1.90	57.71±3.93	59.45±0.67	54.64±2.41

Table 3: Best weighted and average F1 scores on the OPPORTUNITY training set for each augmentation family

Augmentation method	Weighted F_1 score(%)	Average F_1 score(%)
Baseline (Without DA)	91.5	58.56
DGW	89.48	61.21
Time Warping	**93.13**	**63.22**
ctHAR-GAN	88.12	56.43

data. The mathematical definition of FID is as follows:

$$FID(x_r, x_g) = ||\mu_r - \mu_g||^2 + T_r(\psi_r + \psi_g - 2(\psi_r \psi_g)^{\frac{1}{2}}) \qquad (5)$$

In equation 5, $\mu_r \epsilon \mathbb{R}^s$ represents the average of the sensor channels (s) in the real data, while $\mu_g \epsilon \mathbb{R}^s$ represents the average of the sensor channels in the generated data. ψ_r stands for the covariance matrix of the real data features, and ψ_g stands for the covariance matrix of the generated data features. The FID score is calculated from real and synthetic training data. The quality of the generated data is inversely proportional to the FID score, meaning that the better the quality of the generated data, the lower the FID score will be.

To support our findings we also presented the FID score of each class in table 4. It is clear from the table that the FID score per class achieved through the Time Warping RT and DGW method is close and lower than that of ctHAR-GANs. This indicates that the Time Warping RT and DGW PM approach produces more realistic data.

6 CONCLUSION

In this study, we have conducted an extensive analysis of various TSDA methods specifically on HAR Sensor signals. Our experiments demonstrate that the utilization of TSDA techniques can enhance performance compared to not employing them. We explored both fundamental (such as RT and PM) and more advanced methods (like GAN). Regardless of the approach, synthetic data generated through time domain-based TSDA algorithms (i.e. TW and DGW) proved effective in improving model performance. Interestingly, our findings suggest that even simple TSDA techniques like TW outperform complex ones like conditional time series GAN. In the future, we will focus on optimizing the performance of our ctHAR-GAN. To generate more realistic data by the GAN, our focus will be to use a variational Auto Encoder to learn the pattern from the training dataset and provide a more resourceful input (latent space vector) to the generator.

ACKNOWLEDGMENTS

This study was supported in part by the Deutscher Akademischer Austauschdienst (Award No. 91831212) and Bundesministerium für Bildung und Forschung (Grant No. 16KISA074).

REFERENCES

[1] Suja A Alex, NZ Jhanjhi, Mamoona Humayun, Ashraf Osman Ibrahim, and Anas W Abulfaraj. 2022. Deep LSTM Model for Diabetes Prediction with Class Balancing by SMOTE. *Electronics* 11, 17 (2022), 2737.

[2] Chris M Bishop. 1995. Training with noise is equivalent to Tikhonov regularization. *Neural computation* 7, 1 (1995), 108–116.

[3] Rok Blagus and Lara Lusa. 2013. SMOTE for high-dimensional class-imbalanced data. *BMC bioinformatics* 14 (2013), 1–16.

[4] David R Brillinger. 2001. *Time series: data analysis and theory.* SIAM.

[5] Hong Cao, Vincent YF Tan, and John ZF Pang. 2014. A parsimonious mixture of Gaussian trees model for oversampling in imbalanced and multimodal time-series classification. *IEEE transactions on neural networks and learning systems* 25, 12 (2014), 2226–2239.

[6] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.

[7] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh. 2019. The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica* 6, 6 (2019), 1293–1305.

[8] Ankur Debnath, Govind Waghmare, Hardik Wadhwa, Siddhartha Asthana, and Ankur Arora. 2021. Exploring generative data augmentation in multivariate time series forecasting: opportunities and challenges. *Solar-Energy* 137 (2021), 52–560.

[9] Tonya Fields, George Hsieh, and Jules Chenou. 2019. Mitigating drift in time series data with noise augmentation. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 227–230.

[10] Germain Forestier, François Petitjean, Hoang Anh Dau, Geoffrey I Webb, and Eamonn Keogh. 2017. Generating synthetic time series to augment sparse datasets. In *2017 IEEE international conference on data mining (ICDM)*. IEEE, 865–870.

Table 4: Comparison of FID score of synthetic data on OPPORTUNITY

Activity	DGW	Time Warping RT	ctHAR-GAN
Open Door 1	2657.62	**2482.71**	3688.54
Open Door 2	**1068.16**	1382.25	6096.21
Close Door 1	**3495.66**	3910.30	4191.88
Close Door 2	46287.72	**4590.81**	5638.107
Open fridge	2456.42	**2334.60**	3567.15
Close fridge	2497.89	**2118.51**	3772.22
Open Dishwasher	**2544.60**	3048.8	3099.14
Close Dishwasher	**2810.78**	3063.46	3398.070
Open Drawer 1	2546.35	**2328.7**	4141.20
Close Drawer 1	447.30	**323.04**	4687.67
Open Drawer 2	**358.71**	493.71	4422.19
Close Drawer 2	523.50	**416.01**	4656.29
Open Drawer 3	654.20	**423.24**	3274.28
Close Drawer 3	3729.18	**3217.40**	3466.84
Clean Table	**1640.15**	1886.65	5169.010
Drink from Cup	**115.26**	151.50	1591.037
Toggle Switch	5995.38	**5041.99**	5629.57

[11] Jingkun Gao, Xiaomin Song, Qingsong Wen, Pichao Wang, Liang Sun, and Huan Xu. 2020. Robusttad: Robust time series anomaly detection via decomposition and convolutional neural networks. *arXiv preprint arXiv:2002.09545* (2020).

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.

[13] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* 30 (2017).

[14] Yifan Hu and Yu Wang. 2022. BSDGAN: Balancing Sensor Data Generative Adversarial Networks for Human Activity Recognition. *arXiv preprint arXiv:2208.03647* (2022).

[15] Zawar Hussain, Michael Sheng, and Wei Emma Zhang. 2019. Different approaches for human activity recognition: A survey. *arXiv preprint arXiv:1906.05074* (2019).

[16] Zawar Hussain, Quan Z Sheng, and Wei Emma Zhang. 2020. A review and categorization of techniques on device-free human activity recognition. *Journal of Network and Computer Applications* 167 (2020), 102738.

[17] Brian Kenji Iwana and Seiichi Uchida. 2021. An empirical survey of data augmentation for time series classification with neural networks. *Plos one* 16, 7 (2021), e0254841.

[18] Brian Kenji Iwana and Seiichi Uchida. 2021. Time series data augmentation for neural networks by time warping with a discriminative teacher. In *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 3558–3565.

[19] Charmi Jobanputra, Jatna Bavishi, and Nishant Doshi. 2019. Human activity recognition: A survey. *Procedia Computer Science* 155 (2019), 698–703.

[20] Arthur Le Guennec, Simon Malinowski, and Romain Tavenard. 2016. Data augmentation for time series classification using convolutional neural networks. In *ECML/PKDD workshop on advanced analytics and learning on temporal data*.

[21] Xi'ang Li, Jinqi Luo, and Rabih Younes. 2020. ActivityGAN: Generative adversarial networks for data augmentation in sensor-based human activity recognition. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*. 249–254.

[22] Xiao-Li Meng and David A Van Dyk. 1999. Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika* 86, 2 (1999), 301–320.

[23] Mehdi Mirza and Simon Osindero. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014).

[24] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16, 1 (2016), 115.

[25] Khandakar M Rashid and Joseph Louis. 2019. Window-warping: a time series data augmentation of IMU data for construction equipment activity identification. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, Vol. 36. IAARC Publications, 651–657.

[26] Daniel Roggen, Alberto Calatroni, Mirco Rossi, Thomas Holleczek, Kilian Förster, Gerhard Tröster, Paul Lukowicz, David Bannach, Gerald Pirkl, Alois Ferscha, et al. 2010. Collecting complex activity datasets in highly rich networked sensor environments. In *2010 Seventh international conference on networked sensing systems (INSS)*. IEEE, 233–240.

[27] Hiroaki Sakoe and Seibi Chiba. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing* 26, 1 (1978), 43–49.

[28] Kaleb E Smith and Anthony O Smith. 2020. Conditional GAN for timeseries generation. *arXiv preprint arXiv:2006.16477* (2020).

[29] Odongo Steven Eyobu and Dong Seog Han. 2018. Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network. *Sensors* 18, 9 (2018), 2892.

[30] Jubil T Sunny, Sonia Mary George, Jubilant J Kizhakkethottam, Jubil T Sunny, Sonia Mary George, and Jubilant J Kizhakkethottam. 2015. Applications and challenges of human activity recognition using sensors in a smart environment. *IJIRST Int. J. Innov. Res. Sci. Technol* 2 (2015), 50–57.

[31] Naoya Takahashi, Michael Gygli, Beat Pfister, and Luc Van Gool. 2016. Deep convolutional neural networks and data augmentation for acoustic event detection. *arXiv preprint arXiv:1604.07160* (2016).

[32] Terry T Um, Franz MJ Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. 2017. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In *Proceedings of the 19th ACM international conference on multimodal interaction*. 216–220.

[33] Jiwei Wang, Yiqiang Chen, Yang Gu, Yunlong Xiao, and Haonan Pan. 2018. SensoryGANs: An effective generative adversarial framework for sensor-based human activity recognition. In *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[34] Yan Wang, Shuang Cang, and Hongnian Yu. 2019. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Systems with Applications* 137 (2019), 167–190.

[35] Qingsong Wen, Liang Sun, Fan Yang, Xiaomin Song, Jingkun Gao, Xue Wang, and Huan Xu. 2020. Time series data augmentation for deep learning: A survey. *arXiv preprint arXiv:2002.12478* (2020).

[36] Jinsung Yoon, Daniel Jarrett, and Mihaela Van der Schaar. 2019. Time-series generative adversarial networks. *Advances in neural information processing systems* 32 (2019).

[37] Pu Zhao, Chuan Luo, Bo Qiao, Lu Wang, Saravan Rajmohan, Qingwei Lin, and Dongmei Zhang. 2022. T-SMOTE: temporal-oriented synthetic minority oversampling technique for imbalanced time series classification. In *Proceedings of IJCAI*.