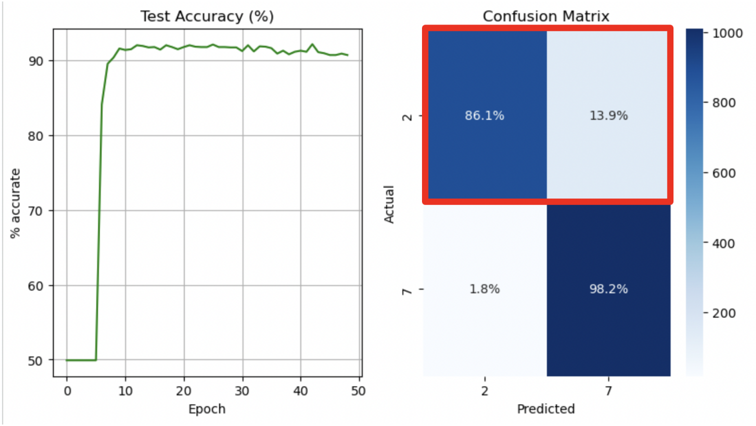
Fair Healthcare Data Using a Conditional DDPM

Keira Behal, Alina Chen, Caleb Fikes, Sophia Xiao

Emory University

June 2023

Problem: Imbalanced Datasets

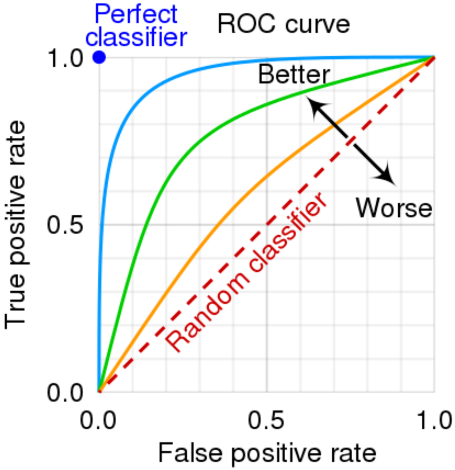


- Models trained on such imbalanced datasets often exhibit a **bias** towards the majority class. (High accuracy for majority class, low accuracy for minority classes.)
- **Can lead to inequitable healthcare outcomes.**

Fairness Metrics

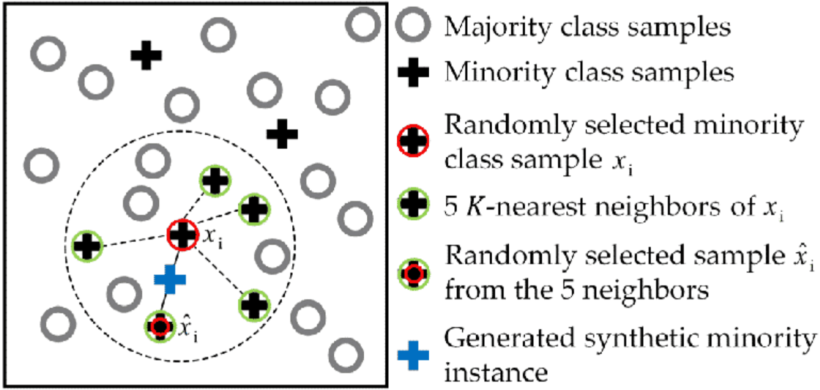
Accuracy is **not** a good indicator of fairness.

Metric	Formula
Precision	$\frac{T_p}{T_p+F_p}$
Recall (Sensitivity)	$\frac{T_p}{T_p+F_n}$
F1 Score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$
AUROC	$\int_0^1 \mathbf{ROC}(x)\, dx$



https://upload.wikimedia.org/wikipedia/commons/thumb/1/13/Roc_curve.svg/440px-Roc_curve.svg.png

Existing Method: SMOTE Algorithm



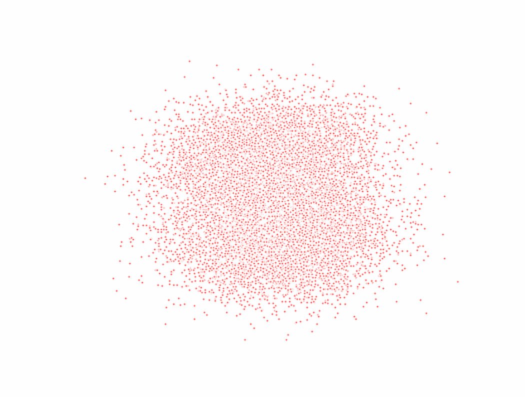
`https://rikunert.com/smote_explained`

Our SMOTE implementation:

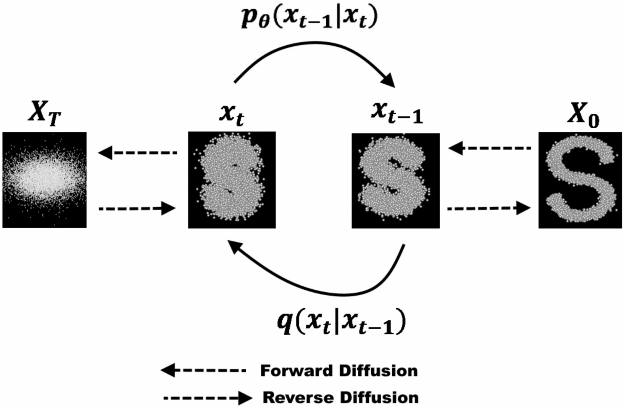
- Number of minority class samples: $T \approx 1000$
- Number of nearest neighbors: $K = 5$
- N %, amount of SMOTE: 90%

Why Denoising Diffusion Probabilistic Model?

https://openai.com/dall-e-2?ref=assemblyai.com



Denoising Diffusion Probabilistic Model



Forward (noising) process:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}).$$

Denoising process:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \sim \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t))$$

https://miro.medium.com/v2/resize:fit:1400/format:webp/1*MLRitrnUdOy6rPtZfNwO4w.png

Forward Diffusion Process

Definition:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

Simplification:

$$q(\mathbf{x}_t|\mathbf{x}_0) \sim \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$$

$$\text{where } \alpha_t := 1-\beta_t \text{ and } \bar{\alpha}t := \Pi_{s=1}^{t}\alpha_s$$

Sampling is easy!

$$\mathbf{x}_t|\mathbf{x}_0 = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$$

Denoising Diffusion Process

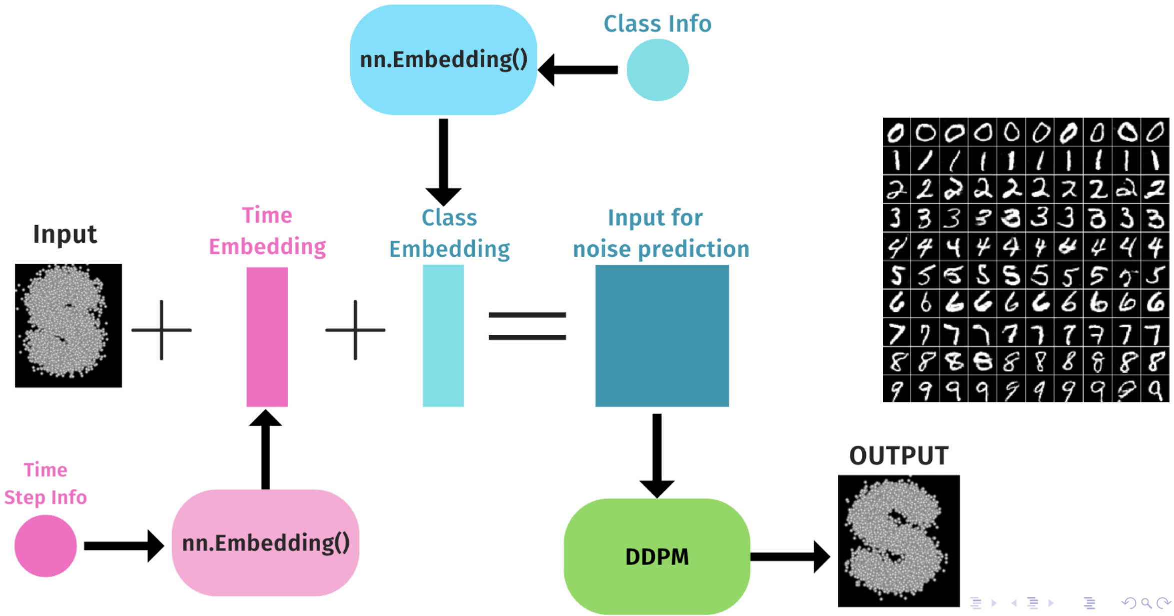
Parameterize Mean:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}(\mathbf{x_t} - \frac{\beta_t}{\sqrt{1-\alpha_t}}\boldsymbol{\epsilon}_\theta(x_t, t))$$

Loss Function:

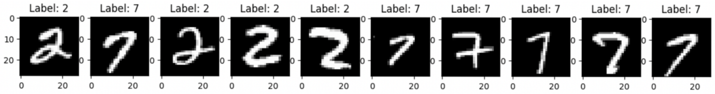
$$L_{\text{simple}} = ||\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}}\boldsymbol{\epsilon}, t)||^2$$

Conditional DDPM AT ONE SPECIFIC TIME STEP

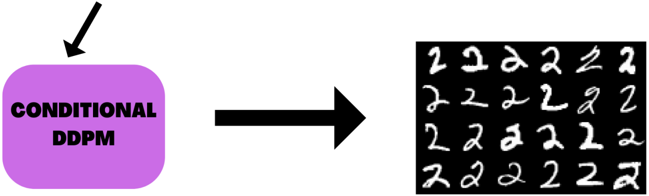


DATA AUGMENTATION PROCESS

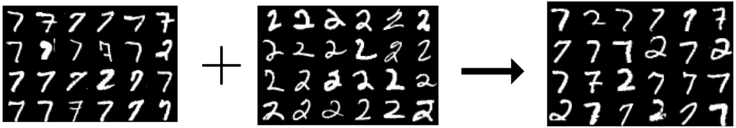
STEP 0



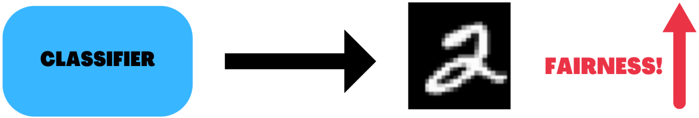
STEP 1



STEP 2



STEP 3



MNIST Experiment

Experiment

Test Conditional DDPM on 2 classes of MNIST

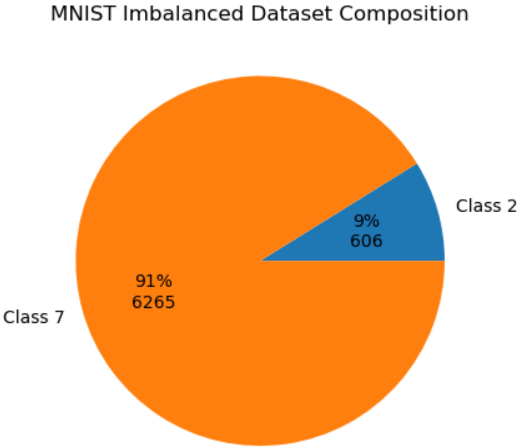
Control Methods

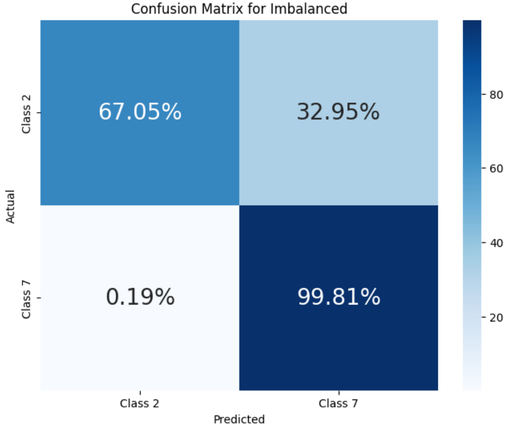
- Unbalanced data: choose 100% of class 7 samples and 10% of class 2 samples
- Balanced data: even split between class 7 and class 2 (Original data)

Augmentation Methods

- Data augmented by SMOTE
- Data augmented by DDPM (Aug_DDPM)
- Fully DDPM-generated data (Full_gen)

Biased Classifier due to training on Imbalanced Data





Confusion Matrices

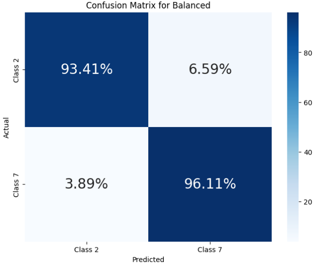


Figure: Original Balanced Data

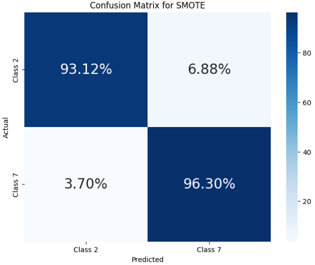


Figure: Augmented by SMOTE

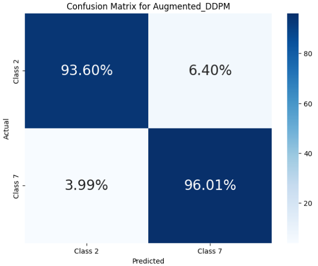


Figure: Augmented by DDPM

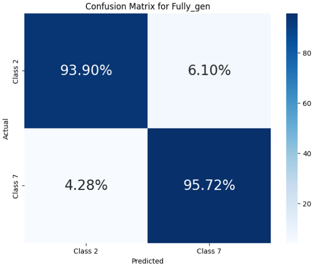


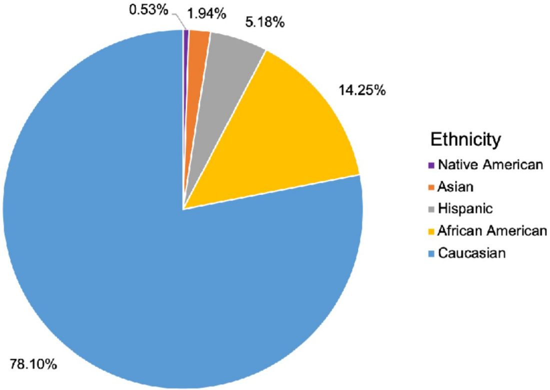
Figure: Fully DDPM-generated Data

Results: Fairness Metrics

	F1	Precision	Recall	AUROC
Unbalanced	0.829	0.751	0.998	0.834
Balanced	0.948	0.936	0.961	0.948
SMOTE	0.947	0.933	0.963	0.947
Aug_DDPM	0.948	0.937	0.960	0.948
Full_gen	0.948	0.940	0.957	0.948

Table: Performance of Augmented Datasets

Next Steps: Electronic Health Records (EHR)



Future Questions:

- Will a Conditional DDPM be superior to existing methods for EHR?
- How can we account for multiple classes in our implementation?
- How do we assess the quality of synthetically generated tabular data?

Bibliography

LFW Face Dataset: http://vis-www.cs.umass.edu/lfw/information
CelebA celebrity face dataset: https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html
Fashion MNIST model example: https://machinelearningmastery.com/how-to-develop-a-cnn-from-scratch-for-fashion-mnist-clothing-classification/
Improved DDPM (Cosine noise schedule, fast sampling, log likelihood optimization): https://arxiv.org/pdf/2102.09672.pdf
GAN paper (ICLR 2019): https://arxiv.org/pdf/1809.11096.pdf
DDPM vs. GAN: https://arxiv.org/pdf/2212.07501.pdf
Unet paper (2015): https://arxiv.org/pdf/1505.04597.pdf
FID introduced (2017): https://arxiv.org/pdf/1706.08500.pdf
Debiassing image enhancement: https://arxiv.org/pdf/2106.10080.pdf
Fairness measures: https://arxiv.org/abs/1808.00023
Call for open science: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8515002/
Fairness of MIMIC-III: https://www.nature.com/articles/s41597-021-01110-7

Bibliography 2

bias removal via language models: https://arxiv.org/pdf/2301.11104v2.pdf
Analyzing Bias in Diffusion-based Face Generation Models:
https://arxiv.org/pdf/2305.06402.pdf
Fair Diffusion: Instructing Text-to-Image Generation Models on Fairness:
https://arxiv.org/abs/2302.10893
TabTransformer Architecture (cs.LG 2020): https://arxiv.org/pdf/2012.06678.pdf
Adversarial Purification (probably unrelated): https://arxiv.org/pdf/2205.07460.pdf
Analyzing bias in face-generation models: https://arxiv.org/pdf/2305.06402.pdf
DDIM paper: https://arxiv.org/pdf/2010.02502.pdf
Debiasing via disentanglement of latent representations:
https://arxiv.org/pdf/2202.00391.pdf
comparison of DDPM and VAE's: https://sander.ai/2022/01/31/diffusion.html
DDPM with Tabular data: https://arxiv.org/pdf/2209.15421.pdf