Applying Deep Learning to Public Health: Using Unbalanced Demographic Data to Predict Thyroid Disorder with TensorFlow

UNIVERSITY OF SAN FRANCISCO
School of Nursing and Health Professions

Anaelia Ovalie, John LaGue, Shih-Yin Chen, and Yasser Attiga
Advisor: Patricia Francis-Lyon, Ph.D.

INFLUENCE HEALTH

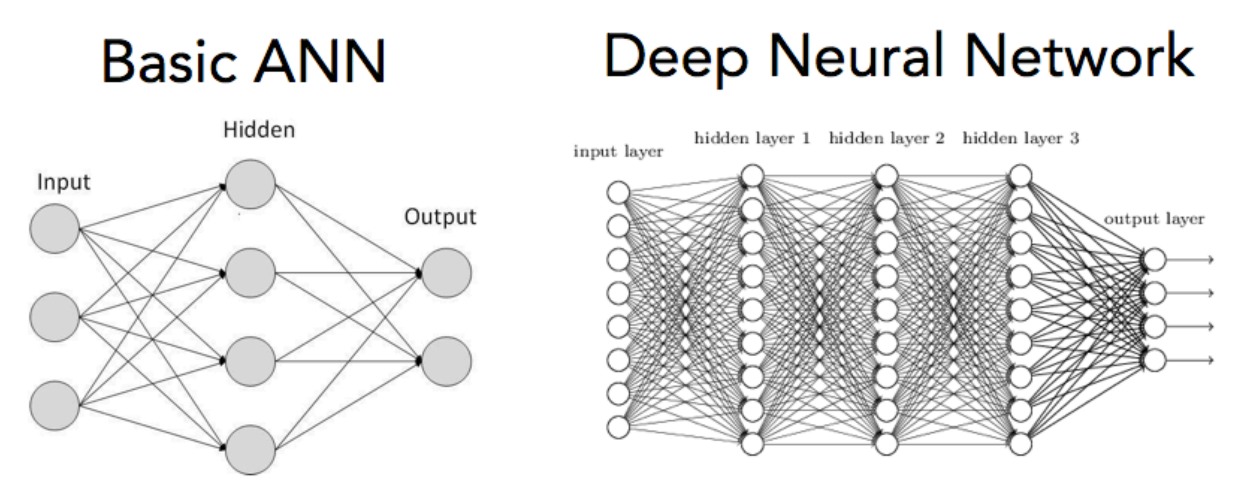
Abstract

- Thyroid disorders affect an estimated 20 million Americans, and up to 60% of them may not realize they have one. To help reach out to at-risk populations, Influence Health developed a model that identifies people who are more likely to have a thyroid disorder.

- **The goal of this project was to create a model using deep learning techniques to see if predictions can be improved over the current Influence Health models.**

- Results: By using TensorFlow's DNN classifier, our group was able to improve on Influence Health's Random Forest models, raising lift scores by 20 – 50 percentage points.

Neural Networks



Artificial Neural Network (ANN)

- Machine learning technique that attempts to 'learn' something from a given data set by mimicking the neurons of a human brain.
- Three main sections of interconnected "neurons:" input, hidden, and output.
- Weights and biases of "neurons" are updated after each epoch.
- Used for tasks such as classifying images as cats or dogs or recognizing handwritten digits on checks.

Google's TensorFlow

- Open source machine learning software library for Python

```
tf.contrib.learn.DNNClassifier.__init__(hidden_units,
feature_columns, model_dir=None, n_classes=2,
weight_column_name=None,optimizer=None, activation_fn=relu,
dropout=None, gradient_clip_norm=None,
enable_centered_bias=False, config=None,
feature_engineering_fn=None, embedding_lr_multipliers=None)
```

- Much of our project involved running numerous models to determine important and optimal parameters
- Activation_fn proved to be the most influential for our project

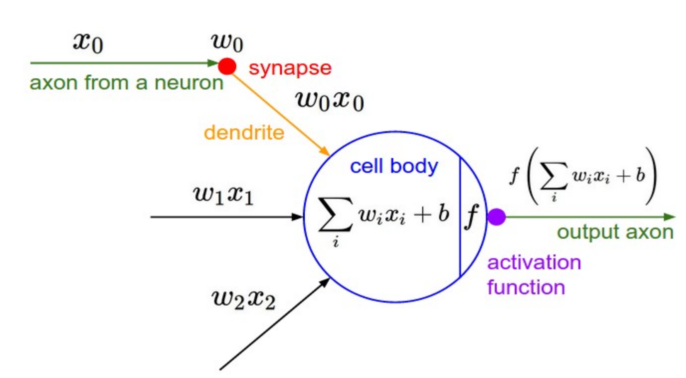
Methods

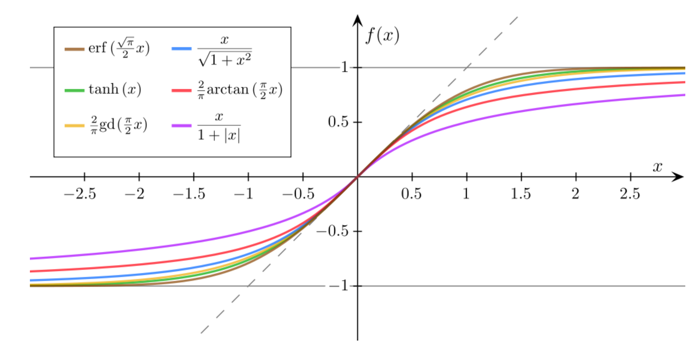
Sampling Methods for Imbalanced Data

- Downsampling:
  - Randomly drop instances from the over-represented class
- Upsampling with Bootstrap:
  - Randomly add copies of instances from the under-represented class
- Upsampling with SMOTE (Synthetic Minority Over-sampling Technique):
  - Creating synthetic samples from the minor class instead of creating copies. Depending upon the amount of over-sampling required, neighbors from the k nearest neighbors are randomly chosen. Selecting two or more similar instances and perturbing an instance one attribute at a time by a random amount within the difference to the neighboring instances.

Activation functions

- Every neuron has two properties: "weight" and "bias" ("threshold")
- The role of the activation function in a neural network is to produce a non-linear decision boundary via non-linear combinations of the weighted inputs





Results

DOWNSAMPLE WITH 0.5 THRESHOLD		
	precision	sensitivity
elu	0.1172	0.7142
relu	0.1141	0.7233
relu6	0.1141	0.7235
sigmoid	0.0788	0.6479
softplus	0.1113	0.7269
softsign	0.1136	0.7300
tanh	0.1181	0.7113

BOOTSTRAP WITH 0.5 THRESHOLD		
	precision	sensitivity
elu	0.1176	0.7134
relu	0.1143	0.7244
relu6	0.1144	0.7234
sigmoid	0.0855	0.6094
softplus	0.1117	0.7246
softsign	0.1153	0.7185
tanh	0.1183	0.7114

SMOTE WITH 0.5 THRESHOLD		
	precision	sensitivity
elu	0.0875	0.8353
relu	0.1124	0.6129
relu6	0.1122	0.6142
sigmoid	NA	NA
softplus	0.0765	0.9387
softsign	0.0833	0.8675
tanh	0.0849	0.8883

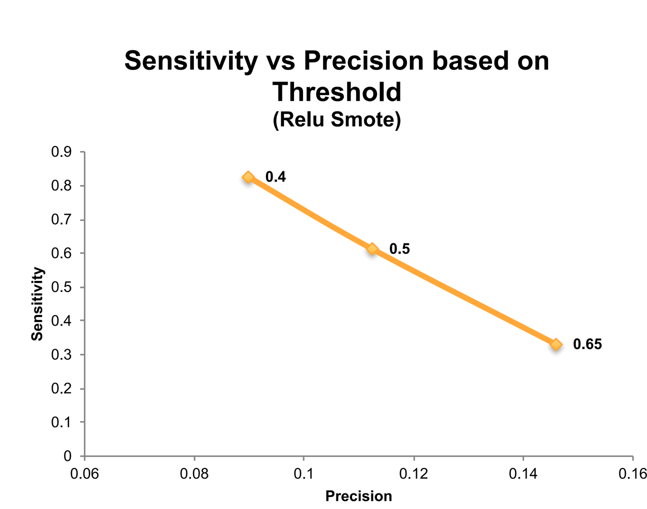
DOWNSAMPLE WITH 0.4 THRESHOLD		
	precision	sensitivity
elu	0.0988	0.8311
relu	0.0971	0.8386
relu6	0.0971	0.8388
sigmoid	0.0610	1.0000
softplus	0.0948	0.8433
softsign	0.0984	0.8304
tanh	0.0987	0.8324

BOOTSTRAP WITH 0.4 THRESHOLD		
	precision	sensitivity
elu	0.0987	0.8313
relu	0.0967	0.8385
relu6	0.0970	0.8380
sigmoid	0.0610	1.0000
softplus	0.0949	0.8422
softsign	0.0987	0.8278
tanh	0.0987	0.8343

SMOTE WITH 0.4 THRESHOLD		
	precision	sensitivity
elu	0.0767	0.9207
relu	0.0898	0.8259
relu6	0.0897	0.8270
sigmoid	NA	NA
softplus	0.0684	0.9819
softsign	0.0761	0.9267
tanh	0.0758	0.9444

Precision and Sensitivity

For each model, as precision improved, sensitivity diminished, and vice versa. This was mostly a function of threshold; when neurons had higher activation thresholds, they were less likely to misclassify someone as positive (i.e. has thyroid disorder). However, that meant they were also predisposed to selecting fewer people, leading to unbalanced confusion matrices.

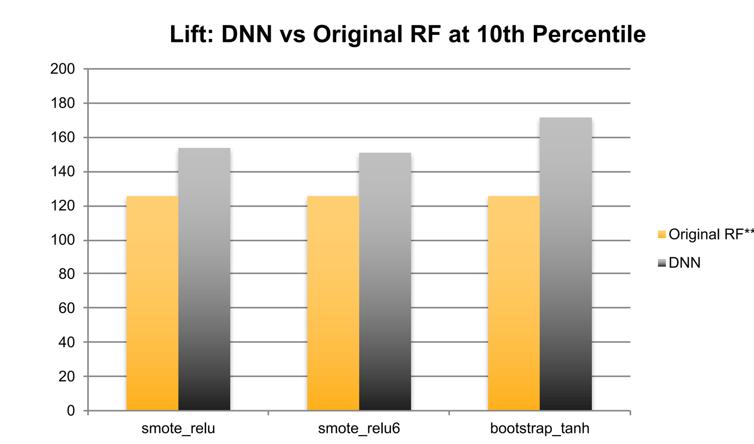


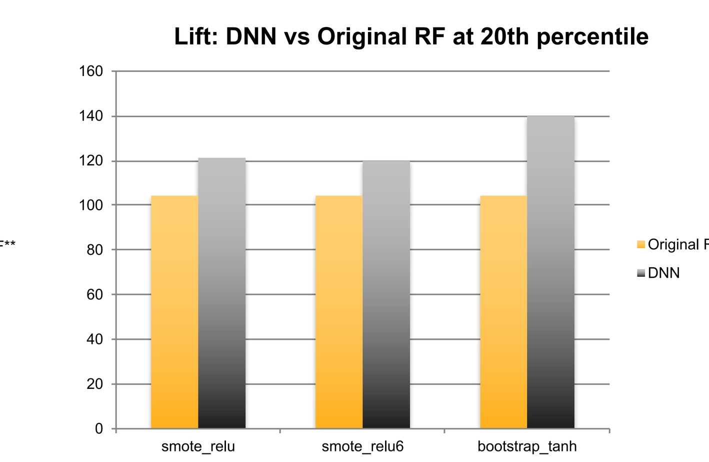
Discussion

The Lift Statistic (borrowed from Influence Health)

Lift measures the performance of the model in predicting target (positive) classes at a certain percentile compared to the average prevalence in the population. It is simply the percent increase in prediction from average response at the top (10th, 20th, and so on) percentiles. For example, suppose a population has an average response rate of 5%, but a certain model has identified a segment with a response rate of 20%. Then that segment would have a lift of 300% ((20-5)/5)*100%.

Using lift as a point of comparison, our models outperform random forest classifiers.





Feature Rankings and Physiological Relevance

Feature Rankings	
age	0.167675
income	0.145367
housemates	0.114853
division	0.110549
education	0.093428
gender	0.090256
payer	0.058479
marital	0.045166
mail	0.039942
POC	0.038789
race	0.035839
donate	0.031434
dwell	0.028226

Common risk factors associated with thyroid disorders vary based on the disease, but they include family history, iodine intake deficiency, age, sex, type 1 diabetes, and radiation to the neck region. Besides age and sex, none of the other risk factors can be extracted from our demographic data. Yet, the DNN model still identified a subset of people with a higher propensity of having a disorder. If DNN classifiers using demographic data can work effectively on thyroid disorder predictions, they could likely work for other diseases as well.

Conclusion

Two main tasks of our analysis:
- Addressing unbalanced data was the key
- Chose parameters to tweak such as activation functions and hidden nodes

For all different models we built, we found out the best model with the highest effectiveness. We evaluated the models base on the precision, recall, and lift. As a result, we found that the demographic data can have some insight on detecting thyroid disorders, and using our deep neural network model, the effectiveness of targeting the potential patients can increase up to 140% with 20th percentile of the population, comparing to random selection.

In the future, if we have more medical data, we can try to find out the relationships among different diseases, and will potentially be able to build more knowledge-driven models.