Privacy-preserving Generative Deep Neural Networks Support Clinical Data Sharing

Yianni Laloudakis
David Morales
Caleb Geniesse
Arjun Balasingam

CS 273B, Group 2, Paper Review

About the Paper

- Title: “Privacy-preserving generative deep neural networks support clinical data sharing”
- Who: B. Jones, Z. Wu, C. Williams, C. Greene @ UPenn
- Submitted to BioRxiv in July 2017

- Claim: *“Deep neural networks can generate shareable biomedical data to allow reanalysis while preserving the privacy of study participants.”*

Introduction & Background

Challenges

- Lots of research showing potential in data-driven medicine
- Many rich datasets, with lots of personal information

- Sharing datasets is challenging
    - Many steps to establish formal collaboration and agree on usage requirements

Prior Work

- Evaluation of benefits of clinical trial data sharing
    - SPRINT Trial Data Challenge lead to personalized treatment and decision support systems
    - Analysis was still screened by data analysis agreements and privacy protection

- SPRINT Challenge
    - Data from trial comparing different strategies to manage systolic blood pressure
    - Published in NEJM in Nov 2015
    - Challenge participants must *apply* for data

Proposed Approach

- Eliminate technical barriers that hinder data sharing
- Leverage techniques from deep neural networks

- Generative Adversarial Networks (GANs)
    - Two deep NNs (generator + discriminator) trained against each other to generate *simulated yet realistic* patient information
    - Generator creates a participant from a set of random numbers
    - Discriminator labels generator output as ‘real’ or ‘generated’
    - Over training epochs, generator learns how to create samples that fool discriminator

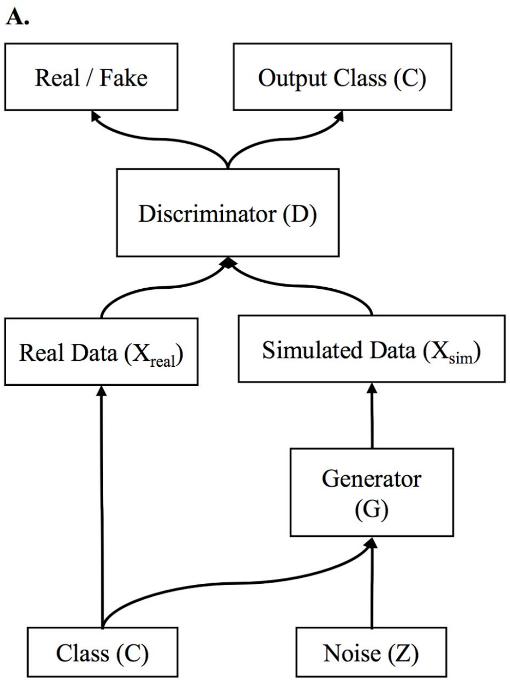
Differential Privacy

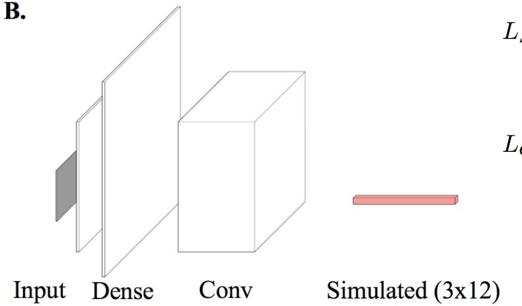
- Incorporated as a constraint on the GANs
    - Limit effect that any particular datapoint has on training
    - Add random noise to participant data

- Generates new individuals without revealing information about any single participant
- Train discriminator with this constraint

Methods

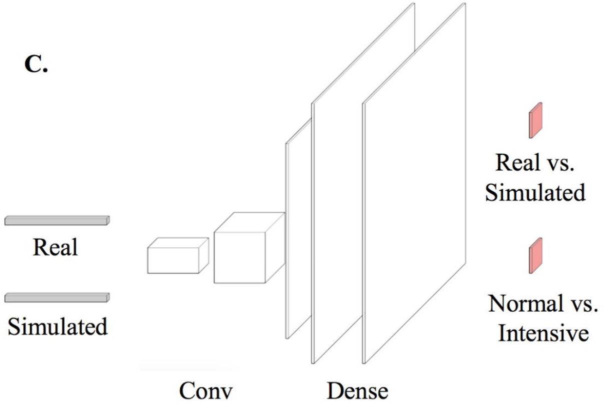
Model Architecture





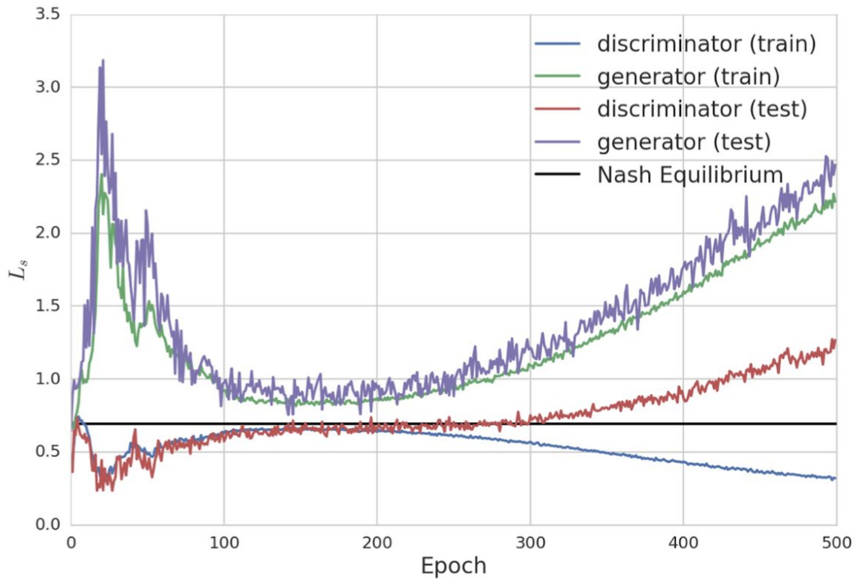
$$L_S = E[\log P(S = real \mid X_{real})] + E[\log P(S = fake \mid X_{fake})]$$

$$L_C = E[\log P(C = c \mid X_{real})] + E[\log P(C = c \mid X_{fake})]$$

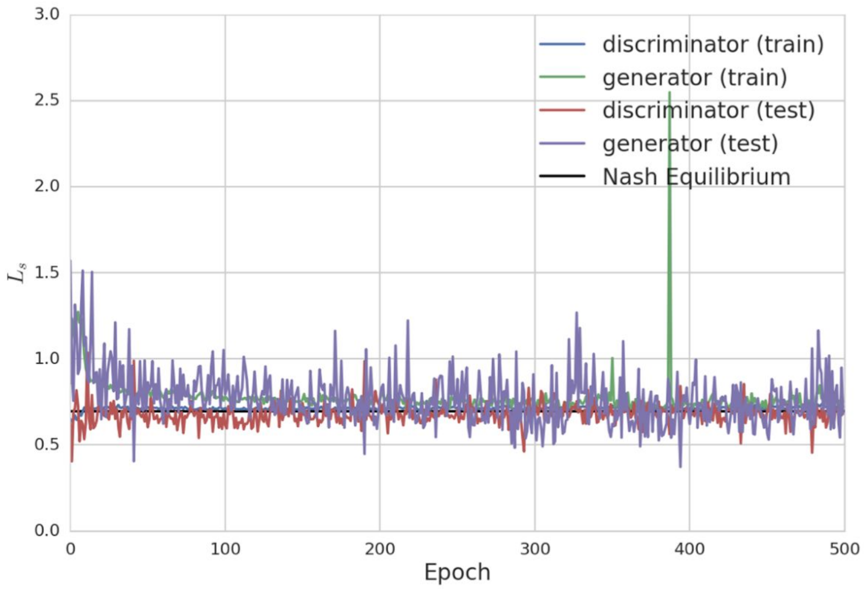


Model Training Process

D.

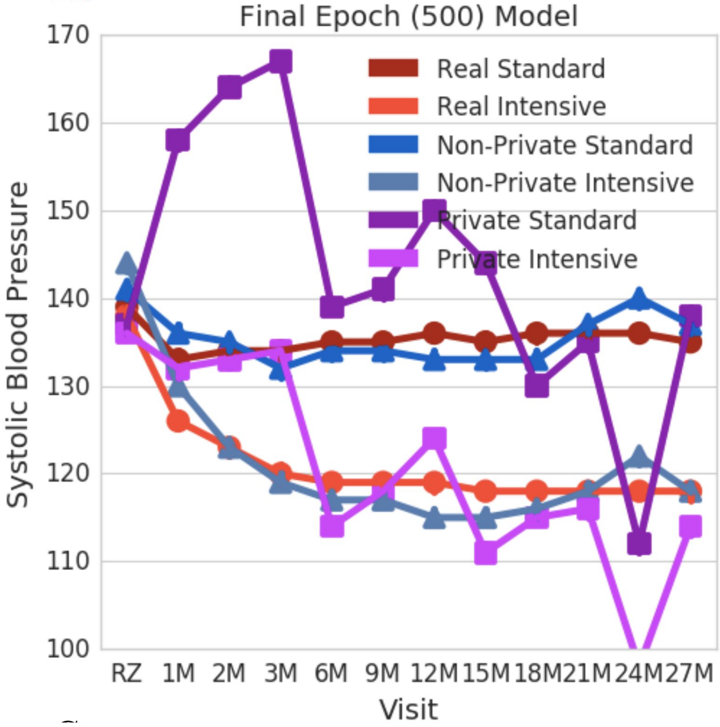


E.

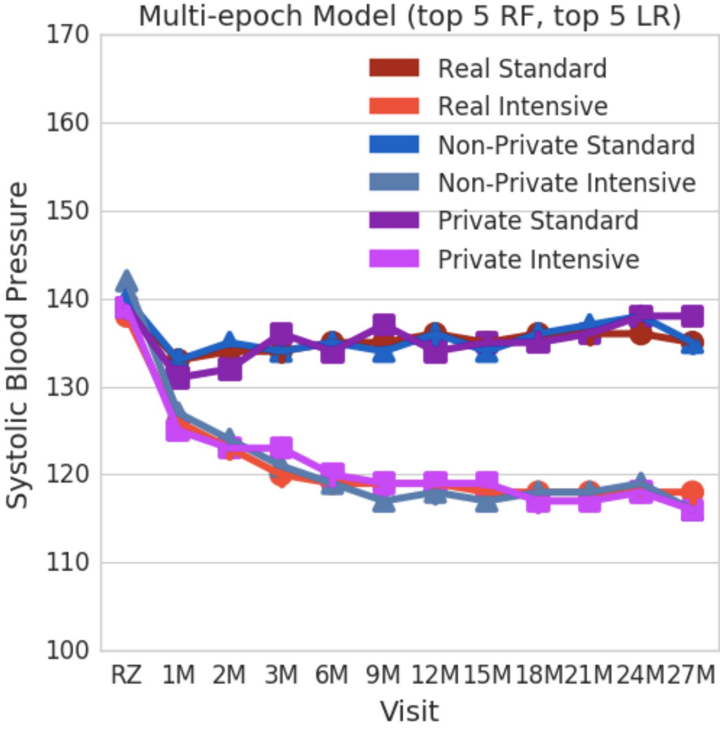


Data Sampling

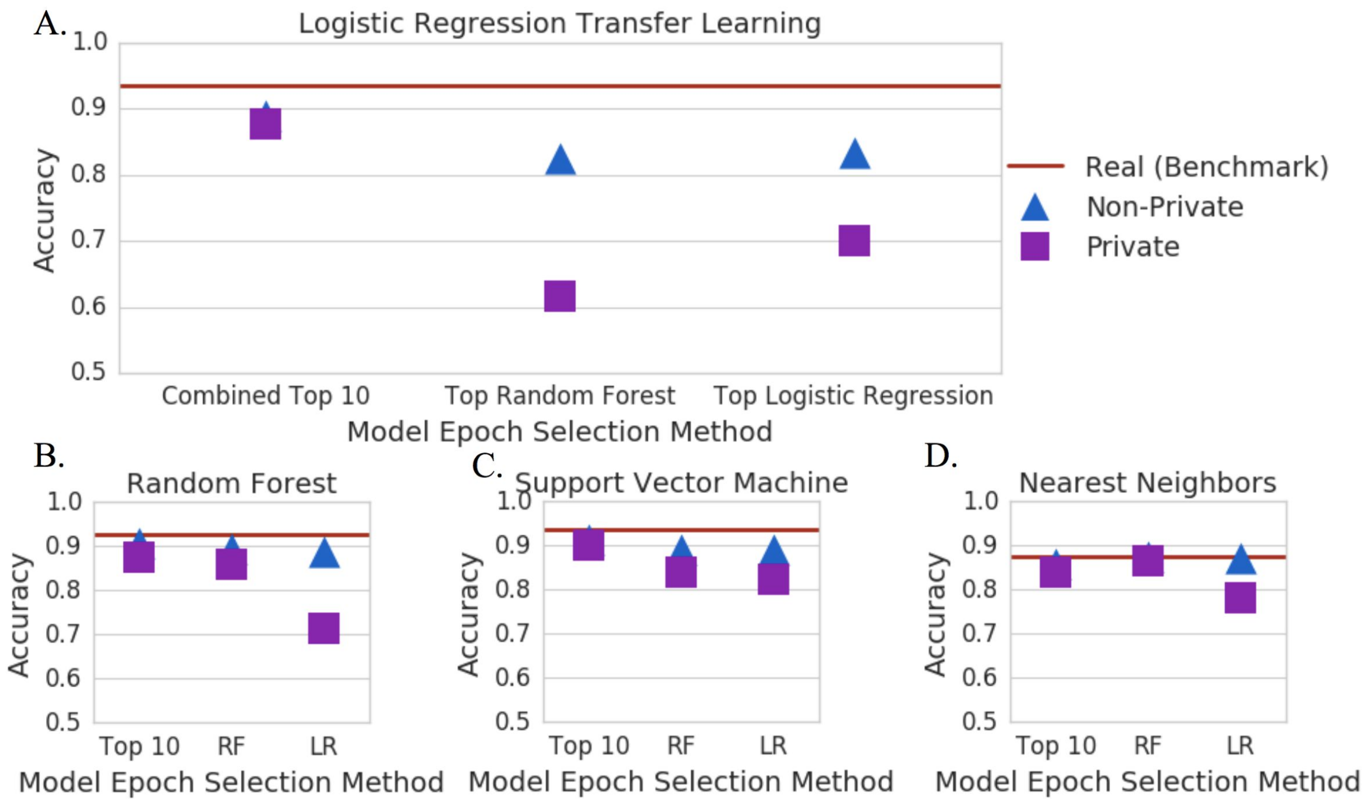
A.



D.



Model Evaluation

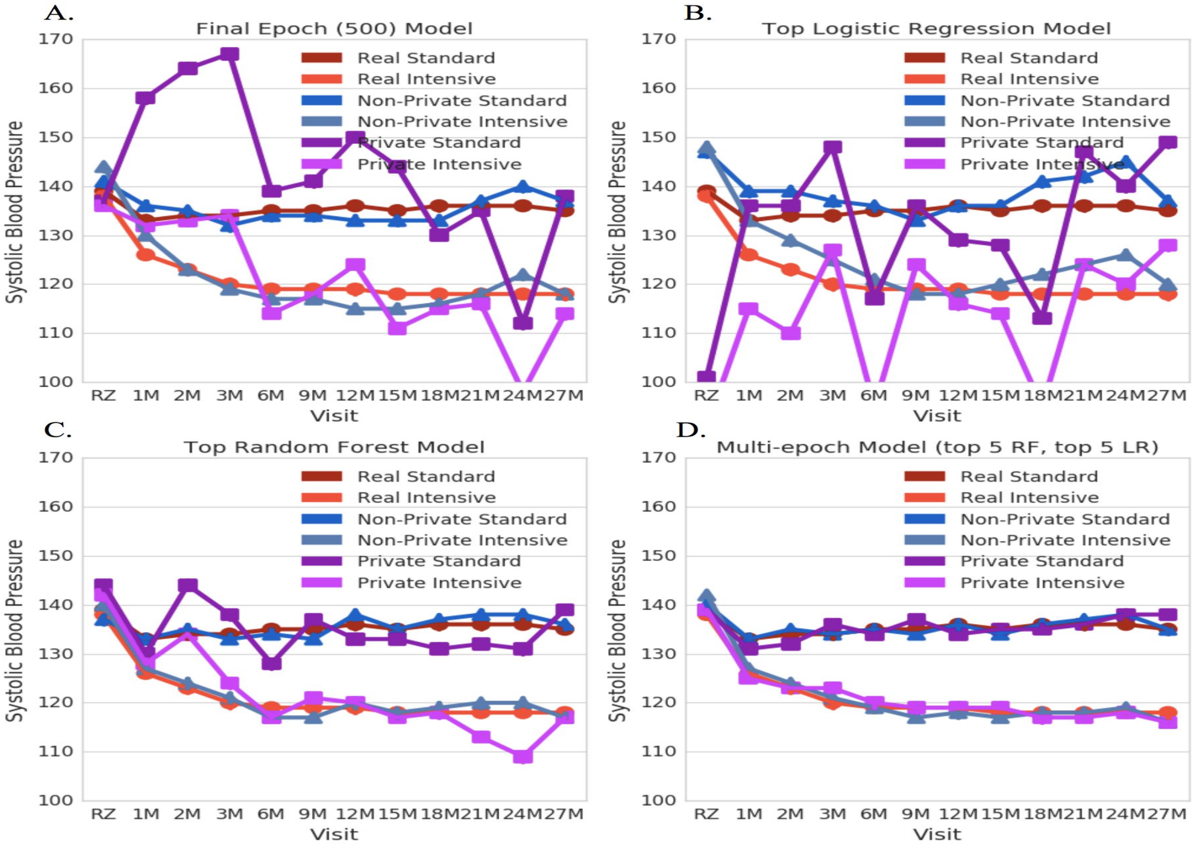


Results

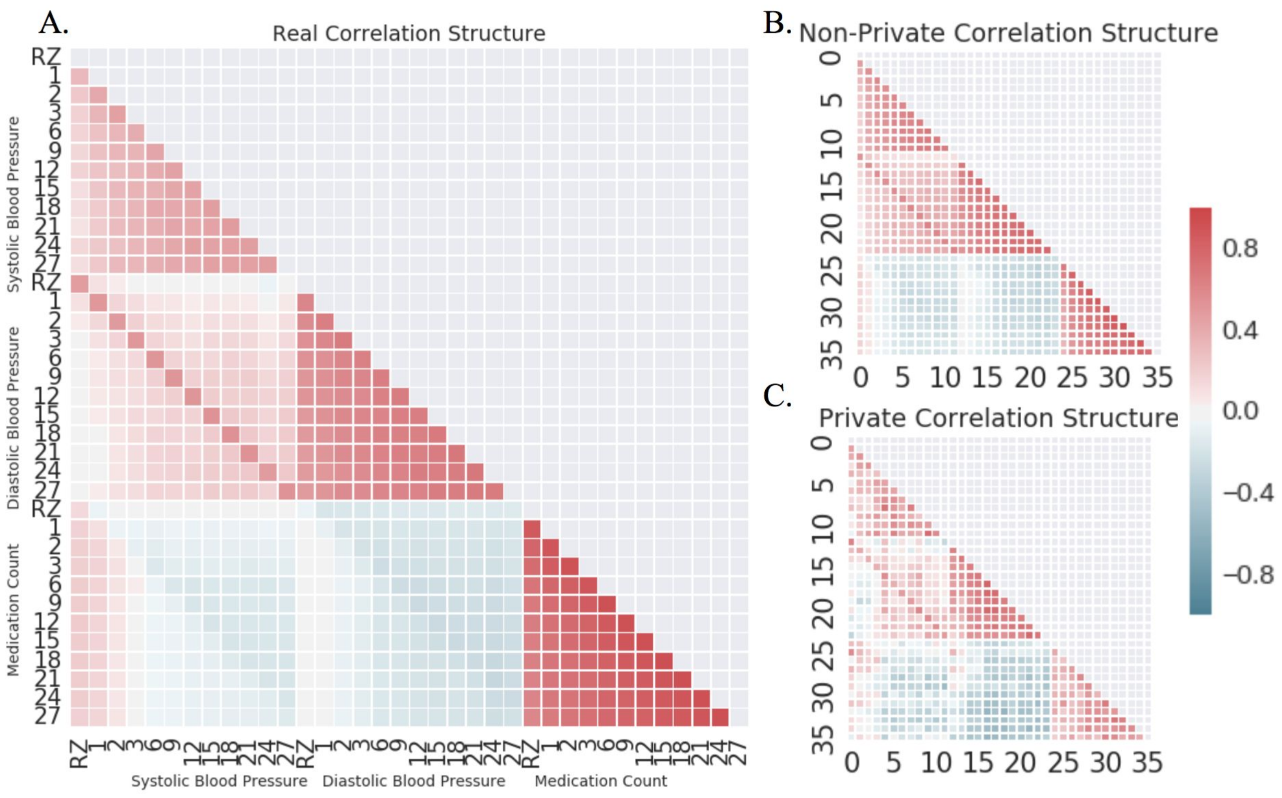
Usefulness Evaluation

- Compare variable distributions between real and simulated data
  - Are simulate variable values consistent with real values?
- Compare correlation structure between variables in real and simulated data
  - Are any relationships between variables maintained?
- Compare machine learning classifiers constructed on real vs. simulated data
  - Can simulated data be used to make classifications on real data?

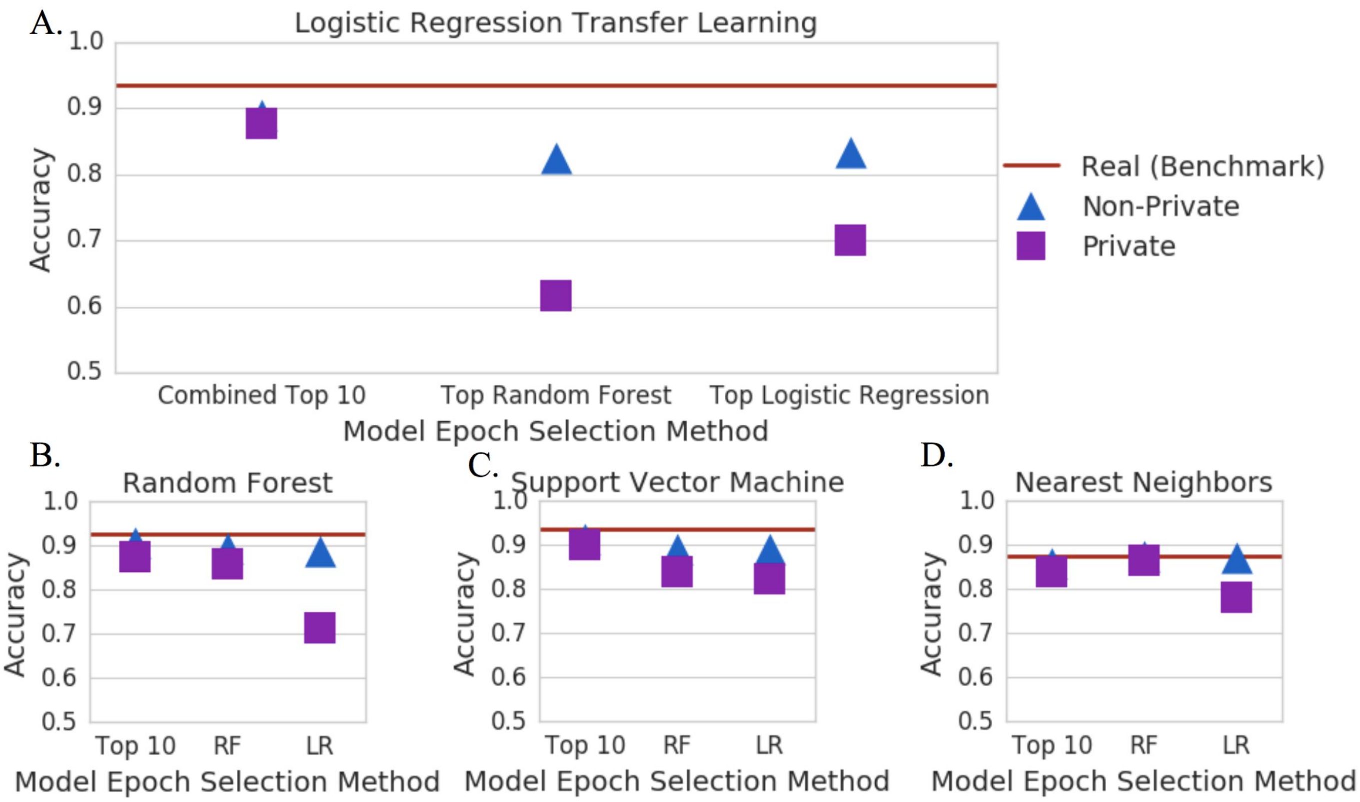
Variable Distribution



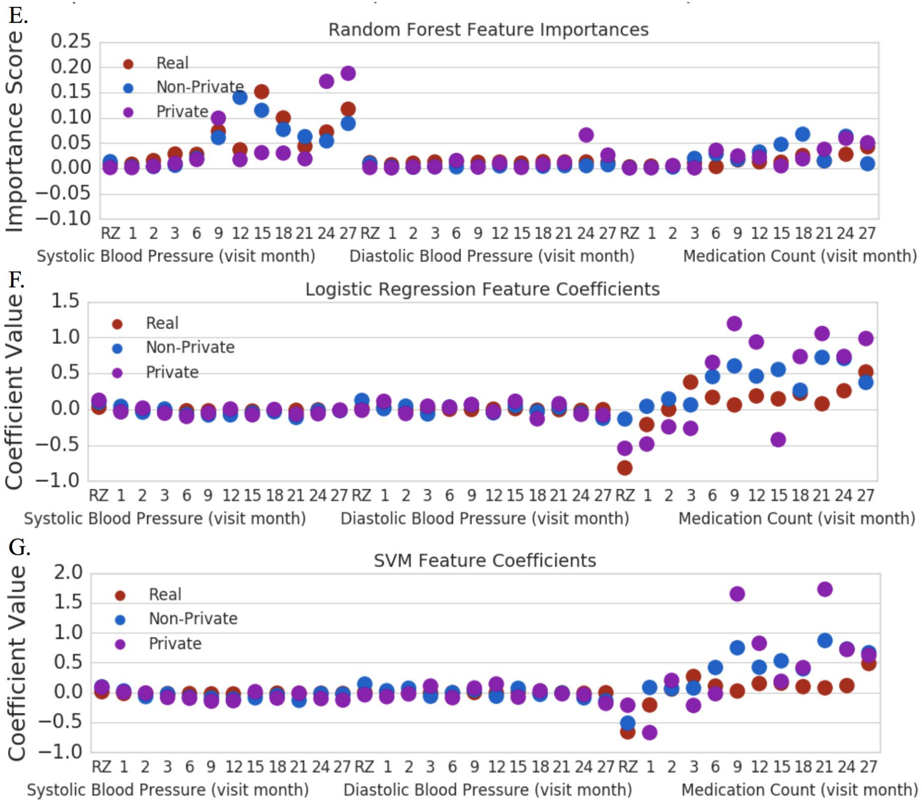
Correlation Between Variables



Transfer Learning Task



Transfer Learning Task (cont'd)



Critique

Technical Issues

Differential Privacy

- Offers plausible deniability
- Losses converge to a noisy equilibrium
- Shrinking gradients may reduce the quality of samples generated

Features

- Discrepancy in the comparison of features
  - Distribution
  - Importance
- Unclear how additional features will affect generated data

General Concerns

Simulated and shareable data may make biomedical analysis easier.

Concerns

1. **Does it actually remove a technology barrier?**
2. **Does it actually reduce privacy risks?**
    - *If so, will patients trust it enough to relinquish rights?*
3. **Does it actually produce useful data?**
    - *If so, will researchers trust the data blindly?*
4. **How well does the model generalize to other datasets?**

Questions? Comments?