Supplementary Materials -
RadialGAN: Leveraging multiple datasets to improve target-specific predictive models using Generative Adversarial Networks

Jinsung Yoon [1] **James Jordon** [2] **Mihaela van der Schaar** [1 2 3]

1. Additional Experiments

Table 1. Statistics of 14 studies

Study Label	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Feature No	77	86	72	77	58	80	41	99	81	73	71	13	47	47
Avg. Overlap (%)	46.6	47.7	49.0	55.9	49.6	55.2	61.9	47.3	52.8	47.5	55.6	71.0	60.6	51.9
Logistic Regression Base														
Target Only AUC	.836	.805	.655	.721	.605	.687	.654	.635	.756	.631	.664	.963	.681	.750
Target Only APR	.192	.426	.373	.122	.290	.194	.295	.122	.210	.251	.420	.899	.260	.445
Multi-layer Perceptron Base														
Target Only AUC	.790	.812	.688	.687	.638	.748	.603	.829	.824	.588	.686	.977	.735	.770
Target Only APR	.179	.392	.417	.117	.320	.234	.254	.270	.245	.218	.431	.924	.263	.436

Table 1 contains a summary of each of the 14 selected datasets. It also contains the AUC and APR achieved by the Target-Only model on each dataset which is used as the baseline in Fig. 1, where we report the results of training a multi-layer perceptron (MLP) instead of the logistic regression used in the main paper. As can be seen from the figure, the results are very similar for the MLP, with RadialGAN outperforming the two benchmarks in almost all datasets.

In Fig. 2 we compare the performance of training a predictive model when additional samples are generated by RadialGAN versus simply having additional true samples (i.e. from the target dataset), which would be the best thing we could possibly have (a model whose goal is to generate additional samples cannot hope to generate samples that are better than true samples). 100 samples are taken from the target dataset, and RadialGAN then generates additional ones using other datasets, whereas True Datasets simply uses additional samples from the target dataset. As can be seen from the figure, the performance improvement from RadialGAN is almost as good as from simply having more true samples, indicating that RadialGAN is able to generate additional samples that provide almost as much information to a predictive model as the true samples would. In the figure, Target-Only is the performance of a predictive model trained only on the initial 100 samples (i.e. it makes no use of any additional samples), and is given to indicate the effect that generating no further samples would have.

[1]University of California, Los Angeles, CA, USA [2]University of Oxford, UK [3]Alan Turing Institute, UK. Correspondence to: Jinsung Yoon <jsyoon0823@gmail.com>.
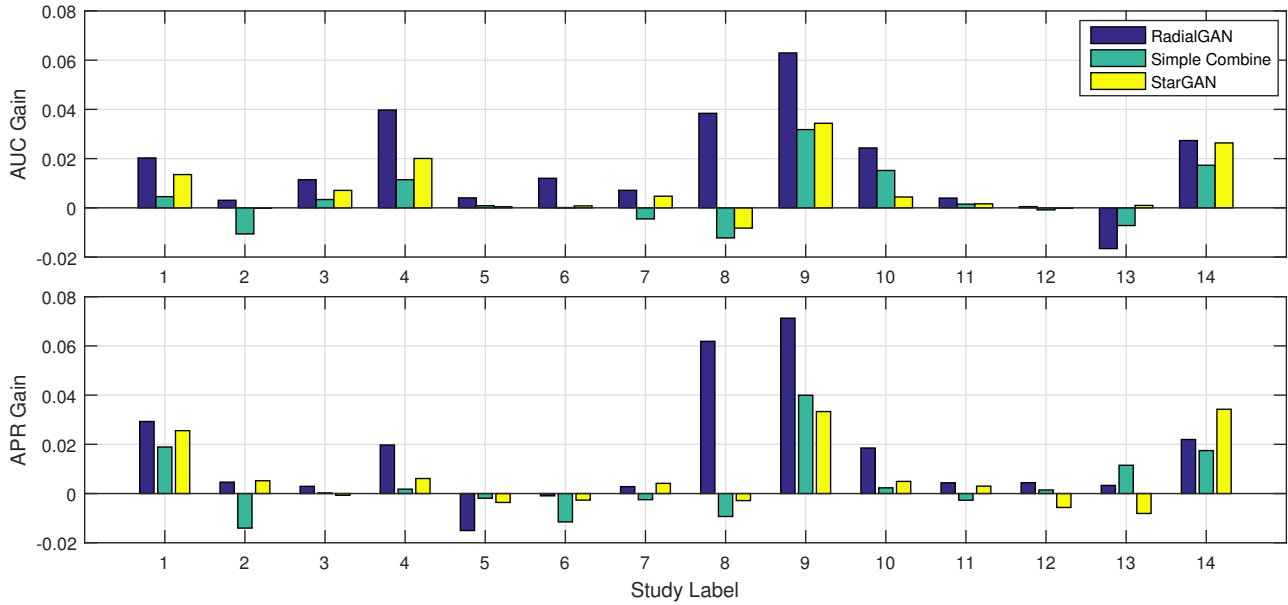
Figure 1. Performance improvements over Target-Only baseline for all 14 studies using a multi-layer perceptron as the predictive model.
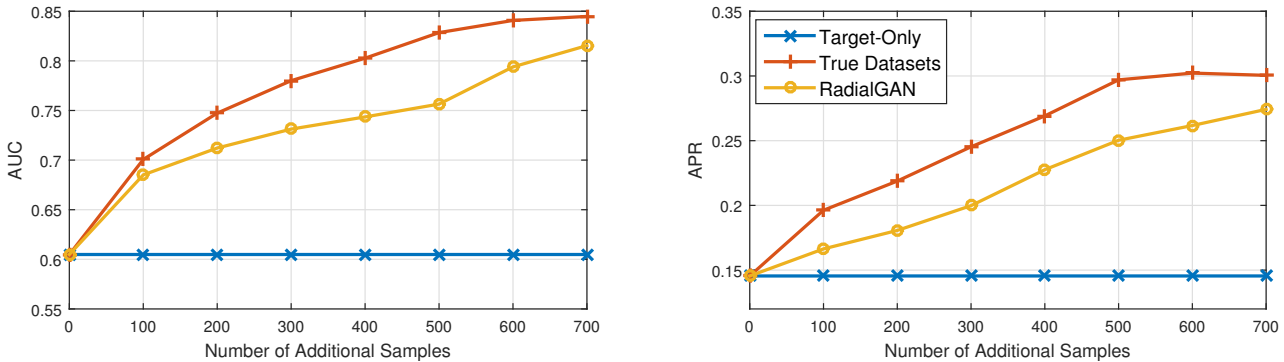

Figure 2. Comparison of the prediction accuracies when using RadialGAN for extra samples versus using extra samples from the target dataset (True Datasets). This figure demonstrate that the quality of samples generated by RadialGAN are almost as good as those from the true datasets. Target-Only is given as reference and is the performance of a predictive model trained using only the initial 100 samples.

2. Hyper-parameter Optimization

In all experiments described in the main manuscript, the depth of the generator and discriminator in all benchmarks is set to 3. The depth of all encoders, decoders and discriminators in RadialGAN is also set to 3. For all the benchmarks and RadialGAN, the number of hidden nodes in the first, second and third hidden layers is d, $d/2$ and d (where d is the dimension of the input), respectively. We use $\tanh$ as the activation function for each layer except for the output layer where we use the sigmoid activation function and the size of all mini-batches is 64 (i.e. $k_G = k_D = 64$). For RadialGAN, we use cross-validation to select λ, β among $\{0.1, 0.5, 1, 2, 10\}$ and also to select the dimension of $\mathcal{Z}$ from $\{\max_i(d_i), \max_i(\frac{d_i}{2}), \max_i(\frac{d_i}{4})\}$ where d_i is the dimension of $\mathcal{X}^{(i)}$. For CycleGAN (Zhu et al., 2017), co-GAN (Mirza & Osindero, 2014), and StarGAN (Choi et al., 2017), we also use cross-validation to select λ among $\{0.1, 0.5, 1, 2, 10\}$. (Wiens et al., 2014) does not have any hyper-parameters to be optimized. We use tensorflow to implement all the benchmarks and RadialGAN.

References

Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., and Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. *arXiv preprint arXiv:1711.09020*, 2017.

Mirza, M. and Osindero, S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

Wiens, J., Guttag, J., and Horvitz, E. A study in transfer learning: leveraging data from multiple hospitals to enhance hospital-specific predictions. *Journal of the American Medical Informatics Association*, 21(4):699–706, 2014.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.