

Appendix A Supplementary Material

Datasets	GAN Variants	Classifiers	Dataset Type	Cross-validation Value	Epochs	Noise	Batch Size	Learning Rate	
BCW	GAN	XGBoost	Synthetic Dataset	CV=10	200	264	128	5,00E-06	
	CGAN				1000	264	128	5,00E-06	
	CTGAN				1000	264	128	5,00E-06	
	CRAMERGAN				1000	264	128	5,00E-06	
	DRAGAN				1000	264	128	5,00E-06	
	WGAN				1000	264	128	5,00E-06	
	GAN		Mixed Dataset		300	264	128	5,00E-06	
	CGAN				300	264	128	5,00E-06	
	CTGAN				700	264	128	5,00E-06	
	CRAMERGAN				1000	264	128	5,00E-06	
	DRAGAN				500	264	128	5,00E-06	
	WGAN				100	264	128	5,00E-06	
LC	GAN	XGBoost	Synthetic Dataset	CV=10	300	100	128	5,00E-06	
	CGAN				300	100	128	5,00E-06	
	CTGAN				300	100	128	5,00E-06	
	CRAMERGAN				1000	100	128	5,00E-06	
	DRAGAN				900	100	128	5,00E-06	
	WGAN				900	100	128	5,00E-06	
	GAN		Mixed Dataset		1000	100	128	5,00E-06	
	CGAN				300	100	128	5,00E-06	
	CTGAN				900	100	128	5,00E-06	
	CRAMERGAN				300	100	128	5,00E-06	
	DRAGAN				100	100	128	5,00E-06	
	WGAN				70	100	128	5,00E-06	
CTG	GAN	XGBoost	Synthetic Dataset	CV=5	100	264	128	1,00E-05	
	CGAN				100	264	128	1,00E-05	
	CTGAN				100	264	128	1,00E-05	
	CRAMERGAN				100	264	128	1,00E-05	
	DRAGAN				100	264	128	1,00E-05	
	WGAN				100	264	128	1,00E-05	
	GAN		Mixed Dataset		100	264	128	5,00E-06	
	CGAN				100	264	128	1,00E-05	
	CTGAN				200	264	128	1,00E-05	
	CRAMERGAN				CV=5	100	264	128	5,00E-06
	DRAGAN				100	264	128	1,00E-05	
	WGAN				CV=10	200	264	128	1,00E-05

Table A1: Optimal Hyperparameters for GAN Models and XGBoost Classifier on Selected Datasets

Datasets	GAN Variants	Classifiers	Dataset Type	Cross-validation Value	Epochs	Noise	Batch Size	Learning Rate
BCW	GAN	SVM	Synthetic Dataset	CV=10	200	264	128	5,00E-06
	CGAN				1000	264	128	5,00E-06
	CTGAN				700	264	128	5,00E-06
	CRAMERGAN				1000	264	128	5,00E-06
	DRAGAN				1000	264	128	5,00E-06
	WGAN				1000	264	128	5,00E-06
	GAN		Mixed Dataset		500	264	128	5,00E-06
	CGAN				300	264	128	5,00E-06
	CTGAN				700	264	128	5,00E-06
	CRAMERGAN				1000	264	128	5,00E-06
	DRAGAN				100	264	128	5,00E-06
	WGAN				300	264	128	5,00E-06
LC	GAN	SVM	Synthetic Dataset	CV=10	100	100	128	5,00E-06
	CGAN				1000	100	128	5,00E-06
	CTGAN				300	100	128	5,00E-06
	CRAMERGAN				100	100	128	5,00E-06
	DRAGAN				300	100	128	5,00E-06
	WGAN				100	100	128	5,00E-06
	GAN		Mixed Dataset		100	100	128	5,00E-06
	CGAN				300	100	128	5,00E-06
	CTGAN				1000	100	128	5,00E-06
	CRAMERGAN				300	100	128	5,00E-06
	DRAGAN				100	100	128	5,00E-06
	WGAN				100	100	128	5,00E-06
CTG	GAN	SVM	Synthetic Dataset	CV=5	100	264	128	1,00E-05
	CGAN				100	264	128	1,00E-05
	CTGAN				1000	264	128	1,00E-05
	CRAMERGAN				1000	264	128	1,00E-05
	DRAGAN				1000	264	128	1,00E-05
	WGAN				100	264	128	1,00E-05
	GAN		Mixed Dataset	CV=10	200	264	128	1,00E-05
	CGAN				100	264	128	1,00E-05
	CTGAN				200	264	128	1,00E-05
	CRAMERGAN			CV=5	200	264	128	1,00E-05
	DRAGAN				200	264	128	5,00E-06
	WGAN				100	264	128	1,00E-05

Table A2: Optimal Hyperparameters for GAN Models and SVM Classifier on Selected Datasets

For the BCW dataset, from the results of pairwise correlations between features could clearly see some pairs of features have a relatively high correlation; for features that are highly correlated with each other, it is suggested to remove highly correlated features to avoid a predictive bias for the information contained in these features. Additionally, features exhibit multicollinearity, which can lead to numerical instability during analysis, also features with right-skewed distributions, which might not provide much valuable information; they are candidate for removal [71][72][73][74].

Dataset	Dropped Features
BCW	id, Unnamed: 32, area_mean, radius_mean, perimeter_worst, radius_worst, area_worst, radius_se, perimeter_se, concave points_mean, compactness_worst, concave points_se, concavity_se, concavity_worst, fractal_dimension_worst, smoothness_mean, symmetry_mean
LC	Patient Id
CTG	FileName, Date, SegFile, b, e, A, B, C, D, E, AD, DE, LD, FS, SUSP, CLASS, DR

Table A3: List of the Dropped Features of Each Dataset

Datasets	Dataset Type	GAN	CGAN	CTGAN	CRAMERGANDRAGAN	WGAN
BCW Dataset	Original DS	569	569	569	569	569
	Synthetic DS	570	641	570	570	570
	Mixed DS	1139	1210	1139	1139	1139
LC Dataset	Original DS	1000	1000	1000	1000	1000
	Synthetic DS	1001	1153	1001	1001	1001
	Mixed DS	2001	2153	2001	2001	2001
CTG Dataset	Original DS	2129	2129	2129	2129	2129
	Synthetic DS	2127	2305	2127	2127	2127
	Mixed DS	4256	4434	4256	4256	4256

Table A4: Comparison of dataset sizes across different GAN models.

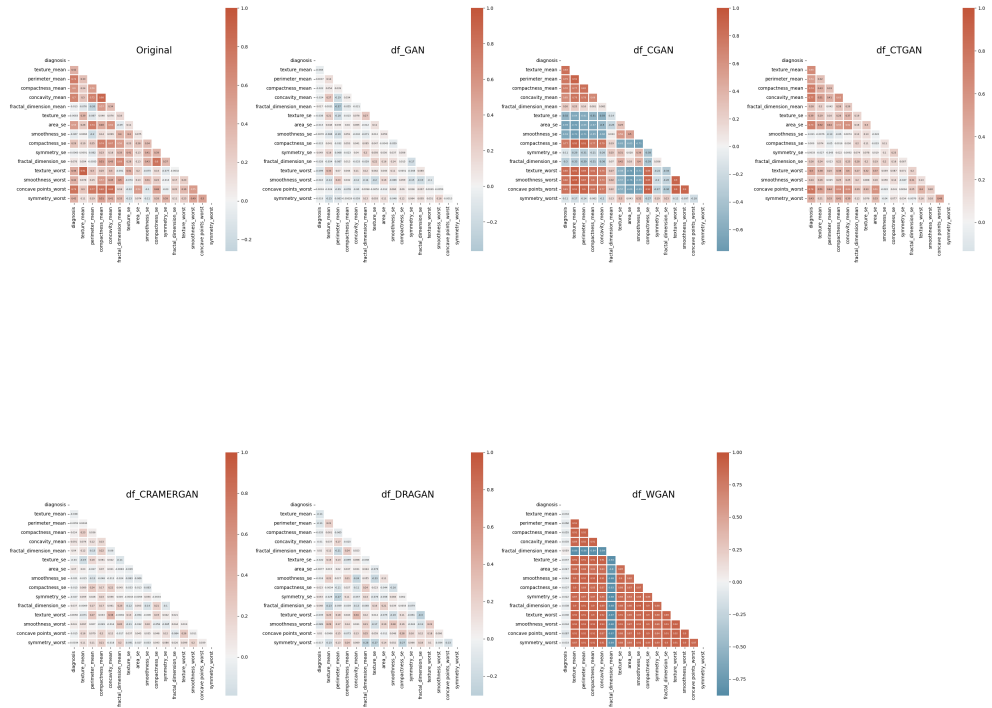


Fig A1: Pearson Correlation Heatmaps: BCW Original and Synthetic Datasets - A Comparative Analysis Across Selected GAN Variants

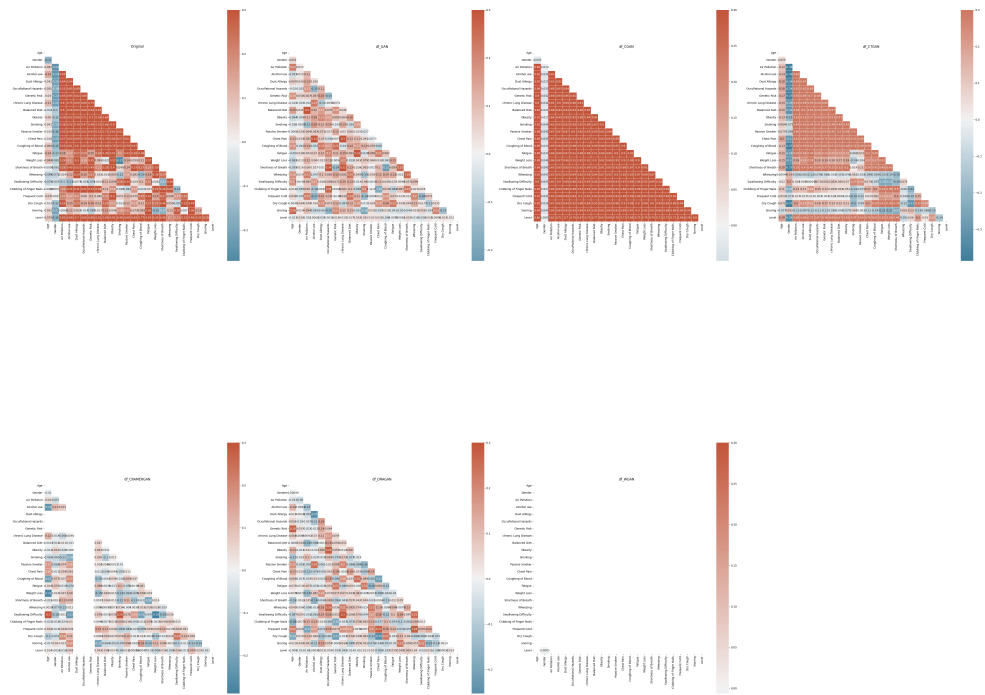


Fig A2: Spearman's Rank Correlation Heatmaps: LC Original and Synthetic Datasets
- A Comparative Analysis Across Selected GAN Variants

