Journal of Theoretical and Applied Information Technology
31st May 2021. Vol.99. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CTGAN VS TGAN? WHICH ONE IS MORE SUITABLE FOR GENERATING SYNTHETIC EEG DATA

[1]MIN JONG CHEON, [2]DONG HEE LEE, [3]JI WOONG PARK, [4]HYE JIN CHOI, [5] JUN SEUCK LEE, [*]OOK LEE

[1]Department of Information Systems, Hanyang University, 222 Wangshimni-ro, Seongdong-gu, 04673 Seoul, South Korea
[2]Department of Information Systems, Hanyang University, 222 Wangshimni-ro, Seongdong-gu, 04673 Seoul, South Korea
[3]Department of Information Systems, Hanyang University, 222 Wangshimni-ro, Seongdong-gu, 04673 Seoul, South Korea
[4]Department of Information Systems, Hanyang University, 222 Wangshimni-ro, Seongdong-gu, 04673 Seoul, South Korea
[5]Department of Information Systems, Hanyang University, 222 Wangshimni-ro, Seongdong-gu, 04673 Seoul, South Korea
[*]Department of Information Systems, Hanyang University, 222 Wangshimni-ro, Seongdong-gu, 04673 Seoul, South Korea

E-mail: [1]jmj2316@hanyang.ac.kr, [2]ryu03153@hanyang.ac.kr, [3]jiwongp94@hanyang.ac.kr, [4]hyyejinn@hanyang.ac.kr, [5]js940912@hanyang.ac.kr, [*]ooklee@hanyang.ac.kr

ABSTRACT

BCI has been an alternative method of communication between a user and a system, and EEG is a representative non-invasive neuroimaging technique in BCI research. However, gathering a large dataset of EEG is difficult due to insufficient conditions. Therefore, a data augmentation is required for the data and a generative adversarial network is a representative model for the augmentation. As the EEG data is a CSV format, we decided to utilize CTGAN and TGAN for creating synthetic data. Our research was conducted through 3 steps. First of all, we compared two datasets from each model through data visualization. Secondly, we conducted a statical method for calculating similarity score. Lastly, we used both data as input data of the machine learning algorithms. Through the first step and second step, we found that the data from CTGAN has higher similarity than TGAN. However, in the last step, the result showed that the result such as accuracy, precision, recall, f1 score showed no significant difference between the two datasets. Furthermore, compared to the original dataset, none of the synthetic datasets showed higher scores. Therefore, we concluded that further research is needed to find out a better method for data augmentation so that the synthetic data could be utilized for the input data of machine learning or deep learning algorithms.

Keywords: *Deep Learning, GAN, EEG, BCI, Data Augmentation, Artificial Intelligence*

1. INTRODUCTION

1.1 Background

Brain Computer Interface (BCI) is a system that connects human brain activity to external technology. BCI demonstrated possibilities in predicting the intentions of people by analyzing brain signals. Many researchers have studied EEG-based emotion recognition methods and have made significant progress[1].

Electroencephalography (EEG) is a test that detects and records electrical activity of the brain. EEG is one of the main non-invasive neuroimaging techniques in neuroscience research[2]. However, EEG-based research has faced difficulties because it requires highly cost devices to produce data with high sampling rates and sensitivities. Because Low-sampling-sensitivity signals are not good enough to be utilized, comprehensive efforts have been made to reconstruct high-sampling-sensitivity EEG signals from existing signals. The upsampling operation is a typical reconstructing method to generate upsampled signals and signals with distinct sensitivity[3]. There are three categories of reconstruction methods: reconstruction by interpolation[4], mathematical modeling[5], deep neural networks[6]. Even though both reconstruction by interpolation and mathematical modeling are assumed to be effective in reconstructing signals, neither could represent brain activity[7].

In contrast to reconstruction by interpolation and mathematical models, reconstruction by deep neural networks (DNNs) demonstrated promising results in representing brain signals by focusing on image signal reconstruction. A class of machine learning frameworks for generating artificial data, Generative Adversarial Networks (GANs) showed outstanding outcomes in generating new brain signals. If the generation of artificial EEG signals can be used as naturalistic data, there would be great advances in the BCI field[7].

1.2 Related Works

Luo, Y., Lu, B.-L. proposed CWGAN framework for EEG data augmentation. Also, qualities of the generated data are evaluated by the three indicators such as discriminator loss, maximum mean discrepancy, two-dimensional mapping[8]. They used two public EEG datasets, namely SEED and DEAP. The emotion recognition models achieved 2.97%, 9.15% and 20.13% improvements on SEED dataset and DEAP dataset for arousal and valence classifications, respectively. K. G. Harmann. Et al. tried an improvement to the Wasserstein GAN training showing increased training stability. Also, they compared different evaluation metrics such as Inception score, Frechet Inception Distance, Euclidean Distance, and Sliced Wasserstein distance. The EEG datasets used for training stemed from a simple motor task in which the subjects were instructed to either rest or move the left hand. The result showed the models trained with their methods performed best for every evaluation metrics[9]. Luo, T.-J. et al. suggested a contemporary deep neural network that uses a GAN/WGAN framework with a TSFMSE-based loss function for LSS-EEG signal reconstruction. Three EEG signal datasets with different sampling rates and sensitivities were used. Results proposeded that the GAN/WGAN frameworks give a significant improvement on the classification performance of EEG signals reconstruction with the same sensitivity, but the classification performance improvements of EEG signal reconstructions with different sensitivity were not significant[7]. Brenninkmeijer,B. proposed adding skip connections to TGAN to increase gradient flow and information retention, and adding WGAN-GP architecture to TGAN. Also, they proposed a metric to evaluate similarity score. Three datasets were used, which were the Census dataset, the Berka Czech Financial dataset, and Creditcard Farud dataset. The census dataset was mostly consisted of categorical values, while the creditcard dataset was consisted of continuous values. The Berka dataset was a mix of both. As a result, TGAN-skip and TGAN-WGAN-GP outperformed TGAN in the similarity scores in the census and credit card dataset.

1.3 Objectives

Due to inadequate conditions, studies had difficulties in gathering sufficient brain signals from EEG tests. Therefore, finding the productive solution to solve those downsides is important. In order to efficiently accumulate enough data of brain signals, data augmentation, which is a method to increase the amount of data by producing artificial training data from existing training data can be used as an effective tool. Furthermore, from the research about related works, we found that using deep learning model for data augmentation is efficient. Generative adversarial network (GAN) is a representative data augmentation tool based on deep learning. Tabular generative adversarial network (TGAN) and conditional tabular generative adversarial network (CTGAN) were used and each method produces 12811 samples of the same size as the original data. As the first evaluation method for data similarity, this paper attempts to visualize data using TableEvaluator library. As a result, we identify various visualizations such as the column-specific cumsum, distribution of real data and synthetic data and column-specific differences between real and synthetic data. Secondly, the similarity score is calculated using the TableEvaluator library for data similarity evaluation. The similarity score is the average of basic statistics, correction column correction, mirror column correction, 1-MAPE estimator, and 1-MAPE PCA, indicating that the higher the value, the more similar the data is to the original data. The similarity score from this experiment indicates that CTGAN produces more similar data than TGAN. Lastly, this paper validates the similarity between real data and synthetic data using machine learning. Of the 12811 samples produced by TGAN and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CTGAN, 70% are used as train dataset, and 30% of the original data is used as test dataset. This paper tries binary classification with a user-defined label as the target, which indicates whether a participant is confused or not. Random Forest, XGBoost, LightGBM, and Catboost algorithms are used for machine learning algorithms. As a result, the efficacy of each algorithm is evaluated through accuracy, precision, recall, f1, AUC, and confusion matrix.

2. MATERIALS AND METHODS

2.1 Data Description

We used 'Confused student EEG brainwave data' dataset, an EEG data from a Kaggle challenge. The dataset can be found online at https://www.kaggle.com/wanghaohan/confused-eeg[11]. 10 college students were assigned to watch 20 videos, 10 of which were pre-labeled as easy and the others as difficult. Easy videos were assumed not to be confusing for college students, such as videos of the introduction of basic algebra or geometry. Difficult videos were expected to confuse a typical college student if a student was not familiar with the video topics like Quantum Mechanics, and Stem Cell Research. A single-channel wireless MindSet that measured activity over the frontal lobe was used for measuring EEG data. Through this equipment, 11 types of signals were measured, which were Attention, Meditation, Raw, Delta, Theta, Alpha1, Alpha2, Beta1, Beta2, Gamma1 and Gamma2. The students watched videos wearing this Mindset. After each session, the students self-rated their confusion level on a scale of 1-7, where one corresponded to the least confusing and seven corresponded to the most confusing. These labels were quantized into two classes, representing whether the students were confused or not in addition to predefined labels of confusion. The normalized two-class label serves as the target label of this research. Figure 1 shows some samples of the dataset.

2.2 TGAN

TGAN, which is based on the GAN algorithm, was introduced by Lei Xu and Kalyan Veeramachaneni for a tabular data augmentation. Since a tabular data contains numerical variables and categorical variables, both numerical and discrete variables need to take separate steps.
For the numerical variables :

1. Each numerical variable in C_i is trained by a GMM((Gaussian Mixture Model) with m components. GMM models a distribution with a weighted sum of m Gaussian distributions. The means and standard deviation of the m Gaussian distributions are $\eta_i^{(1)}, \dots \eta_i^{(m)}$ and $\sigma_i^{(1)}, \dots, \sigma_i^{(m)}$.

2. For the probability of $c_{i,j}$, which is from each of the m Gaussian distributions as a vector $\mu_{i,j}^{(1)}, \dots \mu_{i,j}^{(m)}$. $\mu_{i,j}$ represents a

	Attention	Mediation	Raw	Delta	Theta	Alpha1	Alpha2	Beta1	Beta2	Gamma1	Gamma2	user-definedlabeln
0	56.0	43.0	278.0	302000.0	90600.0	33700.0	24000.0	27900.0	45100.0	33200.0	8290.0	0.0
1	40.0	35.0	-50.0	73800.0	28100.0	1440.0	2240.0	2750.0	3690.0	5290.0	2740.0	0.0
2	47.0	48.0	101.0	758000.0	384000.0	202000.0	62100.0	36300.0	131000.0	57200.0	25400.0	0.0
3	47.0	57.0	-5.0	2010000.0	129000.0	61200.0	17100.0	11500.0	62500.0	50000.0	33900.0	0.0
4	44.0	53.0	-8.0	1010000.0	354000.0	37100.0	88900.0	45300.0	99600.0	44800.0	29700.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...
12806	64.0	38.0	-39.0	128000.0	9950.0	709.0	21700.0	3870.0	39700.0	2600.0	960.0	0.0
12807	61.0	35.0	-275.0	323000.0	797000.0	153000.0	146000.0	39800.0	571000.0	36600.0	10000.0	0.0
12808	60.0	29.0	-426.0	681000.0	154000.0	40100.0	39100.0	11000.0	27000.0	20400.0	2020.0	0.0
12809	60.0	29.0	-84.0	366000.0	27300.0	11400.0	9930.0	1940.0	3280.0	12300.0	1760.0	0.0
12810	64.0	29.0	-49.0	1160000.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]000.0	22100.0	4480.0	0.0

12811 rows × 12 columns

Figure 1. Samples of the raw data

normalized probability distribution over m Gaussian distributions.

3. $c_{i,j}$ is then normalized as $v_{i,j} = (c_{i,j} - \eta_i^{(k)})/2\sigma_i^{(k)}$, where $k = argmax_k \mu_{i,j}^{(k)}$. $v_{i,j}$ is then clipped to [-0.99, 0.99]

For the categorical variables :

1. A $d_{i,j}$, which is a discrete variable D_i is first represented as a $|D_i|$ -dimensional one-hot vector $d_{i,j}$.

2. Noise is added to each dimension as $d_{i,j}^{(k)} \leftarrow d_{i,j}^{(j)} + Uniform(0,\gamma)$, and $\gamma = 0.2$.

3. The representation is then normalized as $d_{i,j} \leftarrow d_{i,j} / \sum_{k=1}^{|D_i|} d_{i,j}^{(k)}$.

GAN is composed of generator and discriminator. The generator G generates the synthetic data and tries to deceive the discriminator, while the discriminator D tries to discriminate whether the data is real data or synthetic data. In the TGAN, a long short term memory(LSTM) algorithm, which is a kind of recurrent neural network(RNN) is utilized for the generator G. A feed-forward neural network is used for the discriminator, which applies batch normalization, LeakyReLU and mini-batch discrimination[12].

2.3 CTGAN

As a existing method of Generator does not consider imbalance of categorical variables, an approach of conditional generator is suggested. Unlike TGAN, a variational Gaussian Mixture Model(VGM) is utilized instead of GMM for the numerical variables, and Wassersten GAN loss function is used for gradient penalty. For the categorical variables, a "training-by-sampling", conditional vector, and generator loss is implemented for solving imbalance problems. In the "training-by-sampling" method, a critic estimates the output of the conditional generator,

	Attention	Mediation	Raw	Delta	Theta	Alpha1	Alpha2	Beta1	Beta2	Gamma1	Gamma2	user-definedlabeln
0	61	64	53	3740	309000	25700	939	8760	9770	56100	4870	1
1	34	51	52	203000	27500	4190	27400	6490	8780	12800	16600	1
2	48	48	-29	298000	7600	4450	13100	8340	3460	30600	5970	0
3	0	0	2050	891000	42000	11200	29200	44000	4900	109000	601	0
4	48	60	-1410	1370000	9320	2410	204000	8340	24500	25900	1170	0
...	...	...	...	...	...	...	...	...	...	...	...	...
12795	56	57	-184	1850000	265000	12300	5640	6390	17200	63300	8170	1
12796	47	56	-75	1750000	82100	87800	41500	15200	6100	3930	3950	0
12797	50	54	37	21300	46100	22400	575000	22000	59900	13900	77000	1
12798	66	47	-14	914000	6270	341000	2560	9070	52800	13900	9820	1
12799	56	0	2050	3470	107000	45100	5640	5770	17200	63300	831	1

12800 rows × 12 columns

Figure 2. Samples of synthetic data from TGAN

	Attention	Mediation	Raw	Delta	Theta	Alpha1	Alpha2	Beta1	Beta2	Gamma1	Gamma2	user-definedlabeln
0	44.588595	31.471641	39.380119	1.691889e+06	6.311720e+05	75557.988505	59934.754746	50385.783144	70341.935794	40856.148669	28269.406062	0.0
1	34.327515	54.192607	22.026602	2.143840e+06	2.097090e+05	49905.744623	4311.901404	96541.974070	26128.858575	48150.010501	19196.369042	1.0
2	41.224834	43.448344	124.112332	1.298842e+05	3.159513e+04	7162.995310	10626.680171	5361.350502	8172.823611	8452.887636	2278.023623	1.0
3	67.378761	58.133450	-5.994545	-2.123398e+04	-1.117595e+03	2632.838769	16884.290312	4067.254891	3870.894035	9629.918024	9352.663532	1.0
4	92.294966	58.210178	45.623242	5.844718e+03	1.278883e+04	39035.180080	15795.325277	28303.250652	15332.644717	10027.099890	3277.436455	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...
12806	48.170934	51.491822	-87.831508	1.861306e+06	6.899968e+05	173343.303279	80344.518425	125826.267088	110593.445482	143238.442610	48436.225676	0.0
12807	90.105514	68.003344	1.507730	-3.083280e+04	2.465952e+04	14820.560950	10952.529066	17263.488834	11661.265924	6321.449177	7650.573913	1.0
12808	22.628139	52.692271	4.998389	1.212968e+06	2.021458e+06	58537.791905	123235.083852	96526.378053	91194.469658	177325.536802	25966.303667	1.0
12809	-2.772931	0.886210	2092.521888	1.367334e+06	7.031539e+04	7516.593041	408768.148662	125017.450416	359349.739212	221342.674079	138088.579020	1.0
12810	47.378771	30.711490	69.948278	6.774485e+05	4.280611e+04	19058.064500	5092.504409	2256.832191	6369.330469	6721.251273	2306.347478	1.0

12811 rows × 12 columns

Figure 3. Samples of synthetic data from CTGAN

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

which evaluates the distance between the learned conditional distribution and the conditional distribution on real data. The conditional vector is introduced to indicate the condition $(D_{i^*} = k^*)$, while $k = 1, \ldots, |D_i|$, and $d_{i,j}$, which is a discrete variable D_i is first represented as a $|D_i|$-dimensional one-hot vector $d_{i,j}$. During training, the conditional generator is allowed to generate any set of one-hot discrete vectors. Furthermore, for penalizing a loss of the generators by adding the cross-entropy between $\hat{d}_{i^*}$ and m_{i^*}, the procedure to produce $\hat{d}_{i^*} = m_{i^*}$ is suggested, which allows the generator to make an exact copy of the given m_{i^*} into $\hat{d}_{i^*}$. The structure of the critic can be described as[13] :

$$h_0 = r_1 \oplus \ldots \oplus r_{pac} \oplus cond_1 \oplus \ldots \oplus cond_{pac}$$
$$h_1 = drop(leaky_{0.2}(FC_{10|r|+10|cond| \to 256}(h_0)))$$
$$h_2 = drop(leaky_{0.2}(FC_{256 \to 256}(h_1)))$$
$$C(\cdot) = FC_{256 \to 1}(h_2)$$

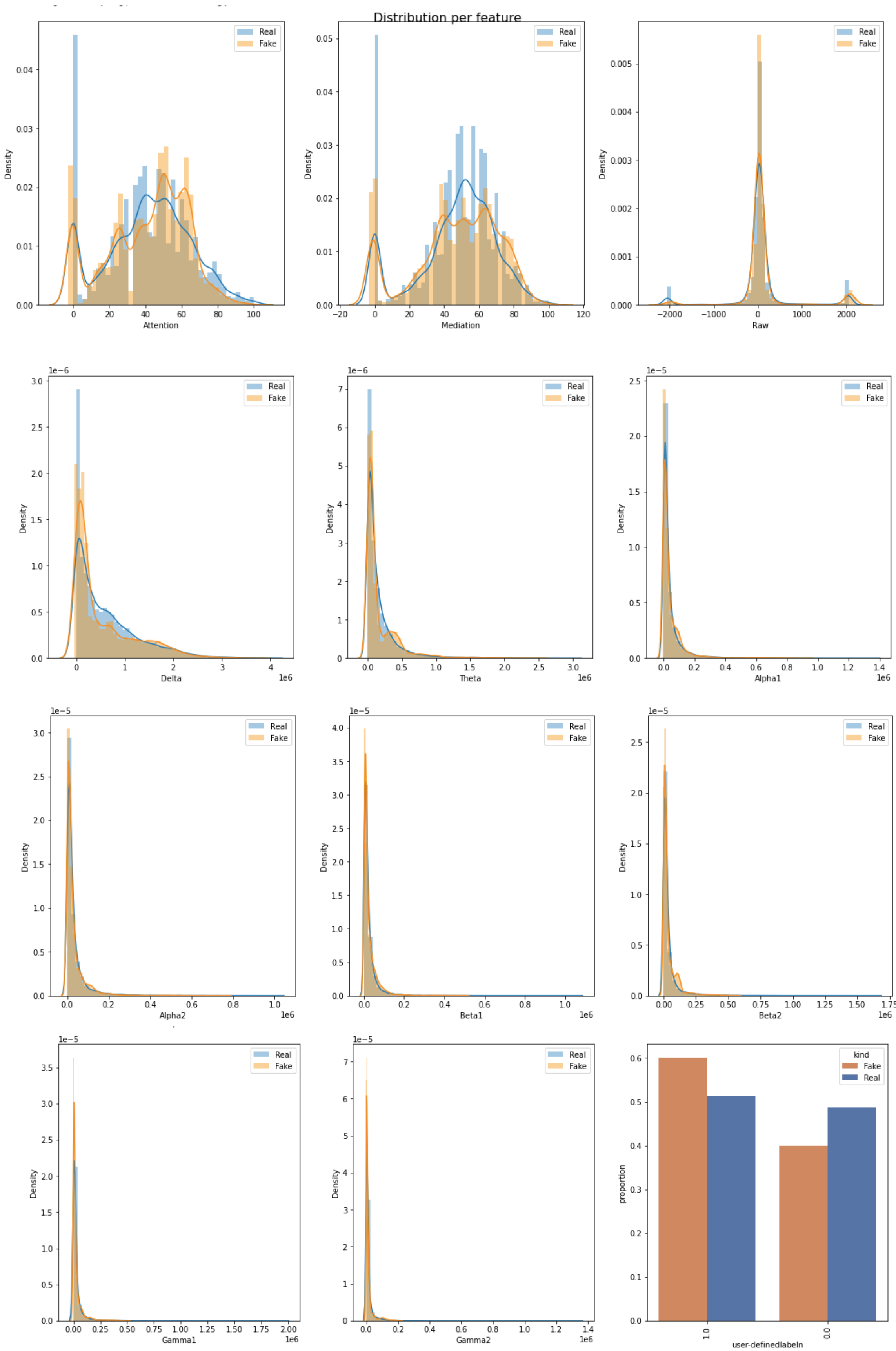
Figure 2 and Figure 3 show some samples of the generated data from TGAN and CTGAN.

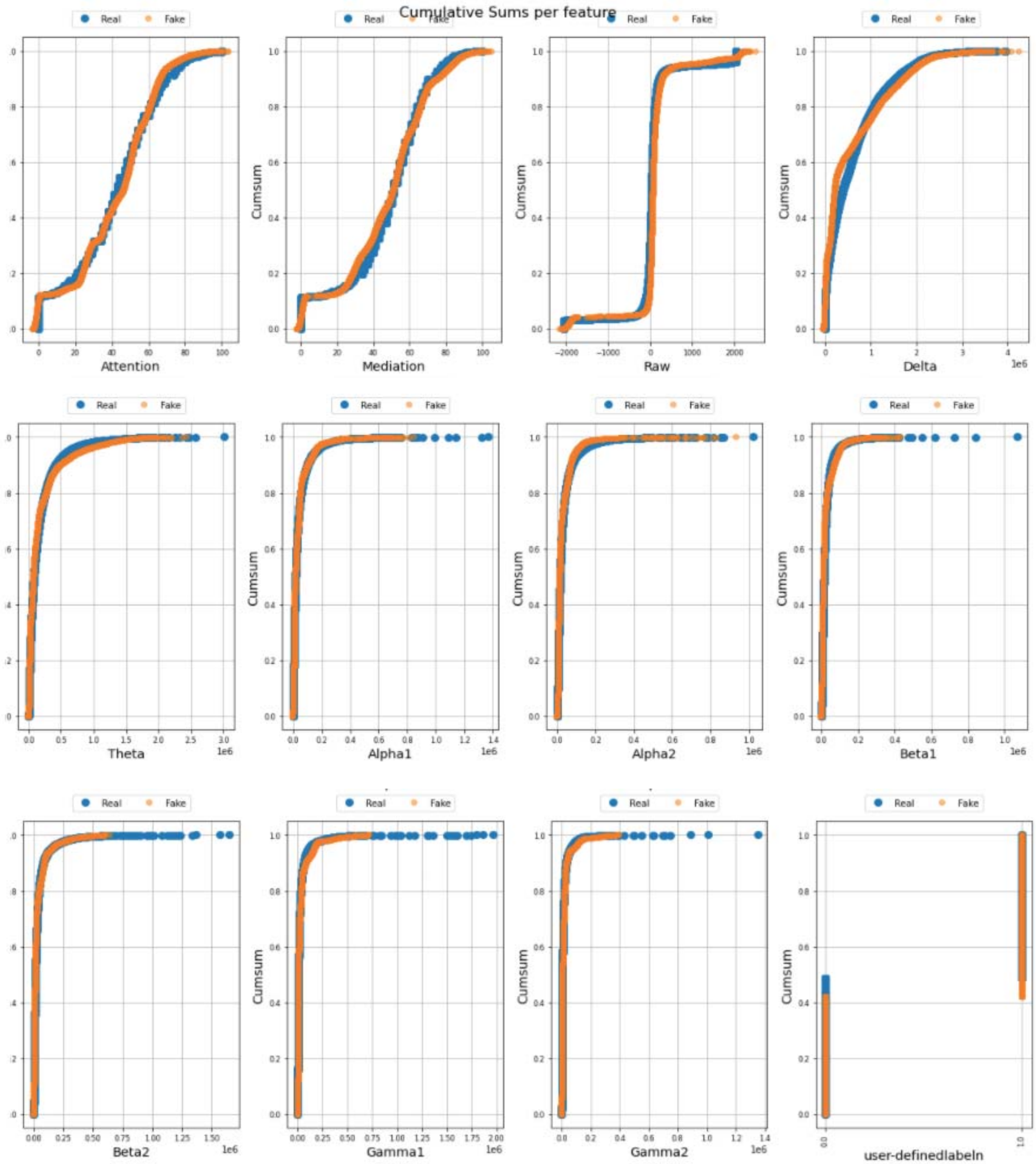
3. RESULTS

3.1 Visualization

In this paper, we carry out experiments to generate EEG data using TGAN and CTGAN and evaluate the efficiency of the generated data. The qualities of the generated data are evaluated by visualization plots using TableEvaluator library and various machine learning algorithms. CTGAN and GAN created 12811 samples respectively and the plots show how real data and synthetic data correlate. As it appears on the visualizations in Figure 4 and Figure 5, they show a feature by feature comparison between the generated data and actual data. Considering the cumulative sums per feature of CTGAN in Figure 4, most of the features in the synthetic data match closely with actual data including attention, medication, delta and user-defined label. The distribution per feature and difference plot of CTGAN also show that the similarity between the original data and the generated data is quite high. Especially in First two components of PCA(Principal Component Analysis) plot, it seems that the distribution of real and synthetic data is almost identical. Likewise, we find that the synthetic data generated by TGAN is also similar to real data even though it's less similar than CTGAN as shown in Figure 5. Although the cumulative sum(cumsum) and distribution per feature plots are similar, the difference of each feature and First two components of PCA plots seem slightly different. To sum up, when visualizing the synthetic data produced by CTGAN and TGAN, the distribution of each feature appears similar, and CTGAN generates the synthetic data more identical to the actual data than TGAN[10]

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195




Cumulative Sums per feature
Real
Fake
Cumsum
Attention
Mediation
Raw
Delta
Theta
Alpha1
Alpha2
Beta1
Beta2
Gamma1
Gamma2
user-definedlabeln
1e6

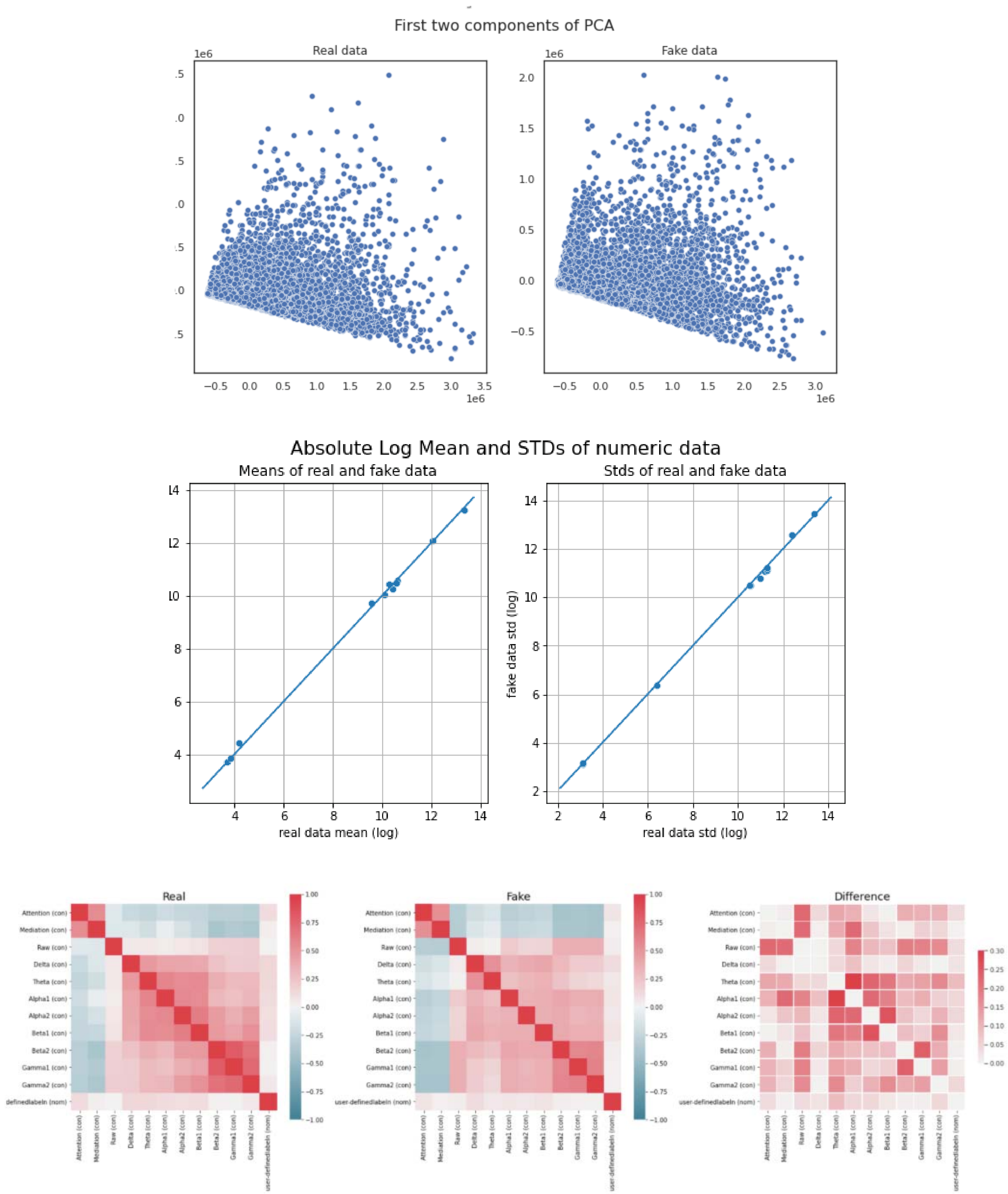
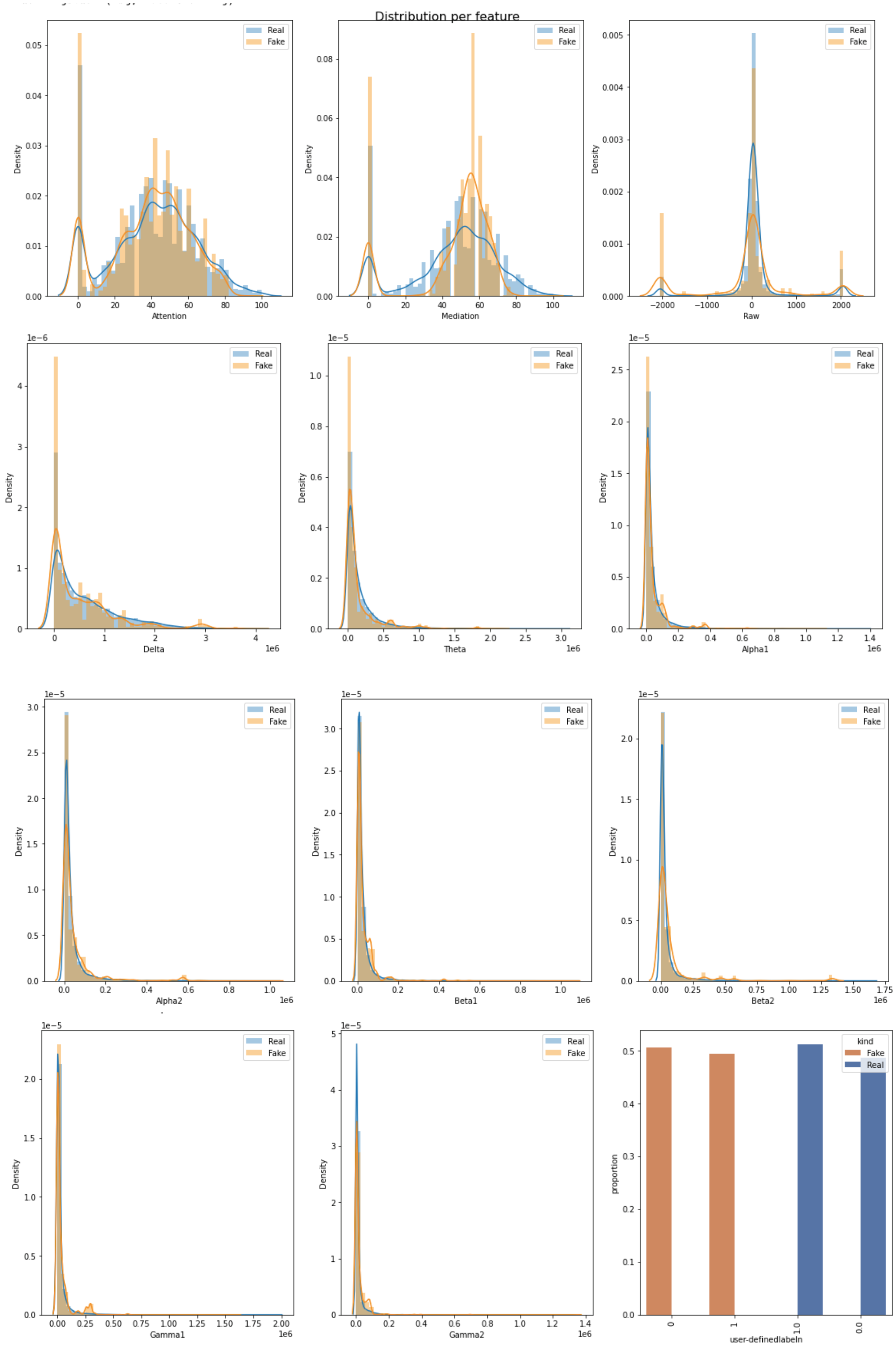
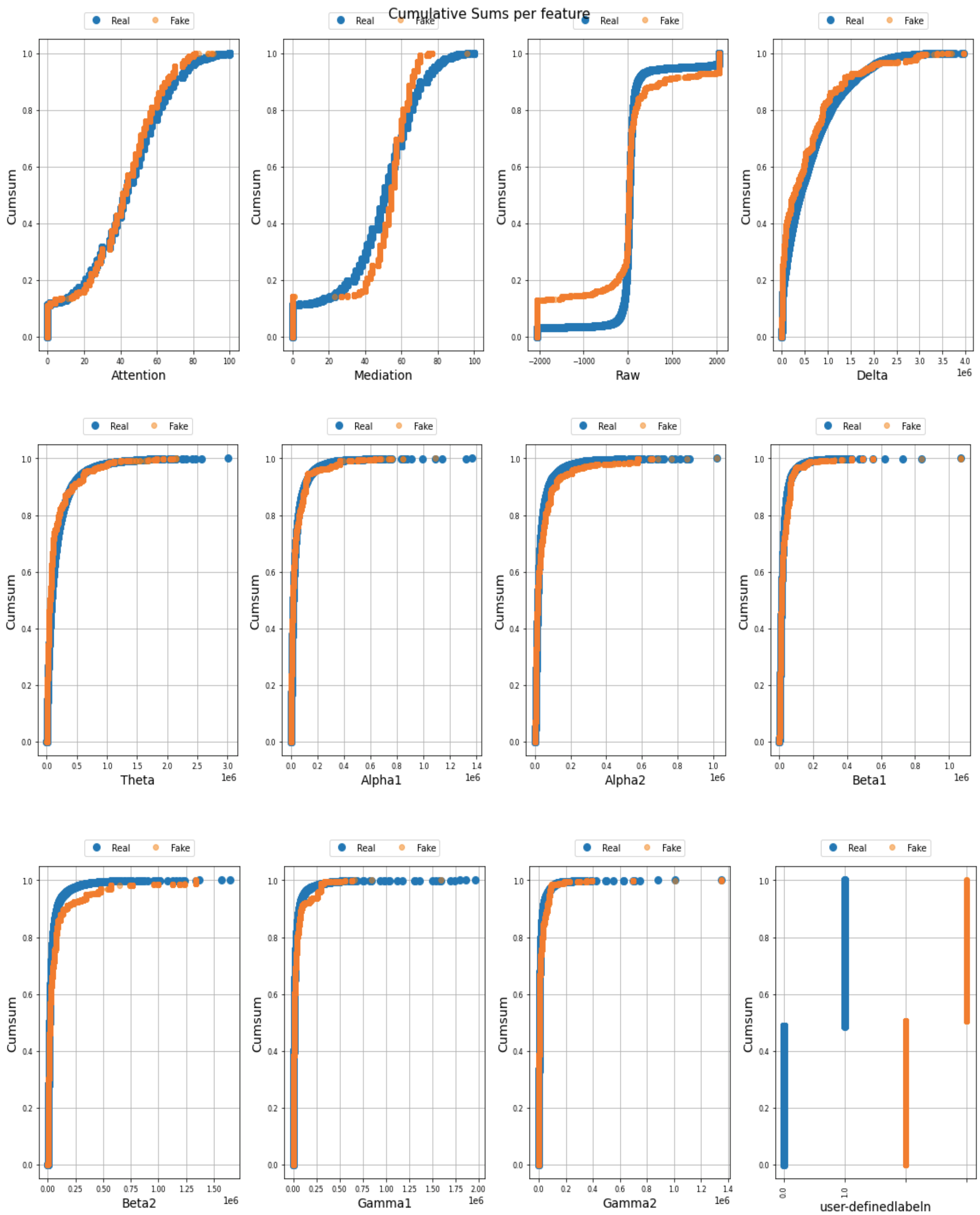


Figure 4. Visualization plots of CTGAN


Distribution per feature
Real
Fake
Density
Attention
Mediation
Raw
Delta
Theta
Alpha1
Alpha2
Beta1
Beta2
Gamma1
Gamma2
1e6
kind
proportion
user-definedlabeln


Cumulative Sums per feature
Real
Fake
Cumsum
Attention
Mediation
Raw
Delta
Theta
Alpha1
Alpha2
Beta1
Beta2
Gamma1
Gamma2
user-definedlabeln
1e6

Journal of Theoretical and Applied Information Technology
31st May 2021. Vol.99. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

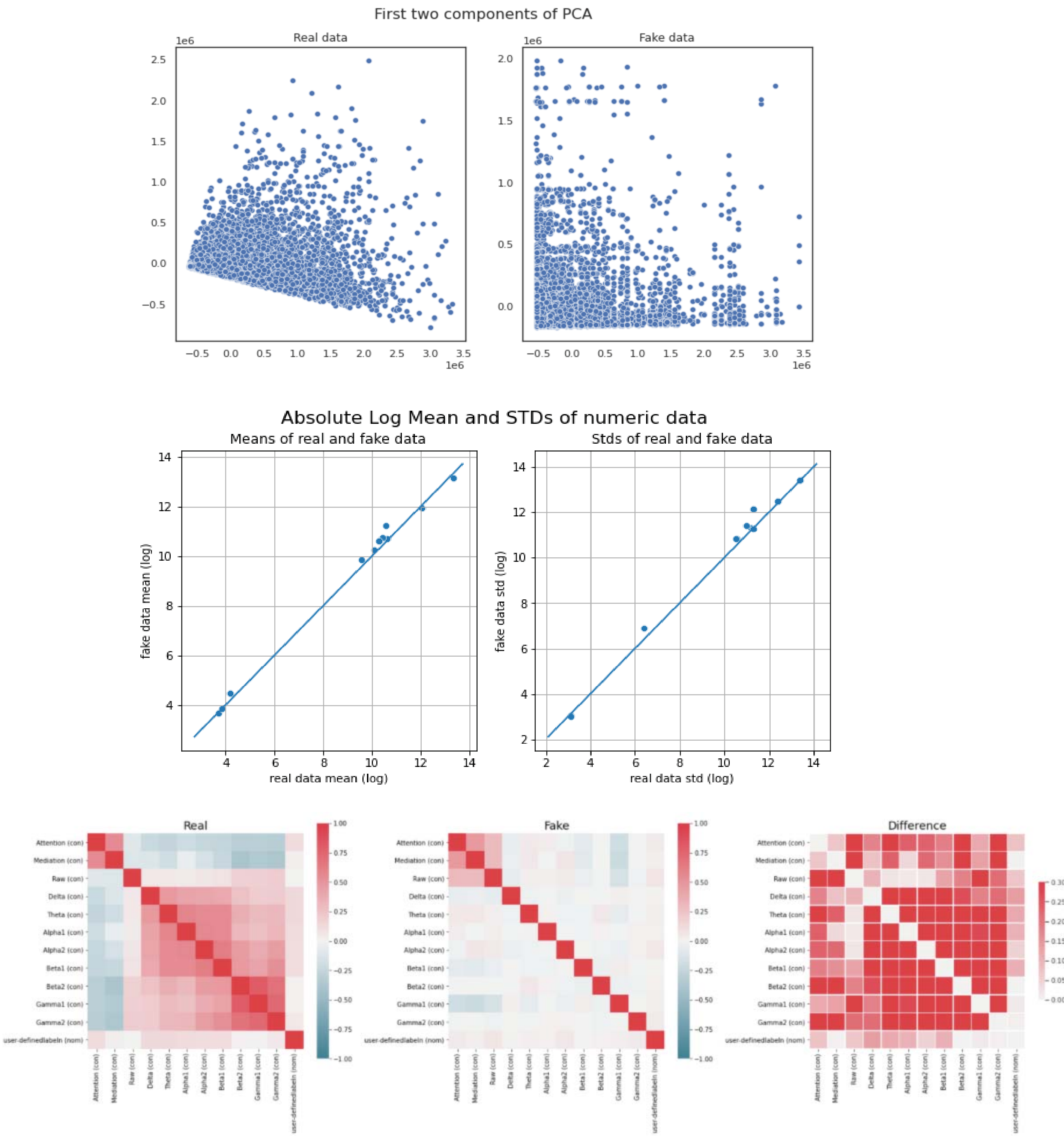


Figure 5. Visualization plots of TGAN

3.2 Similarity Score

We utilize TableEvaluator library for achieving similarity score. The similarity score, which is the average of basic statistics, correction column correction, mirror column correction, 1-MAPE estimator, and 1-MAPE PCA, indicates higher similarity when the score is close to 1. The results show that compared to TGAN, CTGAN gets higher score in basic statistics, correlation column correlations, mean correlation between fake and real columns, 1 – MAPE estimator results and similarity score. It implies that the synthetic data from CTGAN is more similar to the real data than TGAN.

Table 1. Similarity score from CTGAN

Results	Score
Basic Statistics	0.9963
Correlation column correlations	0.9476
Mean Correlation between fake and real columns	0.9393
1 – MAPE Estimator results	0.7250
Similarity Score	0.9021

Table 2. Similarity score from TGAN

Results	Score
Basic Statistics	0.9876
Correlation column correlations	0.0881
Mean Correlation between fake and real columns	0.9351
1 – MAPE Estimator results	0.8552
Similarity Score	0.7165

3.3 Machine Learning Results

As mentioned above, 70% of each 12811 samples produced by tgan and ctgan are used for train dataset and 30% of real data for test dataset. For machine learning algorithms, this paper chooses Random Forest, XGBoost, LightGBM, and Catboost algorithms, and the results are shown in Figure 6 and Figure 7. The expressions of accuracy, precision, recall, f1 and AUC follow the expressions in (1), (2), (3), (4). Based on the f1 score, Catboost on CTGAN data, XGBoost on TGAN data, and Random Forest on real data bring the best results. The confusion matrix for each of these models is shown in Figure 8 and Figure 9. The data generated by TGAN and CTGAN shows similar results in binary classification, and the classification accuracy is generally lower than real data.

$$precision = \frac{TP}{TP + FP} \quad (1)$$

$$recall = \frac{TP}{TP + FN} \quad (2)$$

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (3)$$

$$F1 = \frac{2 * precision * recall}{precision + recall} \quad (4)$$

where TP denotes true – positive, TN denotes true – negative, FP denotes false positive, and FN denotes false negative.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	Random Forest		XGBoost		LightGBM		Catboost	
Trained on	real	synthetic	real	synthetic	real	synthetic	real	synthetic
accuracy	64.75	49.8	61.13	49.17	62.8	49.79	64.2	49.66
precision	72.09	64.45	74.07	68.22	75.01	64.45	74.28	67.96
recall	62.68	49.66	58.67	49.23	60.12	49.66	61.65	49.58
f1	67.06	56.1	65.48	57.19	66.74	56.1	67.38	57.33
AUC	64.78	49.86	61.19	49.26	62.86	49.86	64.25	49.75

Figure 6. Machine learning results from CTGAN

	Random Forest		XGBoost		LightGBM		Catboost	
Trained on	real	synthetic	real	synthetic	real	synthetic	real	synthetic
accuracy	64.75	50.57	61.13	48.75	62.8	49.3	64.2	49.17
precision	72.09	58.23	74.07	68.9	75.01	67.49	74.28	67.49
recall	62.68	50.29	58.67	48.94	60.12	49.31	61.65	49.22
f1	67.06	53.97	65.48	57.23	66.74	56.99	67.38	56.92
AUC	64.78	50.61	61.19	48.85	62.86	49.38	64.25	49.25

Figure 7. Machine learning results from TGAN

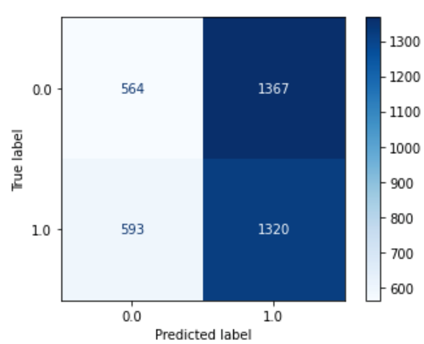


Figure 8: Confusion matrix from CTGAN

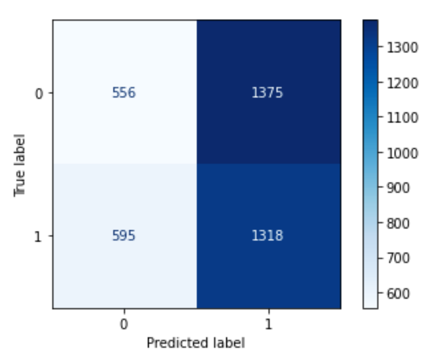


Figure 9: Confusion matrix from TGAN

4. DISCUSSION

4.1 Limitations and Further Considerations

Through the experiments we went through, we found that synthetic data from the CTGAN are more suitable to EEG data than TGAN. However, there is a limitation on our experiment. Even though the visualization steps and similarity steps show us that CTGAN has higher performance in generating synthetic EEG data than TGAN, the machine learning results show that there is no significant difference between them. This means that the visualization and similarity score steps are not directly related to the performance of machine learning. However, as our purpose is to create artificial EEG data and use them as input data for machine learning or deep learning, a further research should be considered. For further research, we can utilize a new GAN which can perform better than TGAN and CTGAN. We can also find another evaluating system which could directly associate to the performance of machine learning or deep learning models.

5. CONCLUSION

Our research shows the possibility of further research for generating synthetic EEG data based on deep learning approach, such as TGAN and CTGAN. Through visualization and similarity score, the EEG data from CTGAN shows higher similarity than TGAN. Unlike the related researches, we tried to use the synthesized dataset as an input data for various machine learning algorithms. However, our study includes a limitation in that when the synthetic data from our experiments are used as the input data for machine learning models, they do not show higher performance compared to the original data. In the future, further research should be conducted to find new deep learning algorithms for generating synthetic data that can perform well as input data.

REFRENCES:

[1] Stegman P, Crawford CS, Andujar M, Nijholt A, Gilbert JE. Brain–Computer interface software: A review and discussion. IEEE Transactions on Human-Machine Systems. 2020;50(2):101–15.

[2] Casson AJ. Wearable EEG and beyond. Biomedical Engineering Letters. 2019;9(1):53–71.

[3] Luo T-jian, Fan Y, Chen L, Guo G, Zhou C. EEG signal RECONSTRUCTION using a Generative adversarial network with Wasserstein distance and Temporal-spatial-frequency Loss. Frontiers in Neuroinformatics. 2020;14.

[4] Erkorkmaz K. Efficient Fitting of the Feed Correction Polynomial for Real-Time Spline Interpolation. Journal of Manufacturing Science and Engineering. 2015;137(4):044501.

[5] Naldi A, Larive RM, Czerwinska U, Urbach S, Montcourrier P, Roy C, et al. Reconstruction and signal propagation analysis of the Syk signaling network in breast cancer cells. PLOS Computational Biology. 2017;13(3).

[6] Jin KH, McCann MT, Froustey E, Unser M. Deep Convolutional Neural Network for Inverse Problems in Imaging. IEEE Transactions on Image Processing. 2017;26(9):4509–22.

[7] Luo T-jian, Fan Y, Chen L, Guo G, Zhou C. EEG Signal Reconstruction Using a Generative Adversarial Network With Wasserstein Distance and Temporal-Spatial-Frequency Loss. Frontiers in Neuroinformatics. 2020;14.

[8] Luo Y, Lu B-L. EEG Data Augmentation for Emotion Recognition Using a Conditional Wasserstein GAN. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2018;

[9] Hartmann, Kay & Schirrmeister, Robin & Ball, Tonio. EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals. 2018;

[10] Brenninkmeijer, Bauke & Hille, Youri & Vries, Arjen. On the Generation and Evaluation of Tabular Data using GANs. 2019;

[11] Wang, H., Li, Y., Hu, X., Yang, Y., Meng, Z., & Chang, K. M. Using EEG to Improve Massive Open Online Courses Feedback Interaction. In AIED Workshops. 2013;

[12] Park N, Mohammadi M, Gorde K, Jajodia S, Park H, Kim Y. Data synthesis based on generative adversarial networks. Proceedings of the VLDB Endowment. 2018;11(10):1071–83.

[13] Xu, L.; Skoularidou, M.; Cuesta-Infante, A.; Veeramachaneni, K. Modeling tabular data using Conditional GAN. In Advances in Neural Information Processing Systems; NIPS, Vancouver Convention Centre: Vancouver,BC, Canada, 2019; pp. 7333–7343.