

Received November 4, 2021, accepted November 18, 2021, date of publication November 23, 2021, date of current version December 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3130421

Multivariate Generative Adversarial Networks and Their Loss Functions for Synthesis of Multichannel ECGs

EOIN BROPHY^{1,2}, MAARTEN DE VOS^{3,4}, GERALDINE BOYLAN², AND TOMÁS WARD^{1,5}, (Senior Member, IEEE)

¹School of Computing, Dublin City University, Dublin, D09 Y074 Ireland

²INFANT Research Centre, University College Cork, Cork, T12 YN60 Ireland

³Department of Electrical Engineering, KU Leuven, 3001 Leuven, Belgium

⁴Department of Development and Regeneration, KU Leuven, 3000 Leuven, Belgium

⁵Insight SFI Research Centre for Data Analytics, Dublin City University, Dublin 9, D09 Y074 Ireland

Corresponding author: Eoin Brophy (eoin.brophy7@mail.dcu.ie)

This work was supported in part by the Science Foundation Ireland under Grant 17/RC-PhD/3482 and Grant SFI/12/RC/2289_P2, and in part by the Flemish Government (AI Research Program).

ABSTRACT Access to medical data is highly regulated due to its sensitive nature, which can constrain communities' ability to utilize these data for research or clinical purposes. Common de-identification techniques to enable the sharing of data may not provide adequate privacy in every circumstance. We investigate the ability of Generative Adversarial Networks (GANs) to generate synthetic, and more significantly, multichannel electrocardiogram signals that are representative of waveforms observed in patients to address these privacy concerns. Successful generation of high-quality synthetic time series data has the potential to act as an effective substitute for actual patient data. For the first time, we demonstrate a range of novel loss functions using our multivariate GAN architecture and analyse their effect on data quality and privacy. We also present the application of multivariate dynamic time warping as a means of evaluating generated time series. Quantitative evidence demonstrates that the inclusion of a penalisation coefficient (Dynamic Time Warping) in the loss function enables our GAN to outperform the other generative models and loss functions explored by 4.9% according to our metrics. This allows for the generation of data that is more representative of the training set and diverse across generated samples, all whilst ensuring sufficient privacy.

INDEX TERMS Generative adversarial networks, ECG, time series.

I. INTRODUCTION

Sharing and using inherently sensitive medical data is becoming increasingly complex, with tightening restrictions that lead to a significant challenge in clinical research and development. As a result, traditional modes of data sharing have become hampered, and efforts are being made by the artificial intelligence (AI) community to overcome these restrictions in ways that respect privacy sensitivities. This is a significant challenge because the development of effective AI requires access to extensive datasets. Such data privacy concerns present researchers and clinicians with an additional set of obstacles in their pursuit of AI-enhanced innovations.

The associate editor coordinating the review of this manuscript and approving it for publication was Tyson Brooks¹.

Addressing these obstacles raises ethical issues, especially regarding how data can be used while ensuring privacy is protected, and public trust is maintained. This poses policy and regulatory challenges for lawmakers and regulators. They must balance safeguarding personal data while not retarding vital innovation and research to improve patient outcomes.

Practitioners with access to sought after data often find themselves working through complex data privacy frameworks and discover that sharing and publishing the information available via the data is highly challenging. For example, personal sensitive data such as medical data intended for secondary purposes like clinical training or research requires anonymisation following its approval for dissemination.

Common methods for the de-identification of data are generalisation, randomisation, or pseudonymisation [1].

However, it has been shown that the de-identification of medical data does not guarantee privacy protection of all individuals in the dataset, and it is possible to re-identify individuals by linkage of data from other sources or residual information [2]. This may result in the inability to share data with further research or clinical institutions. In addition, there is often a shortage of available training data for clinicians and researchers alike, significantly impeding scientific progress, particularly in developing countries.

The generation of synthetic data is one such solution to the presented problem. The goal here lies in producing synthetic physiological data representative of real data gathered during the data collection experiment. However, as stated previously, it is important to note that substantial amounts of data are required to successfully train deep learning models for this purpose. Furthermore, protecting the privacy of the underlying real dataset must also be observed [3], [4].

If these problems can be addressed and overcome, the generated data can be published without breaching privacy and used in further training and research. Increasing access to this type of data will encourage scientific studies and facilitate the upskilling of clinicians, which will in turn aid in preventing or limiting chronic illnesses. This can contribute to a shift in the treatment paradigm from reactive to preventative healthcare.

Capitalising on recent advancements in machine learning and, in particular, deep learning could pave the way for the future of sharing data and disseminating research. The work described in this paper is part of a larger-scoped effort to develop artificial intelligence for use in clinical training and upskilling of medical professionals.

Delaney and Brophy [5], [6] demonstrated that realistic synthetic physiological signals could be generated from a dataset of real signals using deep learning methods. However, that work was limited to single time series. We extend this by exploring the possibility of generating multivariate, strongly-coupled physiological time series and investigating using appropriate evaluation metrics to obtain characteristics in the output present in the training dataset. This is an essential step as multivariate medical time series is not simply a collection of independent time series, each of which can be synthesised independently. An extensive deep coupling between the signals exists and is exemplified in multi-lead electrocardiography (an electrical measure of cardiac activity), also known as ECG. A multi-lead ECG involves measuring the heart's electrical activity via several projections over the body's surface via differential bipolar electrode sets. This produces a tightly coupled time series set that can reconstruct an approximation of the dipole dynamics associated with current flow in the beating heart. This paper focuses on the challenge of synthesising such data using our novel objective functions.

In this paper, we demonstrate the contributions of our method in generating realistic, dependent, multivariate physiological signals while maintaining sufficient levels of privacy in the training dataset. Using the Multivariate

GAN (MVGAN) architecture developed in our preliminary work [6], we explore novel loss functions and their effects on generated data quality. We demonstrate our novel GAN, objective function and evaluation metrics capable of improved multivariate time series data generation for the first time. Finally, we benchmark our generative model against other classical generative models.

II. RELATED WORK

A variety of methods have been used in the past to generate synthetic data. In the medical domain, research has mainly focused on the generation of synthetic Electronic Health Records (EHR) [3], [7]. Of particular relevance for our research are those methods which generate synthetic time series data. Previous approaches include the creation of dynamical models to produce synthetic electrocardiogram signals [4]. These models consist of three coupled ordinary differential equations with the user required to specify the characteristics of the heart rate signals to be generated. Many early methods require expert domain knowledge to generate synthetic data. More recent developments in the machine learning space remove this dependency. For example, WaveNet implemented an auto-regressive neural network that successfully generated synthetic music and speech [8]. In other research, Dahmen and Cook (2019) developed SynSys to produce realistic home sensor data using hidden Markov models and regression models [9].

A significant breakthrough in synthetic data generation was facilitated by the introduction of Generative Adversarial Networks (GANs). GANs do not require input from domain experts and they can be designed to preserve privacy of the training datasets. They were first proposed in the seminal paper by Goodfellow in 2014. A multi-layer perceptron was used for both the discriminator and the generator [10]. Radford *et al.* (2015) subsequently developed the deep convolutional generative adversarial network (DCGAN) to generate synthetic images [11]. A recurrent GAN (RGAN) was first proposed in 2016. The generator contained a recurrent feedback loop that used both the input and hidden states at each time step to generate the final output [12]. Recurrent GANs often utilise Long Short-Term Memory neural networks (LSTMs) in their generative models to avoid the vanishing gradient problem associated with more traditional recurrent networks [13]. Since their inception in 2014, GANs have shown great success in generating high-quality synthetic images which are indistinguishable from the actual images [14]–[16].

While the focus has been on developing GANs for improved image generation, there has been a movement towards using GANs for time series and sequence generation [17]. One such implementation involved the generation of polyphonic music as real-valued continuous sequential data using an LSTM in both the generator and discriminator [18]. In contrast, Yu *et al.* (2017) generated synthetic music by representing 88 distinct pitches with discrete tokens [19]. This GAN, known as SeqGAN, contained an

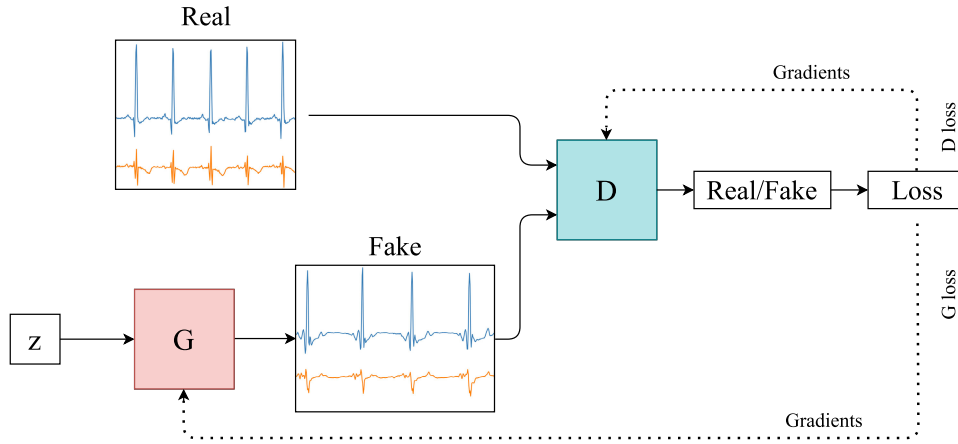


FIGURE 1. Generative adversarial network.

LSTM in the generator with a CNN in the discriminator and outperformed alternative approaches for generating data sequences. GANs were also used to generate single-channel electroencephalogram (EEG) data for motor movement in both the left and right-hand [20]. We are aware of one work that implements both a GAN and a conditional GAN (CGAN) to generate real-valued medical time series data [21]. A CGAN provides additional information to the generator and the discriminator to aid the creation of synthetic data [22]. More recent attempts to generate synthetic ECG used bidirectional LSTMs in the generator and convolutional neural networks in the discriminator [5], [23]. While these works are focused on the generation of medical data, they generate independent, single-channel time series data. Extending on this, we develop a GAN architecture for dependent multivariate medical time series generation. Furthermore, we improve the quality of our generated multichannel ECG through the development of novel loss functions. We compare them with other common loss functions that have been previously explored in the time series GAN literature [17].

III. GENERATIVE ADVERSARIAL NETWORKS

A GAN consists of a generator and a discriminator. The generator G is a neural network that takes random noise $z \in \mathbb{R}^r$ and generates synthetic data. The discriminator D is a neural network that determines if the generated data is real or fake. The generator aims to maximise the failure rate of the discriminator while the discriminator aims to minimise it, see Figure 1. The GAN model converges when the Nash equilibrium is reached. The two networks are locked in a two-player minimax game defined by the value function $V(G, D)$ (1), where $D(x)$ is the probability that x comes from the real data rather than the generated data [10].

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

IV. MULTIVARIATE DYNAMIC TIME WARPING

For a GAN to be considered successful, not only should it converge during training, but it should also learn the distribution of the training data. Dynamic Time Warping (DTW) is used to measure the similarity or distance between two time series sequences and can be implemented as a univariate sequential data classifier. The single-dimensional DTW cumulative distance function defined in (2) is used to find the path that minimises the warping cost. Here $d(q_i, c_j)$ is the squared Euclidean distance between the i^{th} data point of the univariate time series Q and j^{th} data point of the univariate time series C . $D(i, j)$ represents the n -by- n matrix constructed by the squared Euclidean distance between points q_i and c_j where n is the length of the sequence.

$$D(i, j) = d(q_i, c_j) + \min\{D(i-1, j-1), D(i-1, j), D(i, j-1)\} \quad (2)$$

To adapt to the multivariate Dynamic Time Warping (MVDTW) case we redefine $d(q_i, c_j)$ as the cumulative squared Euclidean distances of M data points as in [24]. M is defined as the number of time series that make up the multi-dimensional time series; for this work, the number of individual time series is two ($M = 2$). Q and C are two separate multivariate time series, both with $M = 2$. $q_{i,m}$ is the i^{th} data point in the m^{th} dimension of one multivariate time series Q and $c_{j,m}$ is the j^{th} data point in the m^{th} dimension of the other multivariate time series C , $d(q_i, c_j)$ now becomes:

$$d(q_i, c_j) = \sum_{m=1}^M (q_{i,m} - c_{j,m})^2 \quad (3)$$

Therefore we can now define the cumulative distance for MVDTW as in equation (4). This allows us to find the distance that minimises the warping path and calculate MVDTW. In turn, we can calculate the similarity between

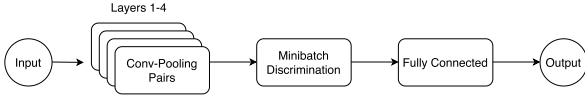


FIGURE 2. Block diagram of discriminator architecture.

our generated data and training data.

$$D(i, j) = \sum_{m=1}^M (q_{i,m} - c_{j,m})^2 + \min \{D(i-1, j-1), D(i-1, j), D(i, j-1)\} \quad (4)$$

V. MODEL DESIGN

This section presents our MVGAN model for generating synthetic, dependent, multivariate physiological time series data. Structurally, our model builds on the architectures of our previous preliminary work. We increase the limited sequence length of 187 in [5], [6] to a more realistic length of 500 sample points. This yields a time series more representative of digitised ECG for the time windows considered (5 seconds at a realistic sampling rate). This length is arbitrary and can be varied through the discriminator to produce data sequences of differing sizes. In terms of generating multichannel data, we increase the number of features available at the input and output of our model. This enables the model to generate realistic, coupled multivariate time series data; this has not been done in previous work. Extending on the earlier models, we also implement 2-dimensional convolution-pooling layers and include a minibatch discrimination layer in the discriminator to prevent mode collapse. The optimiser also has noise introduced to its gradients to create a differentially private GAN model (GAN-DP).

A. GENERATOR

The generator consists of a two-layer stacked LSTM with 50 hidden units in each layer and a fully connected layer at the output. With the extra expected features at the input of the *torch.nn.LSTM* class, this architecture facilitates the generation of multivariate time series data and can scale up to more channels as needed.

B. DISCRIMINATOR

The discriminator is a four-layer 2-dimensional convolutional neural network, a minibatch discrimination layer, a fully connected layer and a sigmoid activation function. Noise was added to the gradient of the optimiser to ensure differential privacy for the GAN-DP model. See Figure 2 for a block diagram of the discriminator and Table 1 for an example of the model parameters.

VI. LOSS FUNCTIONS

Keeping with the same architecture for the MVGAN, we explore novel loss functions by implementing the Loss Sensitive GAN's (LS-GAN) objective function [25] and

TABLE 1. Parameters for the convolution-pooling pairs in the discriminator.

Layer	Input Size	Feature Maps	Filter	Stride	Output Size
C1	2*187	3*2	3	1	3*2*185
P1	3*2*185	3*2	3	1	3*2*183
C2	3*2*183	5*2	3	1	5*2*181
P2	5*2*181	5*2	3	2	5*2*90
C3	5*2*90	8*2	3	2	8*2*44
P3	10*2*44	8*2	3	2	8*2*21
C4	10*2*21	10*2	5	2	10*2*9
P4	10*2*9	10*2	5	2	10*2*3

tailoring it to our multivariate time series generation problem. When the distance between a generated and real multivariate sample becomes small, the GAN will stop increasing the difference $L\theta(zG) - L\theta(x)$ between their losses. The LS-GAN optimizes $L\theta$ and $G\phi$ alternately by seeking an equilibrium (θ_*, ϕ_*) such that θ_* minimizes (5).

$$S(\theta, \phi^*) = \mathbb{E}_{x \sim p_{data}(x)} [L_\theta(x) + \lambda] + \mathbb{E}_{x \sim p_{data}(x), z \sim p_z(z)} [\Delta(x, zG) + L_\theta(x) - L_\theta(zG)] \quad (5)$$

In exploring other loss functions, we investigate the LSGAN with (6) and without (7) an additional penalisation term in the discriminator. This term is the MVDTW, and it penalises the generator if the distance between the multivariate real and generated samples is large. This loss term holds if $1 \leq MVDTW(x, G(z))$. The generator's objective function remains unchanged (8). Here, a is the label for the generated samples, b is the label for the real samples, and c is the hyperparameter that G wants D to recognise the generated samples as real samples.

$$\min_D V_{LSGAN}(D) = 1/2 * \mathbb{E}_{x \sim p_{data}(x)} [(D(x) - b)^2] + 1/2 * \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - a)^2] \quad (6)$$

$$\min_D V_{LSGAN-DTW}(D) = 1/2 * \mathbb{E}_{x \sim p_{data}(x)} [(D(x) - b)^2] + 1/2 * \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - a)^2] + \mathbb{E}_{x \sim p_{data}(x), z \sim p_z(z)} \times [1 - 1/\log(MVDTW(x, G(z)))] \quad (7)$$

$$\min_G V_{LSGAN}(G) = 1/2 * \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - c)^2] \quad (8)$$

The following objective function (9,10) takes the MVDTW of the probability that a sample is either real or fake along with the adversarial ground truth. The adversarial ground truth is an array of either 0's or 1's. In this case, the MVDTW computes the distance between the probabilities and ground truth. In essence, this function computes the squared euclidean distance, and it is retained in this paper as it produces both qualitatively plausible and quantitatively competitive samples.

$$\min_D V_{DTWGAN}(D) = 1/2 * MVDTW_{x \sim p_{data}(x)} [(D(x), valid)] + 1/2 * MVDTW_{z \sim p_z(z)} [(D(G(z)), fake)] \quad (9)$$

$$\min_G V_{DTWGAN}(G) = 1/2 * DTW_{z \sim p_z(z)}[(D(G(z)), valid)] \quad (10)$$

Given these loss functions we now have 5 MVGAN variants as follows:

- 1) Least Squares GAN (LSGAN) (6), (8)
- 2) Least Squares GAN with DTW term (LSGAN-DTW) (7), (8)
- 3) GAN with DTW criterion (DTWGAN) (9), (10)
- 4) Loss Sensitive GAN (LS-GAN) (5)
- 5) Loss Sensitive GAN with DTW distance measure (LS-GAN-DTW) (5)

VII. MATERIALS AND METHODS

A. DATASETS

Multichannel ECG records are signals from two or more leads simultaneously and are frequently used in place of single-channel ECG to give a more complete understanding of the cardiac state. To demonstrate this MVGAN architecture effectively generates multichannel ECG, we have used two datasets in this work. The first openly available dataset is the MIT-BIH Normal Sinus Rhythm (NSR) Database, which includes 18 long-term ECG recordings of subjects found to have had no significant arrhythmia. Recordings were collected at Boston's Beth Israel Hospital and digitised at 128 Hz. Subjects include five men aged 26 to 45, and 13 women, aged 20 to 50.

The second dataset used is the publicly available MIT-BIH Arrhythmia (ARR) dataset [26]. This database contains 48 half-hour long recordings of two-channel ambulatory ECG. Both normal ECG and a range of uncommon but clinically significant ECG irregularities are included in this dataset. The authors of the data collection experiment digitised the recordings at 360 Hz. Each of the records was analysed by two cardiologists to provide reference annotations for every beat. For this dataset, a modified limb lead II (MLII) was used for recording one channel and a unipolar chest lead, also called precordial, or V lead, was used to measure the other channel. V1 was the most common chest lead used, but in some cases, V2, V4, or V5 was used.

In both cases, the datasets are open source and freely available on PhysioNet [27]. Figure 3 shows an example trace of a classic ECG expected from the datasets. The multichannel lead configuration illustrates the dependencies present in the signals that we are seeking to replicate.

B. DATA PREPROCESSING

The datasets were pulled from PhysioNet and loaded using Python's *wfdb* library. Before training our GANs, the datasets required preprocessing in R-peak alignment, segmentation and downsampling. These steps are detailed in the following subsections.

1) R-PEAK ALIGNMENT

Successfully generating dependent multivariate time series requires the training data to retain its inherent dependencies. Fortunately, the ECG channels are already concurrent before

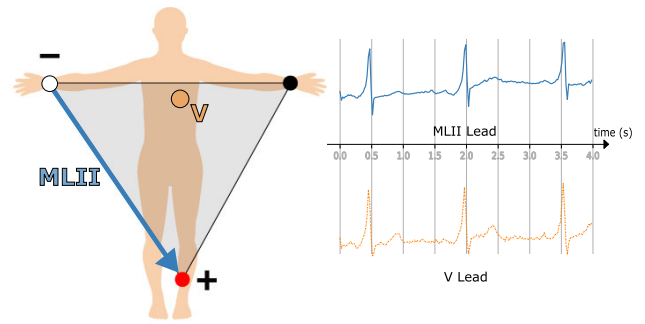


FIGURE 3. Lead configuration used for that data collection experiment with corresponding expected ECG trace.

any preprocessing steps. An R-peak detector provided by *wfdb*'s processing module was used on each of the ECG records. Aligning an R-peak in the centre of every training sample ensured a more effective training set as the QRS complexes occupy similar locations in the sequences. The QRS complex represents ventricular depolarisation and is a combination of the Q, R and S waves in the cardiac cycle.

2) RESAMPLING

The signals were then resampled from their original sampling frequency of 128Hz (NSR dataset) and 360Hz (ARR dataset) to 100Hz using SciPy's *signal.resample*.

3) SEGMENTATION

Following resampling, we normalised and segmented the recordings into smaller samples, each consisting of 5 seconds of data for both leads. Naturally, these samples will not contain the same QRS complexes as the cardiac cycle has natural variability. The length of the data was varied from our previous works [5], [6] to demonstrate the scalability of our GAN architecture. An example of the multichannel input data is shown below in Figure 4 with an artificial offset on the lead two for visualisation purposes.

C. TRAINING

For every loss function explored, the GAN was trained for 50 epochs. For each epoch, the entire training set was divided into batches of 50 multivariate samples. The RMSprop optimiser was used with a learning rate of $\alpha = 0.0002$ as it is computationally efficient and works well for this deep learning model. The GAN variants were trained without minibatch discrimination (MBD), and no mode collapse was observed. We have shown previously that the inclusion of MBD layers can be used with this architecture to prevent mode collapse [6]. In addition, noise was introduced into the gradients of the discriminator optimiser to ensure a differentially private network [28].

D. EVALUATION

1) QUALITY

Maximum Mean Discrepancy (MMD) and multivariate Dynamic Time Warping were used to assess generated data quality. MMD is used here to reinforce the DTW results and

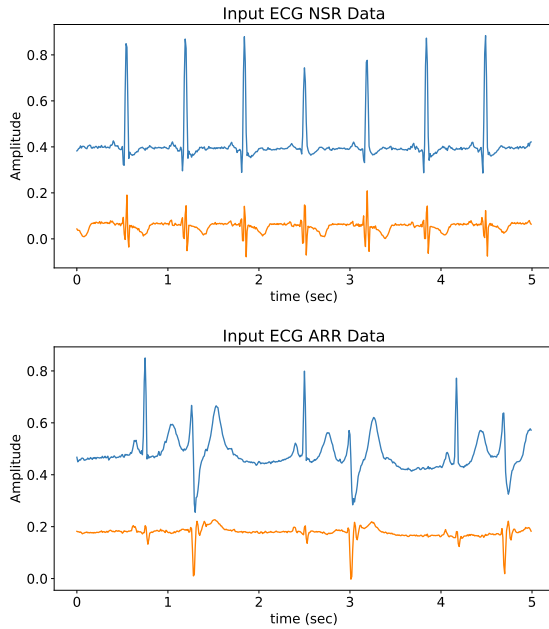


FIGURE 4. Example of preprocessed multichannel ECG from NSR dataset (top) and ARR dataset (bottom) used in training.

to demonstrate that the GAN iteratively learns and generates data from a distribution more representative of the training data distribution.

Multivariate, dependent DTW was calculated to determine similarity measures across the dependent signals in the generated data against the training data. We have shown in the past that the MVDTW method can be used to evaluate generated data from time series GANs [6]. Generated data from the trained generator was compared against the complete training set for evaluation. The evaluation results were averaged over several runs of the model.

2) PRIVACY

Membership inference attacks observe the behaviour of our GAN and attempt to predict examples that were used to train it. A membership inference attack was run to assess presence disclosure. Presence disclosure occurs if it is possible to determine that a particular record was used to train a GAN by observing the generated samples. The sample size r was varied between [1000,10000] training records while the threshold ϵ ranged from [0.05,0.5] of the mean Euclidean distance between all samples. A synthetic dataset of 1000 generated samples was used for this test.

E. BENCHMARKING

To further evaluate and demonstrate the advantages of our GAN, we benchmark our results against current, well-known generative modelling methods. Using the same training dataset, we implemented a multivariate Variational Autoencoder (VAE) and LSTM as a means of generating the type of multichannel data that the proposed GAN is capable of generating. It is important to note that these methods are

usually implemented with single-channel time series data. Here we adopt these methods from a single time series to the multichannel context for the first time to create a benchmark comparison.

To compare how closely the distribution and distance of the generated data match that of the training data, we implemented two time series classifiers alongside our MVDTW and MMD metrics. Support Vector Classification (SVC) and LSTM were the two classifiers of choice. We classify the generated and training data using these models; a classification rate closer to 0.5 demonstrates the classifiers have difficulty distinguishing actual data from the generated data. The poorer the performance of the classifier, the closer the generated data is to the training data. We also compare the data generated from our differentially private GAN to the GAN without DP.

Following this, we run a membership inference attack on the generated data for the GAN and GAN-DP to observe what difference, if any, the differential privacy offers. This series of experiments allows us to understand which model generates realistic multivariate time series signals and which models preserve the underlying privacy of the training data most effectively.

VIII. RESULTS

In this section, we focus on the data generated by the GAN without differential privacy unless explicitly stated otherwise. Qualitative examples of high-quality generated ECG for each GAN can be found from Figure 5 to Figure 9. It becomes apparent that the LS-GAN-DTW generates the best qualitative results for the NSR and ARR datasets. The other variant models appear to have successfully generated realistic, multivariate and dependent ECG data. For visualisation purposes, an offset is again artificially introduced to lead II (orange).

The results shown in Figures 5 through 9 demonstrate that this architecture can successfully generate realistic ECG samples. Lead I is shown in blue and lead II in orange with an artificial offset introduced for visualisation purposes. It appears that for the ARR dataset the GAN models generate noisy ECG, but given the diverse nature of this dataset, the GANs generate good quality data, as is evident in the metrics that follow. However, a qualitative evaluation cannot be considered a complete evaluation of GAN performance due to the lack of a suitable objective function to measure data quality. We address this challenge in the following section.

A. EVALUATION

1) QUALITY

Visually, and therefore from a qualitative perspective, the multi-lead ECG synthesised is of high quality; however, we augment this assessment through the development of suitable objective quantitative metrics. We demonstrate the results for these metrics in this section.

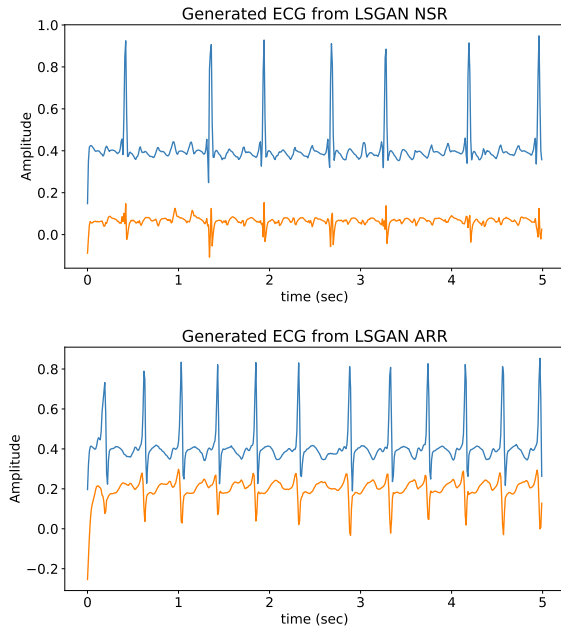


FIGURE 5. Examples of generated multichannel ECG from LSGAN for NSR dataset (top) and ARR dataset (bottom).

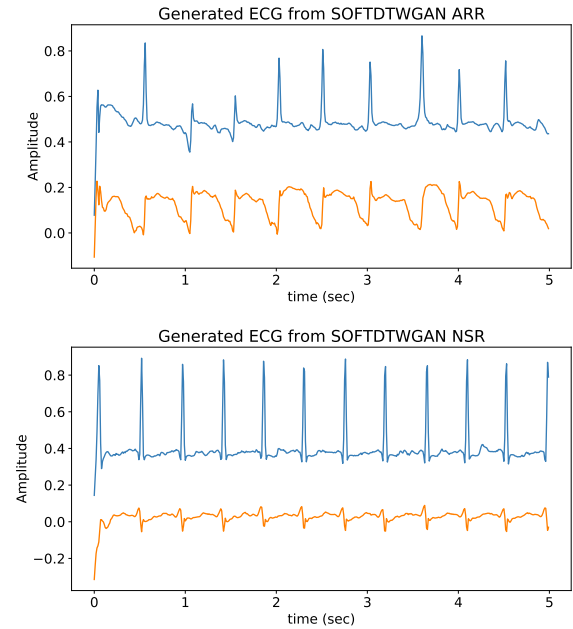


FIGURE 7. Examples of generated multichannel ECG from DTWGAN for NSR dataset (top) and ARR dataset (bottom).

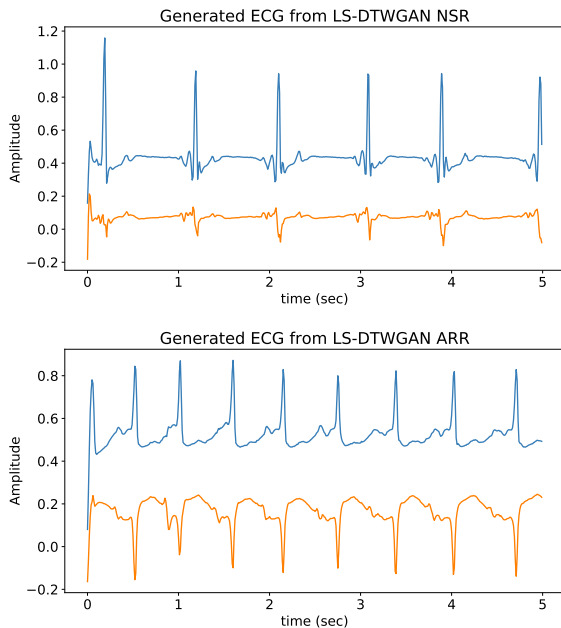


FIGURE 6. Examples of generated multichannel ECG from LSGAN-DTW for NSR dataset (top) and ARR dataset (bottom).

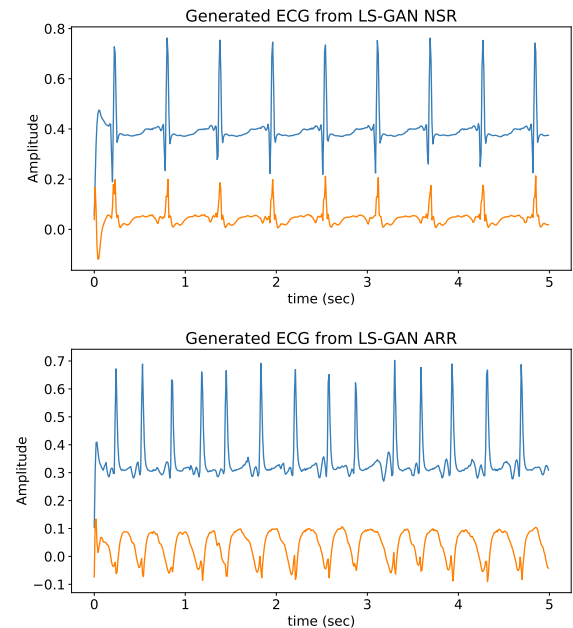


FIGURE 8. Examples of generated multichannel ECG from LS-GAN for NSR dataset (top) and ARR dataset (bottom).

Maximum Mean Discrepancy results in Tables 2 and 3 demonstrate that as the GAN iterates through the training process, it is generating data from a distribution that is more representative of the training data distribution.

Results for DTW extended to multivariate time series can be seen in Tables 2 and 3. The distance measures between the dependent generated signals and the dependent training signals are reducing throughout the training process, indicating that the proposed GAN has successfully captured

the multivariate training data distribution. Although the LSGAN appears to produce the best quantitative results for the NSR dataset according to the metrics used in this paper, the DTWGAN produces an improved MMD for the ARR dataset. The best performing GAN is shown in Figure 10. Over both datasets, normalising DTW and MMD results, the best performing model is the LSGAN-DTW. The LSGAN-DTW shows a 4.9% improvement over the LSGAN and 4.5% improvement over the DTWGAN. As a result of

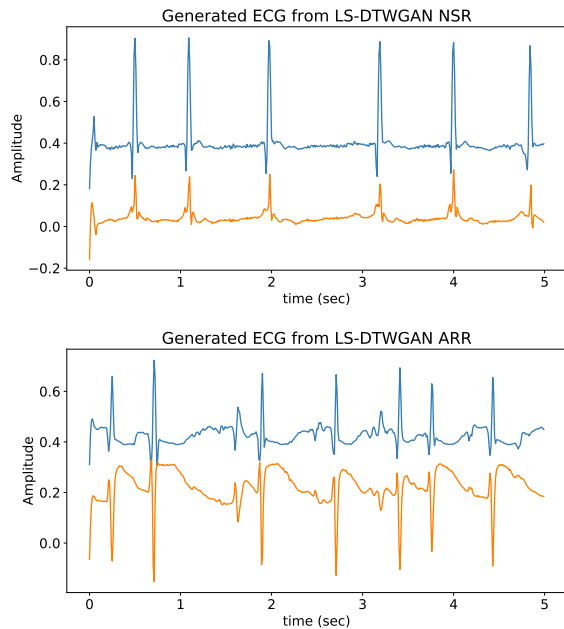


FIGURE 9. Examples of generated multichannel ECG from LS-GAN-DTW for NSR dataset (top) and ARR dataset (bottom).

TABLE 2. Metrics NSR.

Modelling Method	Evaluation	
	MVDTW	MMD
LSGAN	3.0598	0.0057
LSGAN-DTW	3.3069	0.0106
DTWGAN	3.5463	0.0203
LS-GAN	6.7879	0.0194
LS-GAN-DTW	3.3229	0.0086

TABLE 3. Metrics ARR.

Modelling Method	Evaluation	
	MVDTW	MMD
LSGAN	4.0101	0.0568
LSGAN-DTW	4.3463	0.0195
DTWGAN	4.5918	0.0157
LS-GAN	7.4448	0.1917
LS-GAN-DTW	4.2674	0.0747

the LSGAN-DTW being the overall best performing GAN, the results that follow will be reported for this variant unless explicitly stated otherwise.

2) PRIVACY

In terms of privacy, Figure 11 and Figure 12 shows the presence disclosure (averaged over both datasets) for a membership inference attack on the LSGAN-DTW and LSGAN-DTW-DP, respectively. The number of training records identified is very low (Recall), with approximately 0% correctly identified for $\epsilon \leq 0.4 * \text{mean distance}$. However, as the threshold ϵ increases above this boundary, the number of records correctly identified as training records increases independently of the sample size r for the GAN without differential privacy. The GAN-DP preserves the training data

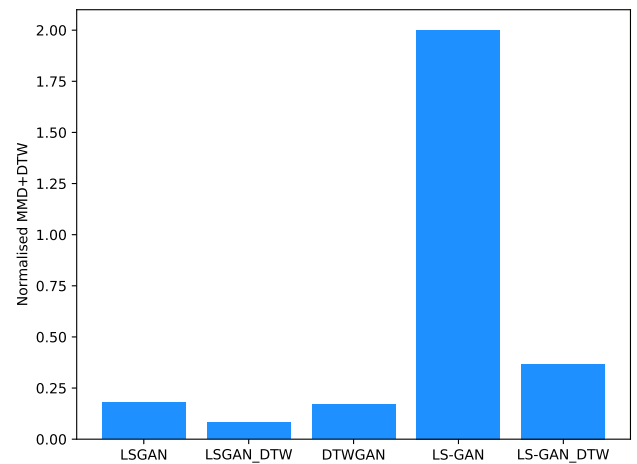


FIGURE 10. Combined, normalised MMD and MVDTW results demonstrate the best performing GAN across both datasets to be the LSGAN-DTW.

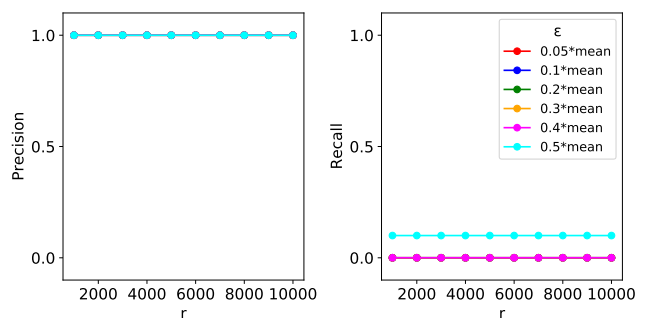


FIGURE 11. Presence disclosure of GAN following a membership inference attack with increasing ϵ .

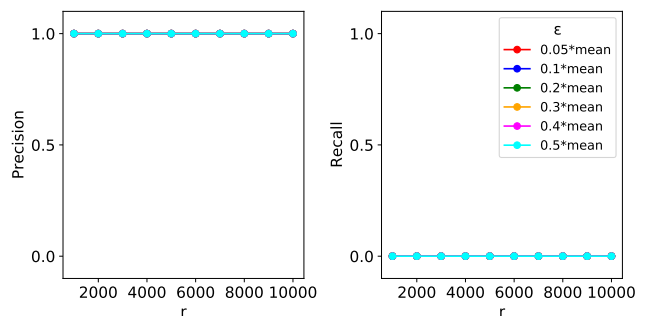


FIGURE 12. Presence disclosure of GAN-DP following a membership inference attack with increasing ϵ .

privacy with no training records identified for $\epsilon \leq 0.5 * \text{mean distance}$. Precision is approximately 100% for all ϵ and r , which means once an attacker deems that a sample originates from the training set, it is almost always correctly attributed to the training set. Overall, this result tells us that for our generated data, an attacker will have a challenging time correctly identifying if a sample has originated from the training set. Therefore, this GAN architecture and loss function can generate data similar in distribution to the training set while maintaining sufficient privacy of the data.

TABLE 4. Classifier accuracy for the generative models averaged over both datasets.

Modelling Method	Classifier Accuracy		
	SVC	LSTM	Average
VAE	0.9	0.68	0.79
LSTM	0.95	0.62	0.78
LSGAN-DTW	0.4	0.62	0.51
LSGAN-DTW-DP	0.56	0.86	0.71

TABLE 5. Evaluation metrics for other generative models averaged over both datasets.

Modelling Method	Evaluation Performance	
	MVDTW	MMD
VAE	47.14	0.96
LSTM	27.87	1.09
LSGAN-DTW	3.8266	0.01505
LSGAN-DTW-DP	13.36	0.29

B. BENCHMARKING RESULTS

These experiments offer a benchmark to compare our GAN to other generative models. Table 4 shows the classifier accuracy averaged over both datasets for each of the generative models introduced in Section VII-E. SVC and LSTM were the two time series classifiers used in these benchmarking tests. A lower classifier accuracy indicates the classifier had difficulty distinguishing the classes, which in this case were the real and generated data. The modelling method that generated the most similar data to the real data was the LSGAN-DTW.

To complement the results shown in Table 4, evaluation metrics were computed for each of the modelling methods. MVDTW and MMD were calculated as in Section VIII-A1 and the results for which are shown in Table 5 below, averaged over both datasets. Smaller distances for MVDTW reflect time series that are more similar to each other and for MMD indicate that the real and generated data distributions are closer. As can be seen, LSGAN-DTW has lower distances for MVDTW and MMD, followed by LSGAN-DTWGAN-DP. This quantitatively demonstrates that the data generated using the GANs are more representative of the real data compared with that of the other generative and time series modelling methods.

IX. DISCUSSION & CONCLUSION

The multivariate GAN proposed in this work has demonstrated a capability for generating high-quality, dependent, multichannel ECG signals. Our introduction of the DTW penalisation term in the GAN objective function leads to a more robust design that avoids mode collapse without the need for MBD layers and results in the generation of diverse multichannel time series. We also introduced a new quantitative method for the assessment of output quality for multichannel time series GANs, namely MVDTW. These quantitative methods can complement qualitative evaluation and, in the context of this paper, confirm the strong performance of the proposed GAN. Ideally, rather than solely relying on classical and novel metrics, we could enlist the

help of a trained physician to classify samples of generated data to determine how accurate the signal traces are, as we have done in our previous work [29]. This forms an avenue for our future work. Finally, given the nature of the data, it would be interesting to implement conditional models to generate normal and pathological data, which would enable future researchers to generate ECG based on their needs.

To address the growing privacy concerns with sensitive personal data such as physiological or medical data, we demonstrated the ability of the LSGAN-DTW, and in particular, the LSGAN-DTWGAN-DP, to sufficiently conserve the confidentiality of the underlying training data. Implementation of a membership inference attack demonstrated promising results for data privacy with these GANs; protecting and isolating the training set from the generated data ensures that a certain level of privacy is maintained. With the addition of a differentially private GAN architecture, we can generate data and ensure that the privacy of the training data is not violated.

We also presented benchmark experimental results for showcasing the advantages that the LSGAN-DTW holds over other generative time series modelling methods. Most of these well-known methods are tailored explicitly to univariate signals, whereas our methods can be scaled up to multivariate use cases, which include strong coupling between time series. Not only is the proposed method capable of generating multivariate medical time series data, but it also generates data from a closer distribution and distance to that of the training data in comparison to the other generative modelling methods utilised in this paper.

Multivariate time series data presents an opportunity for the application of GANs in tackling the data shortage and sharing problem in medical research. In terms of our motivating challenge, the successful generation of diverse samples of multichannel and dependent physiological data means we have the potential to use this technology for clinical training and research applications. With that goal in mind, we have shown, for the first time, a GAN design capable of generating high-quality dependent multichannel physiological time series with quality similar to that present in clinically relevant data repositories.

ACKNOWLEDGMENT

The authors would like to thank the support of NVIDIA Corporation with the donation of the Titan Xp used for this research.

REFERENCES

- [1] K. El Emam, S. Rodgers, and B. Malin, "Anonymising and sharing individual patient data," *BMJ*, vol. 350, p. h1139, Mar. 2015.
- [2] K. El Emam, E. Jonker, L. Arbuckle, and B. Malin, "A systematic review of re-identification attacks on health data," *PLoS ONE*, vol. 6, no. 12, Dec. 2011, Art. no. e28071.
- [3] S. McLachlan, K. Dube, and T. Gallagher, "Using the CareMap with health incidents statistics for generating the realistic synthetic electronic healthcare record," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Oct. 2016, pp. 439–448.

- [4] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289–294, Mar. 2003.
- [5] A. M. Delaney, E. Brophy, and T. E. Ward, "Synthesis of realistic ECG using generative adversarial networks," 2019, *arXiv:1909.09150*.
- [6] E. Brophy, "Synthesis of dependent multichannel ECG using generative adversarial networks," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 3229–3232, doi: [10.1145/3340531.3418509](https://doi.org/10.1145/3340531.3418509).
- [7] Y. Park and J. Ghosh, "Perturbed Gibbs samplers for synthetic data release," 2013, *arXiv:1312.5370*.
- [8] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "WaveNet: A generative model for raw audio," 2016, *arXiv:1609.03499*.
- [9] J. Dahmen and D. Cook, "SynSys: A synthetic data generation system for healthcare applications," *Sensors*, vol. 19, no. 5, pp. 1–11, 2019.
- [10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2014, pp. 2672–2680. [Online]. Available: <http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>
- [11] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.
- [12] D. J. Im, C. D. Kim, H. Jiang, and R. Memisevic, "Generating images with recurrent adversarial networks," 2016, *arXiv:1602.05110*.
- [13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [14] J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," 2017, *arXiv:1709.01872*.
- [15] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.
- [16] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *Proc. 33rd Int. Conf. Mach. Learn., (ICML)*, vol. 3, 2016, pp. 1681–1690.
- [17] E. Brophy, Z. Wang, Q. She, and T. Ward, "Generative adversarial networks in time series: A survey and taxonomy," 2021, *arXiv:2107.11098*.
- [18] O. Mogren, "C-RNN-GAN: Continuous recurrent neural networks with adversarial training," 2016, *arXiv:1611.09904*.
- [19] L. Yu, W. Zhang, J. Wang, and Y. Yu, "SeqGAN: Sequence generative adversarial nets with policy gradient," in *Proc. 31st AAAI Conf. Artif. Intell. (AAAI)*, 2017, pp. 2852–2858.
- [20] K. G. Hartmann, R. T. Schirrmeyer, and T. Ball, "EEG-GAN: Generative adversarial networks for electroencephalographic (EEG) brain signals," 2018, *arXiv:1806.01875*.
- [21] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional GANs," 2017, *arXiv:1706.02633*.
- [22] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
- [23] F. Zhu, F. Ye, Y. Fu, Q. Liu, and B. Shen, "Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network," *Sci. Rep.*, vol. 9, no. 1, pp. 1–11, Dec. 2019.
- [24] M. Shokoohi-Yekta, B. Hu, H. Jin, J. Wang, and E. Keogh, "Generalizing dtw to the multi-dimensional case requires an adaptive approach," *Data Mining Knowl. Discovery*, vol. 31, no. 1, pp. 1–31, 2017.
- [25] G.-J. Qi, "Loss-sensitive generative adversarial networks on Lipschitz densities," 2017, *arXiv:1701.06264*.
- [26] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May/Jun. 2001.
- [27] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, Jun. 2000, doi: [10.1161/01.CIR.101.23.e215](https://doi.org/10.1161/01.CIR.101.23.e215). [Online]. Available: <http://circ.ahajournals.org/content/101/23/e215.full>
- [28] L. Xie, K. Lin, S. Wang, F. Wang, and J. Zhou, "Differentially private generative adversarial network," 2018, *arXiv:1802.06739*.
- [29] J. Weldon, T. Ward, and E. Brophy, "Generation of synthetic electronic health records using a federated GAN," 2021, *arXiv:2109.02543*.



data and, in turn, advancing clinical research in the relevant fields, such as in the neonatal ICU. He won the Apple Best Paper Award from the WristSense Workshop at PerCom 2020.



signatures of patient health from data acquired via wearable sensors and the incorporation of smart analytics into unobtrusive systems. He has won several innovation prizes, among which was the prestigious Mobile Brain Body Monitoring Prize, in 2017. In 2019, he was awarded the Martin Black Prize for the Best Paper in physiological measurements.



across Europe. She was also the Coordinator of the EU funded NEMO study, Europe's first multi-centred dose-finding, and safety study of bumetanide for the treatment of seizures in newborn infants (<http://www.nemo-europe.com>).



new ways of sensing and interpreting our physiology and behavior. He has authored more than 240 peer-reviewed scientific publications. His current research interests include application of machine learning to improve human health, performance, and decision making.

...