Problem Statement

- Build a binary classifier that can detect if an MRI Scan is defaced or not.
  - Can serve as a lightweight deployable tool that researchers can use as a check!
- Build a generative model that can learn to deface as well as reface MRI Scans
  - Showcases that current methods of anonymization are partially reversible to some extent
  - May not provide adequate protection to one's identity

Dataset

- IXI Dataset (http://brain-development.org/ixi-dataset/)
- Around 600 MRI Scans from healthy & normal subjects
- Augmentation Techniques such as:
    - Flips, Rotations, Gaussian Blur, Contrast Normalization, Zoom, etc
- Corresponding Defaced Scan generated
    - With *pydeface* library!
    - Gives us 2 classes → Defaced and Non-Defaced
        - Required for Binary Classification
        - Required for training GANs

Approaches Used

- For Binary Classification
  - Baseline:
    - 6 Logistic Regression [for each 2D scan] Models with Voting amongst them
  - Advanced:
    - 6 Random Forests Classifiers
    - 6 Kenelised SVMs [rbf]
    - 6-branch CNNs that takes the 6 scans parallely and outputs a prediction

Approaches Used

- For Generation of defaced / refaced images
    - Explored CycleGANs and Pix2Pix as this is an Image to Image translation problem
    - Chose Pix2Pix because:
        - Less model complexity
        - Highly correlated dataset - a defaced scan is available for each non-defaced scan
        - Applied only on slice - 2 preprocessing due to computational overheads

Results - I

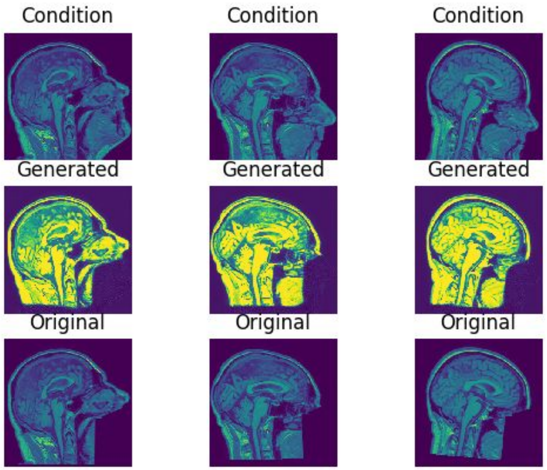
	Logistic Regression	Random Forests	Kernelized [rbf] SVMs	6-branch CNN
Accuracy	0.6367	0.8589	0.8461	0.8846
Precision	0.6290	0.8387	0.8292	0.8591
Recall	0.6667	0.8888	0.8717	0.8923
F1-Score	0.6473	0.8630	0.8499	0.8753

Note: CNNs were trained on 32x32 downsampled images instead of 256x256 images as in other models due to computational overhead. We expect to see greater accuracy if 256x256 images were used.

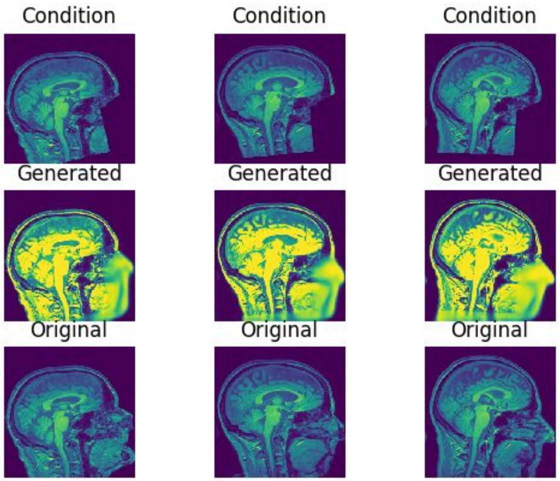
Results - II

For GANs

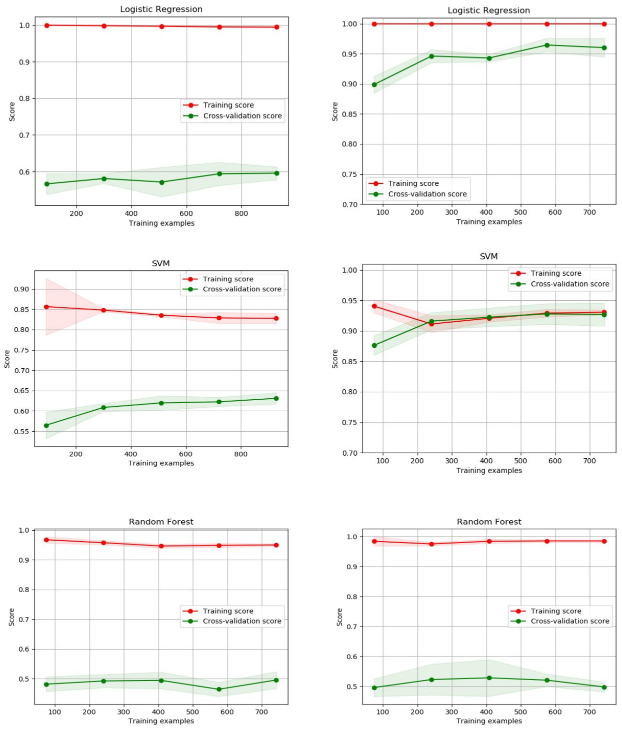
Defacing



Refacing

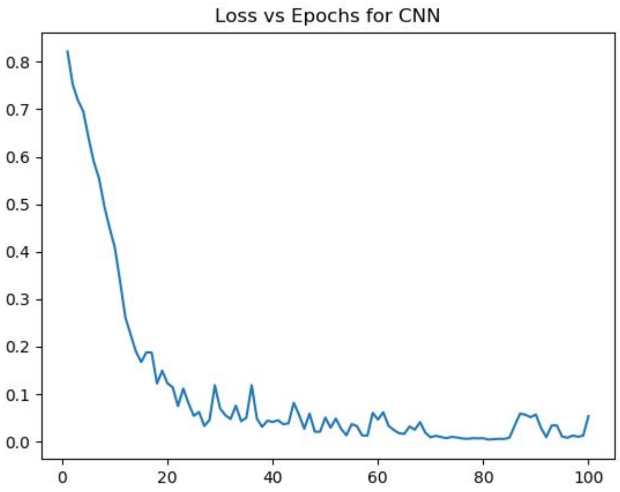


Analysis - I



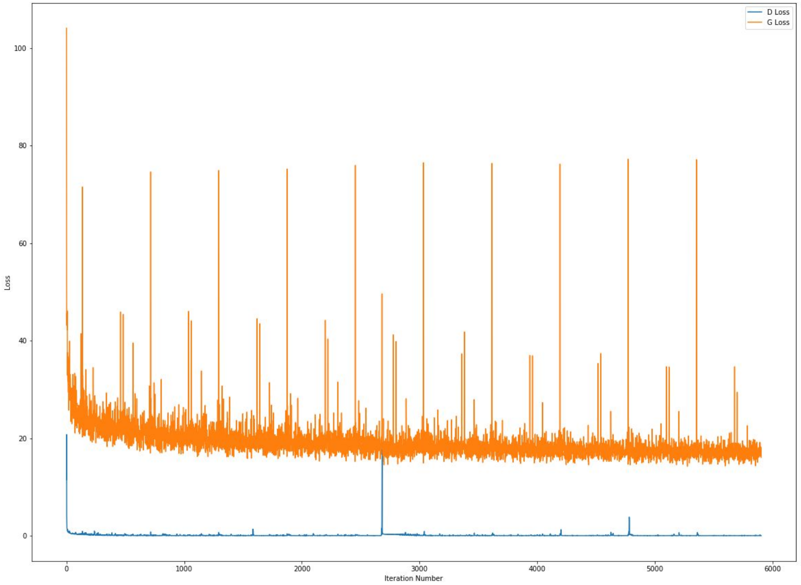
- Models like Linear Regression, SVM and Random Forest are not good for classifying images.
- Because they consider each pixel as independent unit.
- Therefore, any change in pixel values or pixel position would affect each of the models adversely.
- All of these models tend to overfit the data.
- This happens because they are only learning the weights corresponding to the pixel values present in the particular image.
- These learned weights do not generalize well to unseen data samples.

Analysis - II



- We find that CNN performs considerably better than the other three models.
- This happens because CNN architecture allows sharing of weights among different features (pixels).
- Thus CNN achieves Translational and Distortion invariance.
- Therefore, CNN is able to generalize better on the unseen data samples

Analysis - III



- Discriminator loss is 0.005203 after 10 epochs
- Generator loss is 16.0513 after 10 epochs
- Discriminator accuracy is 98%

The generator of the Pix2Pix model is able to learn to reface / deface, but the discriminator is still able to distinguish between the generated data and the original data. We suspect that this is due to the yellowish hue present in the generated images.

Individual Contribution

Madhur - GANs + CNNs + Data Augmentation / Loader

Mudit - GANs + CNNs + Kernelized SVM

Chaitanya - GANs + Random Forests + Logistic Regression