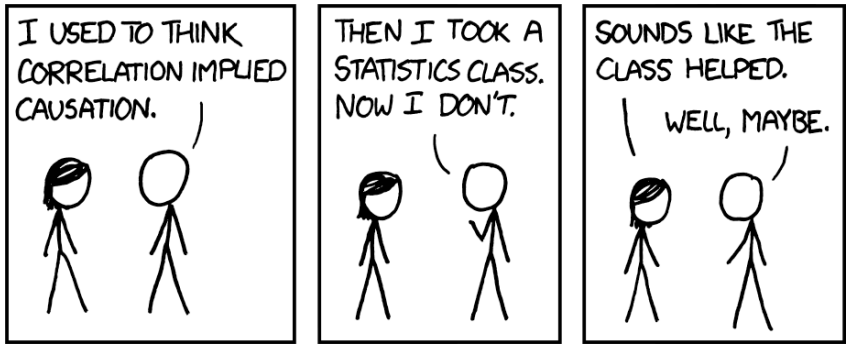
CLEMSON
PUBLIC HEALTH SCIENCES

# Tabular Data 3 : Statistics for Machine Learning Using the National Inpatient Sample Dataset

**MinJae Woo**

Assistant Professor of Health Informatics
Clemson University



## Alternative Title for Today's Session:

# The Encyclopedia of Common Statistical Mistakes

PART 1: COMMON STATISTICAL MISTAKES DURING PREPROCESSING

- ✓ Missing Data and Imputation
- ✓ Variable Selection
- ✓ Imbalanced Dataset

PART 2: COMMON STATISTICAL MISTAKES DURING MODEL TRAINING

- ✓ Benchmark and Selection of Predictive Models

PART 3: COMMON STATISTICAL MISTAKES DURING MODEL EVALUATION

- ✓ Selection of Performance Metrics
- ✓ Imbalanced Dataset
- ✓ Absence of Failure Analysis
- ✓ Too-good-to-be-true Performance

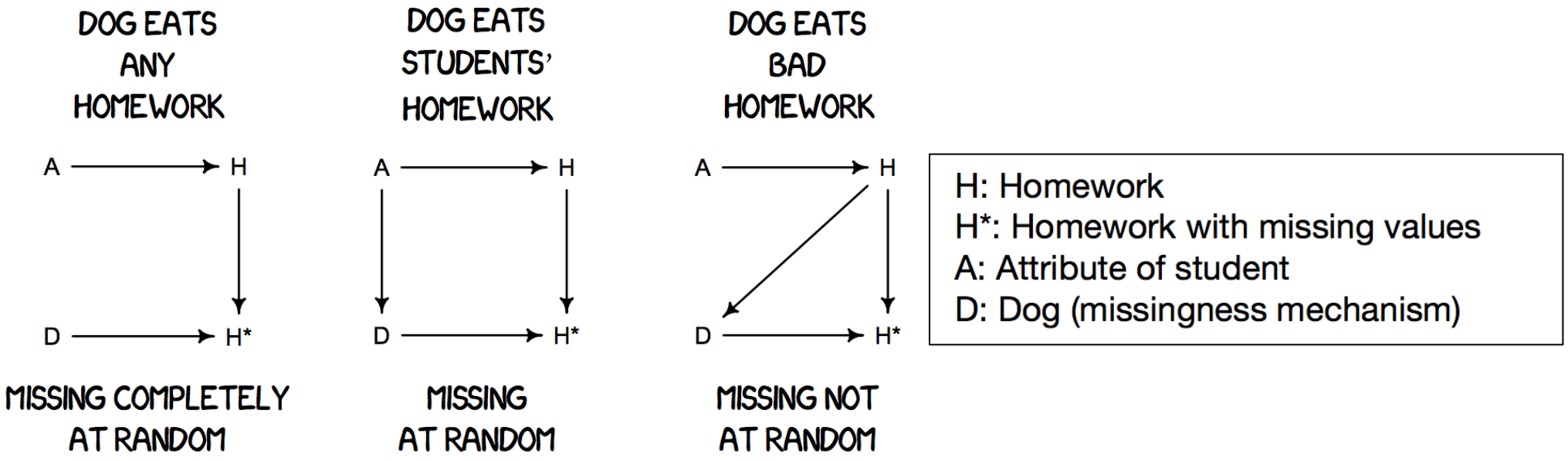
PART 4: COMMON STATISTICAL MISTAKES DURING DISPARITY ANALTYSIS

- ✓ Univariate Approach
- ✓ Our Approach with EMBED and NIS

# PART 1

## Common Statistical Mistakes During Preprocessing

# Missing Data - Introduction




1. From a statistical point of view, the most important consideration for missing data is the type of missingness pattern.
2. There are two missingness pattern types:
    - Missing Completely at Random (**MCAR**): The missing data points occur entirely at random
    - Missing Not at Random (MNAR): The value of the variable that's missing is related to the reason it's missing (e.g., depressed patients are not likely to answer a questionnaire "Are you depressed?")

Image Source: https://x.com/rlmcelreath

# Statistical Imputation Using Regression

**Ideally:**

| Patient | Glucose Level | Patient Age | Sugar Intake | Carb Intake |
|---|---|---|---|---|
| Patient 1 | 85 | 43 | 55 | 88 |
| Patient 2 | 110 | 35 | 36 | 153 |
| Patient 3 | 97 | 67 | missing | 42 |
| Patient 4 | 115 | 52 | 88 | 120 |
| Patient 5 | 81 | 23 | 34 | 27 |
| Patient 6 | 90 | 32 | 57 | 58 |

**In Reality:**

| Patient | Glucose Level | Patient Age | Sugar Intake | Carb Intake |
|---|---|---|---|---|
| Patient 1 | 85 | 43 | 55 | 88 |
| Patient 2 | 110 | 35 | 36 | 153 |
| Patient 3 | 97 | 67 | missing | 42 |
| Patient 4 | 115 | missing | missing | 120 |
| Patient 5 | 81 | 23 | 34 | 27 |
| Patient 6 | 90 | 32 | 57 | 58 |

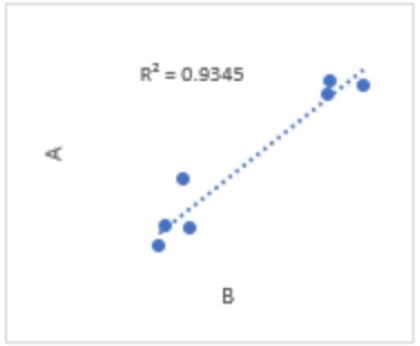
1. (Left) In order to use regression imputation:
    - Set Sugar Intake as the dependent variable and all other variables as independent variables
    - Then, use data from patients 1, 2, 4, 5, and 6 to build a regression model.
    - Apply the model and input the values of GL, PA, and CI to calculate Sugar Intake for Patient 3.
2. (Right) What would be the procedure for using regression imputation for this particular missing pattern?

CLEMSON PUBLIC HEALTH SCIENCES

# Missing Data - Introduction

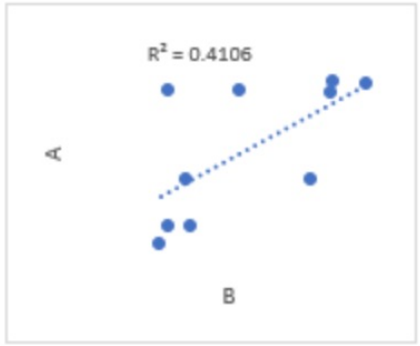
Missing data is in red. There is a strong correlation between A and B, so let's try to impute A using B and C.

| A | B | C |
|---|---|---|
| 0.93 | 1.40 | 1.53 |
| 0.24 | 0.46 | 0.76 |
| | 0.80 | |
| 0.95 | 1.24 | 1.46 |
| 0.23 | 0.57 | |
| 0.90 | | 1.28 |
| 0.15 | 0.42 | |
| 0.47 | 0.54 | 0.63 |
| | 1.14 | |
| 0.89 | 1.23 | 1.45 |



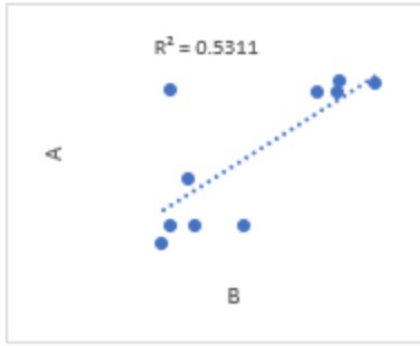

Missing data is filled in randomly. This dillutes the correlations, but allows us to impute using all available data.

| A | B | C |
|---|---|---|
| 0.93 | 1.40 | 1.53 |
| 0.24 | 0.46 | 0.76 |
| 0.90 | 0.80 | 1.53 |
| 0.95 | 1.24 | 1.46 |
| 0.23 | 0.57 | 1.28 |
| 0.90 | 0.46 | 1.28 |
| 0.15 | 0.42 | 1.53 |
| 0.47 | 0.54 | 0.63 |
| 0.47 | 1.14 | 1.28 |
| 0.89 | 1.23 | 1.45 |



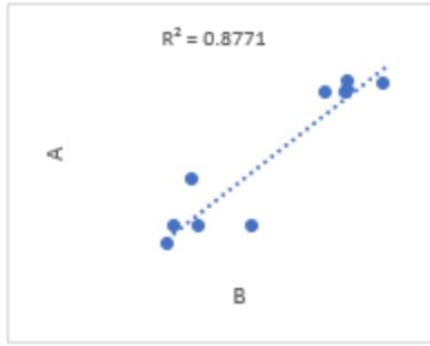

A random forest is used to predict A with B and C. Notice the correlation between A and B improved.

| A | B | C |
|---|---|---|
| 0.93 | 1.40 | 1.53 |
| 0.24 | 0.46 | 0.76 |
| 0.24 | 0.80 | 1.53 |
| 0.95 | 1.24 | 1.46 |
| 0.23 | 0.57 | 1.28 |
| 0.90 | 0.46 | 1.28 |
| 0.15 | 0.42 | 1.53 |
| 0.47 | 0.54 | 0.63 |
| 0.89 | 1.14 | 1.28 |
| 0.89 | 1.23 | 1.45 |




After Imputing B using A and C, we have achieved a correlation between A and B much closer to the original data.

| A | B | C |
|---|---|---|
| 0.93 | 1.40 | 1.53 |
| 0.24 | 0.46 | 0.76 |
| 0.24 | 0.80 | 1.53 |
| 0.95 | 1.24 | 1.46 |
| 0.23 | 0.57 | 1.28 |
| 0.90 | 1.24 | 1.28 |
| 0.15 | 0.42 | 1.53 |
| 0.47 | 0.54 | 0.63 |
| 0.89 | 1.14 | 1.28 |
| 0.89 | 1.23 | 1.45 |




Source: cran.r-project.org

CLEMSON PUBLIC HEALTH SCIENCES

# Missing Data - Mistakes

| Patient | Glucose Level | Patient Age | Sugar Intake | Carb Intake |
|---|---|---|---|---|
| Patient 1 | 85 | 43 | 55 | 88 |
| Patient 2 | 110 | 35 | 36 | 153 |
| Patient 3 | Missing | 67 | 67 | 42 |
| Patient 4 | 115 | 52 | 88 | Missing |
| Patient 5 | 90 | 23 | 34 | Missing |
| Patient 6 | Missing | Missing | 63 | Missing |
| Patient 7 | 90 | Missing | Missing | Missing |
| Average Value | 98 | 36.7 | 57.2 | 94.3 |

| Patient | Glucose Level | Patient Age | Sugar Intake | Carb Intake |
|---|---|---|---|---|
| Patient 1 | 85 | 43 | 55 | 88 |
| Patient 2 | 110 | 35 | 36 | 153 |
| Patient 3 | **98** | 67 | 67 | 42 |
| Patient 4 | 115 | 52 | 88 | **94.3** |
| Patient 5 | 90 | 23 | 34 | **94.3** |
| Patient 6 | **98** | **36.7** | 63 | **94.3** |
| Patient 7 | 90 | **36.7** | **57.2** | **94.3** |

1. Performing imputation can severely distort the distribution of this variable, leading to an underestimation of the standard deviation, which requires caution in the context of descriptive analytics.

CLEMSON PUBLIC HEALTH SCIENCES

# Variable Selection - Introduction

| Patient | Glucose Level | Patient Age | Sugar Intake | Carb Intake |
|---|---|---|---|---|
| Patient 1 | 85 | 43 | 55 | 88 |
| Patient 2 | 110 | 35 | 36 | 153 |
| Patient 3 | 93 | 67 | 67 | 42 |
| Patient 4 | 115 | 52 | 88 | 103 |
| Patient 5 | 90 | 23 | 34 | 94 |
| Patient 6 | 97 | 36 | 63 | 69 |
| Patient 7 | 90 | 44 | 57 | 132 |

Clustering (K=2)

Group 1: Patient 1, Patient 5, Patient 6

Group 2: Patient 2, Patient 3, Patient 4, Patient 7

| Variable | Patient 1 | Patient 2 | Patient 3 | Patient 4 | Patient 5 | Patient 6 | Patient 7 |
|---|---|---|---|---|---|---|---|
| Glucose Level | 85 | 110 | 93 | 115 | 90 | 97 | 90 |
| Patient Age | 43 | 35 | 67 | 52 | 23 | 36 | 44 |
| Sugar Intake | 55 | 36 | 67 | 88 | 34 | 63 | 57 |
| Carb Intake | 88 | 153 | 42 | 103 | 94 | 69 | 132 |

Clustering (K=2)

Group 1: Glucose Level, Sugar Intake, Carb Intake

Group 2: Patient Age

# Variable Selection - Mistakes

| Features | ICD-9-CM | Importance | Early detection model excluded |
|---|---|---|---|
| Normal delivery | 650 | 0.1758 | Yes |
| Other obstetric trauma | 665 | 0.0585 | Yes |
| Spontaneous abortion (miscarriage) | 634 | 0.0139 | Yes |
| Other problems associated with amniotic cavity and membranes | 658 | 0.0124 | Yes |
| Antepartum hemorrhage, abruptio placentae | 641.2 | 0.0106 | Yes |

1. Please, please, please be sure to consult with someone with domain expertise after running the automated variable selection such as K best selection or unsupervised variable clustering.
2. Be careful with ICD codes and be aware that those are designed primarily for billing purposes.
3. The most common suggestion by many IEEE/ACM reviewers is to use embeddings for these codes (instead of one hot coding), but we believe it may harm the transparency and expandability of predictive models.

# Imbalanced Dataset - Introduction

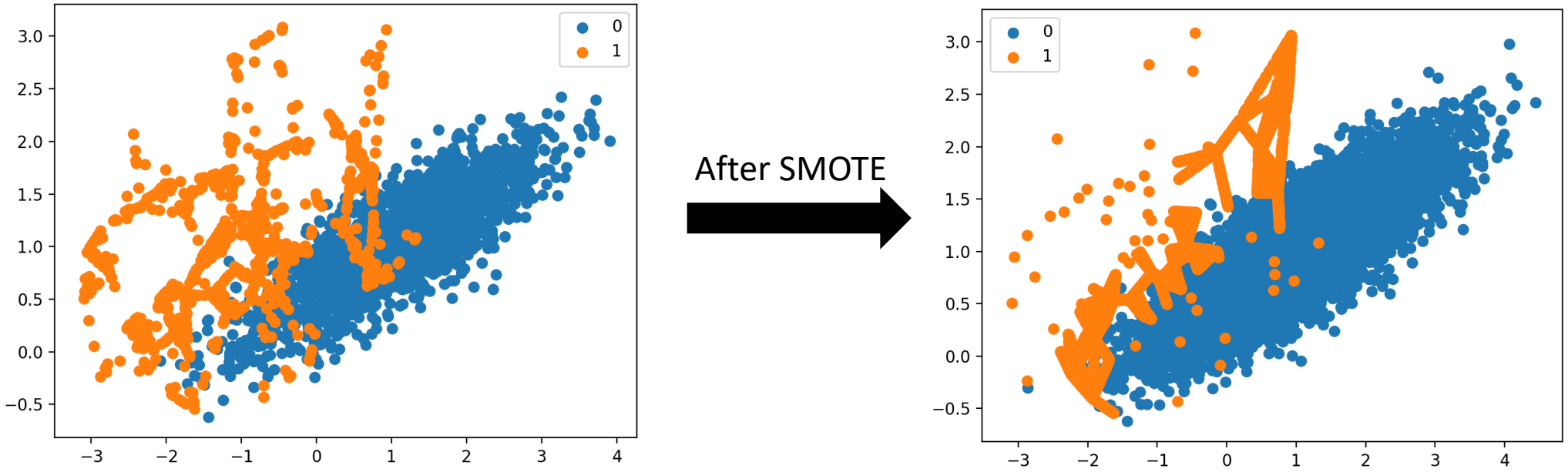
**Table 1. Distribution of patient records used for model development**

| | Count (Percentage) | | | |
|---|---|---|---|---|
| **Characteristics** | **Total** | **Training set** | **Validation set** | **Test set** |
| All records | 6,561,385 (100.00%) | 3,936,831 (100.00%) | 1,312,277 (100.00%) | 1,312,277 (100.00%) |
| **Label** | | | | |
| PPH | 179,210 (2.73%) | 107,370 (2.73%) | 36,055 (2.75%) | 35,785 (2.73%) |
| Non-PPH | 6,382,175 (97.27%) | 3,829,461 (97.27%) | 1,276,222 (97.25%) | 1,276,492 (97.27%) |

1. If one label is more common than the other, the dataset is considered imbalanced.
2. Although there is no concrete definition, a minority class smaller than 5% is typically considered imbalanced from statistical point of view.
    - However, you will start experiencing challenges when the minority class drops below 10%.
3. Remember, you are dealing with a completely different situation here.

CLEMSON PUBLIC HEALTH SCIENCES

# Under- and Over-sampling for Imbalanced Dataset



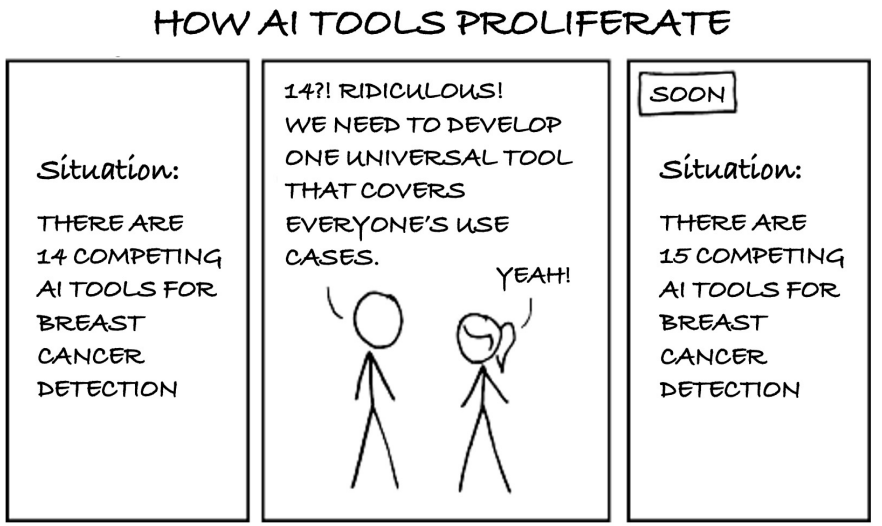

1. Synthetic Minority Oversampling Technique (SMOTE) has been the most popular off-the-shelve option in many studies.
2. Beware that the use of SMOTE can lead to amplification of existing biases.

**Algorithm 5.1** Balance-SMOTE.

**balanceSMOTE**$(TC)$
initialize $Risk \leftarrow \emptyset$
for each $x \in T$ where $y(x) = 1$ do
$Risk(x) \leftarrow 1$
$S(x) \leftarrow \emptyset$
for each $c \in C$ do
$r_c(x) \leftarrow$ **assignRisk**$(x,$ c$)$ to
$Risk(x) = Risk(x) \times r_c(x)$
$S(x) \leftarrow S(x) \cup$ **assignSubgroup**$(x, c)$
end for
return $Risk(x), S(x)$
initialize $B_{pos} \leftarrow \emptyset, B_{neg} \leftarrow \emptyset$
for each $s \in S$ do
$D_s^1 \leftarrow \{x \in T | S(x) = s \; and \; y(x) = 1\}$
$n_s \leftarrow$ **countRow**$(s)$
$w_s \leftarrow Risk(s)$
$N_s = n_s \times w_s$
$B_{pos} \leftarrow B_{pos} \cup$ **oversample**$(D_s, w_s)$
$B_{neg} \leftarrow B_{neg} \cup$ **undersample**$(D_s, N_s)$
end for
return $T_{BSMOTE} \leftarrow B_{pos} \cup B_{neg}$

# PART 2

# Common Statistical Mistakes During Model Training



# Selection of Predictive Model - Introduction

| Classifier | Dataset | AUC | Accuracy | Precision | Recall | F1 Score | Disparity Score |
|---|---|---|---|---|---|---|---|
| **Baseline Logistic Regression** | Test | 0.623 | 0.725 | 0.044 | 0.442 | 0.080 | |
| **Baseline Naïve Bayes** | Test | 0.616 | 0.923 | 0.077 | 0.167 | 0.105 | |
| **Baseline XGBoost** | Test | 0.624 | 0.763 | 0.047 | 0.400 | 0.083 | 4.756 |
| **Baseline XGBoost** | Holdout | 0.628 | 0.763 | 0.047 | 0.406 | 0.085 | 4.564 |
| **Enhanced XGBoost** | Test | 0.591 | 0.728 | 0.039 | 0.382 | 0.071 | 2.875 |
| **Enhanced XGBoost** | Holdout | 0.591 | 0.726 | 0.038 | 0.377 | 0.069 | 3.011 |

1. The current state-of-the-art predictive models fall into two categories: models using boosting and models using neural networks.
2. Models using boosting:
   - The most popular one is XGBoost. Models using neural networks are generally trickier to train or fine-tune.
3. Models using neural networks:
   - The most popular is TabTransformer, suggested by NVIDIA, although some people prefer TabNet, suggested by Google.
4. (Highly-biased personal opinion) It seems that XGBoost outperforms TabNet in many use cases. This may be related to the fact that NN-based models are harder to optimize than boosting-based models, making it appear that XGBoost outperforms TabNet.

# Selection of Predictive Model - Mistake

| Classifier | Dataset | AUC | Accuracy | Precision | Recall | F1 Score | Disparity Score |
|---|---|---|---|---|---|---|---|
| **Baseline Logistic Regression** | Test | 0.623 | 0.725 | 0.044 | 0.442 | 0.080 | |
| **Baseline Naïve Bayes** | Test | 0.616 | 0.923 | 0.077 | 0.167 | 0.105 | |
| **Baseline XGBoost** | Test | 0.624 | 0.763 | 0.047 | 0.400 | 0.083 | 4.756 |
| **Baseline XGBoost** | Holdout | 0.628 | 0.763 | 0.047 | 0.406 | 0.085 | 4.564 |
| **Enhanced XGBoost** | Test | 0.591 | 0.728 | 0.039 | 0.382 | 0.071 | 2.875 |
| **Enhanced XGBoost** | Holdout | 0.591 | 0.726 | 0.038 | 0.377 | 0.069 | 3.011 |

1. The most common mistake in predictive modeling using machine learning is to use machine learning.

2. Carefully crafted traditional models (e.g., regression and simple decision trees) can offer comparable performance to state-of-the-art machine learning models in many cases, while facilitating explanability and easier maintenance. This is especially true for many tabular datasets that are not overly complex.

3. Most common mistake arise from lack of domain expertise; again, make sure to consult with your domain expert
   - "We developed an AI screening model to detect asymptomatic COVID-19 patients using CT images."

# Selection of Predictive Model - Mistake

| Classifier | Dataset | AUC | Accuracy | Precision | Recall | F1 Score | Disparity Score |
|---|---|---|---|---|---|---|---|
| **Baseline Logistic Regression** | Test | 0.623 | 0.725 | 0.044 | 0.442 | 0.080 | |
| **Baseline Naïve Bayes** | Test | 0.616 | 0.923 | 0.077 | 0.167 | 0.105 | |
| **Baseline XGBoost** | Test | 0.624 | 0.763 | 0.047 | 0.400 | 0.083 | 4.756 |
| **Baseline XGBoost** | Holdout | 0.628 | 0.763 | 0.047 | 0.406 | 0.085 | 4.564 |
| **Enhanced XGBoost** | Test | 0.591 | 0.728 | 0.039 | 0.382 | 0.071 | 2.875 |
| **Enhanced XGBoost** | Holdout | 0.591 | 0.726 | 0.038 | 0.377 | 0.069 | 3.011 |

1. If machine learning models were to be used, you will need a benchmark to justify your selection.

2. Be cautious with comparing boosting-based models and NN-based models. The results can very easily be cherry-picked, such as adopting favorable preprocessing process for a specific type of model. As a reviewer, I always try to look for evidence that the authors were aware and mindful of the nature of comparison.

# PART 3

## Common Statistical Mistakes During Model Evaluation

# Selection of Performance Metrics

| Model | Acc | RMSE | TPR | FPR | Prec | Rec | F1-score | AUC | InfoS |
|---|---|---|---|---|---|---|---|---|---|
| ANN | 71.7 | 0.4534 | 0.44 | 0.16 | 0.53 | 0.44 | 0.48 | 0.7 | 48.11 |
| Boost | 75.5 | 0.4324 | 0.27 | 0.04 | 0.74 | 0.27 | 0.4 | 0.59 | 34.28 |
| KNN | 72.4 | 0.5101 | 0.32 | 0.1 | 0.56 | 0.32 | 0.41 | 0.63 | 43.37 |
| Bagg | 71 | 0.4494 | 0.37 | 0.14 | 0.52 | 0.37 | 0.43 | 0.6 | 22.34 |
| SVM | 67.8 | 0.4518 | 0.17 | 0.1 | 0.4 | 0.17 | 0.23 | 0.63 | 11.3 |

1. The above example features different models and performance metrics applied to the same dataset.
2. For classification, the most common metrics are accuracy, F1-score, and AUC.
3. Some metrics have a trade-off relationship with others (e.g., precision and recall).

# Selection of Performance Metrics

| Model | Acc | RMSE | TPR | FPR | Prec | Rec | F1-score | AUC | InfoS |
|---|---|---|---|---|---|---|---|---|---|
| ANN | 3 | 2 | 1 | 1 | 3 | 1 | 1 | 1 | 1 |
| Boost | 1 | 5 | 4 | 4 | 1 | 4 | 4 | 4 | 3 |
| KNN | 2 | 1 | 3 | 3 | 2 | 3 | 3 | 2 | 2 |
| Bagg | 4 | 4 | 2 | 2 | 4 | 2 | 2 | 3 | 4 |
| SVM | 5 | 3 | 5 | 3 | 5 | 5 | 5 | 2 | 5 |

1. Depending on which performance metric is chosen, the best model can change.
2. Some papers introduce their own performance metrics along with their proposed models, which requires caution when interpreting the results.
3. Use caution when interpreting confidence intervals for performance metrics.

# Imbalanced Dataset

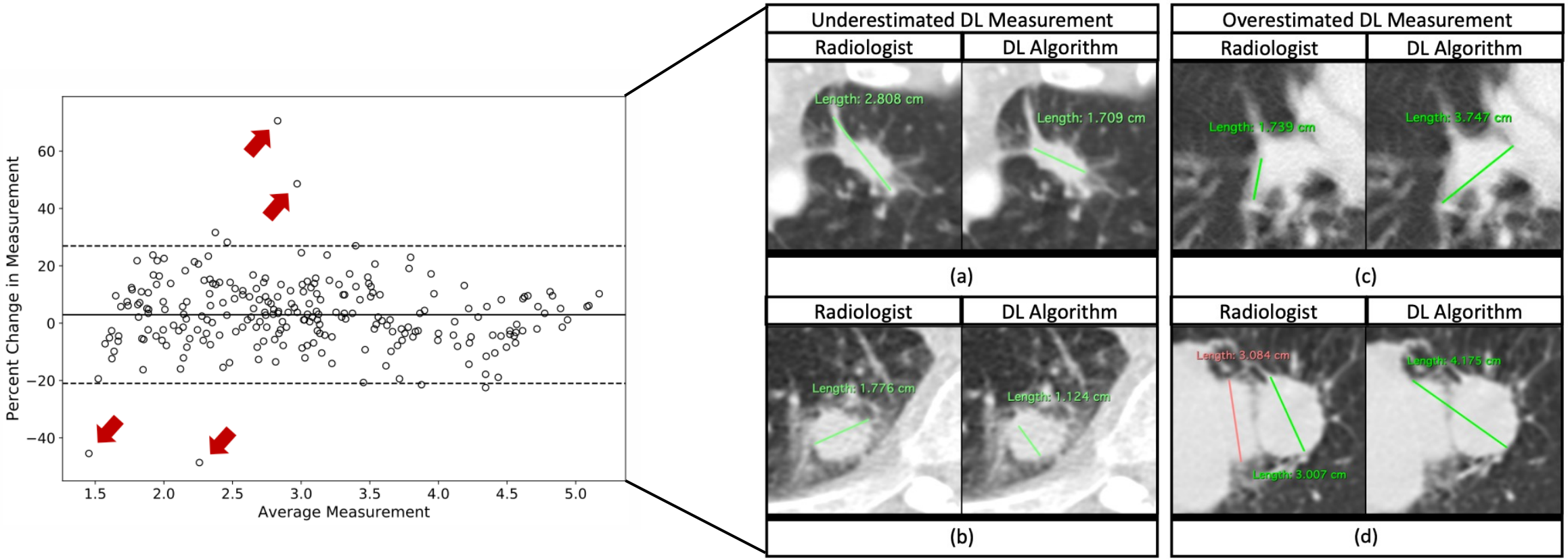
| | | Actual | |
|---|---|---|---|
| | | Pos | Neg |
| Predicted | Pos | 990 | 10 |
| | Neg | 0 | 0 |

Accuracy: 0.99

1. The most common mistake during model evaluation is not considering the imbalanced distribution of positives and negatives, when applicable.
2. Especially in highly imbalanced datasets, your accuracy, AUC, and F1 score can be misleadingly high even if the model does not perform at all on the minority class.
3. A lot of models involving imbalanced datasets are screening models, where it is generally appropriate to prioritize recall (also known as sensitivity).

CLEMSON PUBLIC HEALTH SCIENCES

# Absence of Failure (Case) Analysis




1. There should be either a failure case study (presented above) or a failure analysis (to identify the cause of failure), followed by a performance evaluation.

CLEMSON PUBLIC HEALTH SCIENCES

# High (Maybe Too High) Performance

An actual example of an assignment submitted by my doctoral student

| Models | Precision | Recall | F1 | Accuracy | AUC for ROC |
|---|---|---|---|---|---|
| **XG Boosted Trees** | **0.98** | **0.97** | **0.98** | **0.98** | **0.99** |
| **TabNet (DNN)** | **0.97** | **0.96** | **0.96** | **0.97** | **0.99** |
| ANN (5-Layer) | 0.97 | 0.96 | 0.96 | 0.96 | 0.99 |
| Random Forest Trees | 0.95 | 0.91 | 0.93 | 0.93 | 0.98 |
| Logistic Regres-sion | 0.81 | 0.80 | 0.80 | 0.82 | 0.92 |

TABLE 1: Metrics for comparison of models

**6.2. Evaluation of Performance and Explainability**

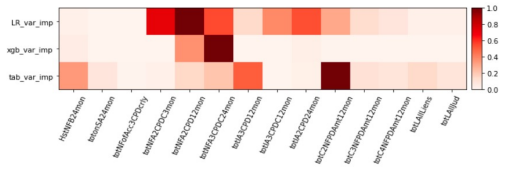



Figure 4: Global Feature importance color map for Logistic Regression, XG Boosted Tress, and TabNet

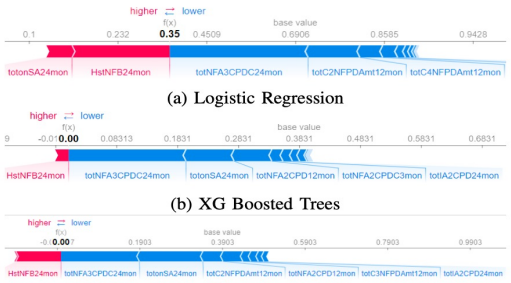



Figure 5: Local Feature importance for an *MPID* whose *worst_status* is 0


1. Having predictors that are supposed to be the dependent variable in a predictive model happens more frequently than you might think (e.g., using patient charges as a predictor in a model predicting length of stay).
2. Remember that an AUC of 0.99 is like an error message indicating one of the following:
   (a) Reverse Causality (b) Data Leakage (c) An overly simple task with a trivial solution or (d) You deserve a Nobel Prize.

# PART 4

# Common Statistical Mistakes During Disparity Analysis

# Univariate Approach

Research shows that pulse oximeters may provide inaccurate readings for Black patients and people with darker skin tones.

Will they also be inaccurate for (1) white patients with darker skin tones and (2) Black patients with lighter skin tones?

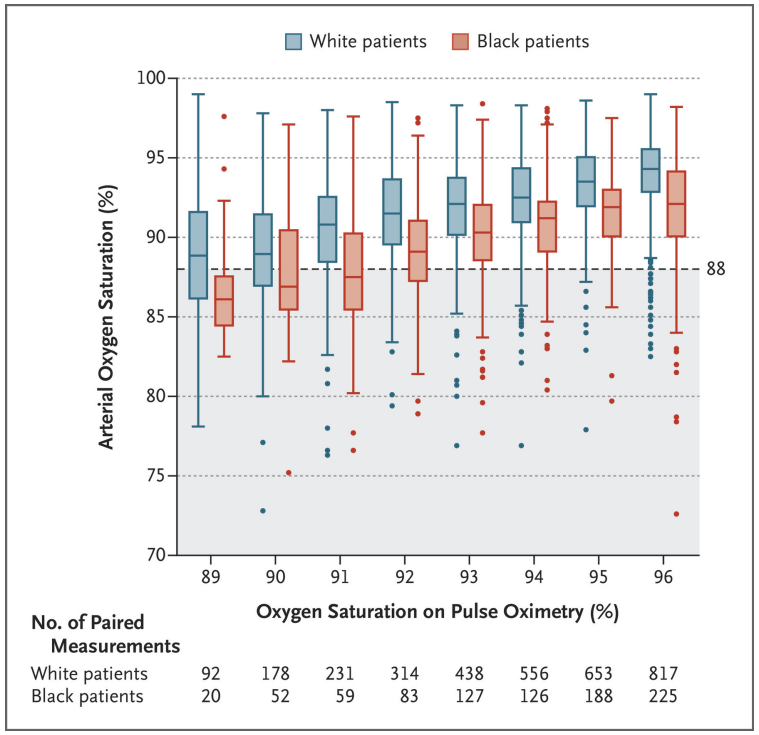



Image Source: New England Journal Medicine 2020; 383:2477-2478

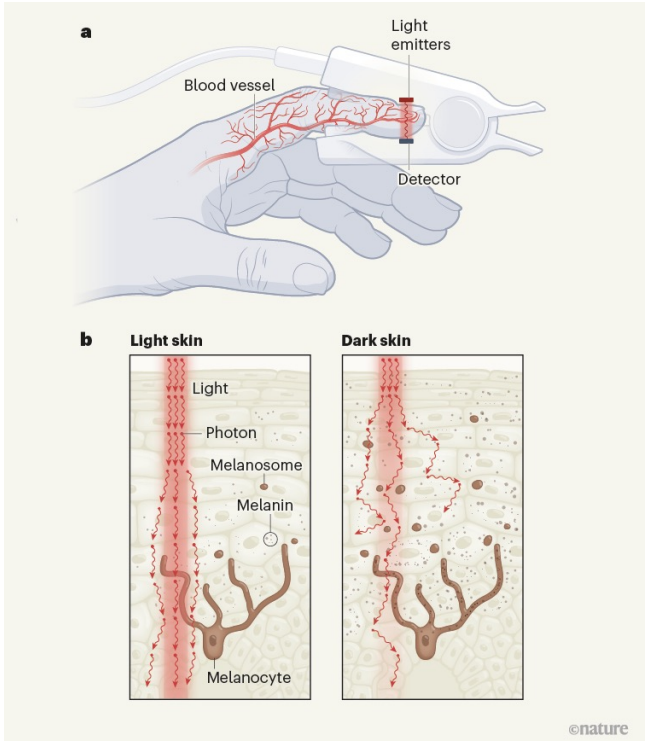



Image Source: Nature 610(7932):449-451

# Our Multivariate Approach with EMBED

**Table 4: Evaluation of Model Performance Using Multivariate Regression to Assess Risk Ratio of False Negative Predictions by Subgroups, Compared to Versus Univariate Evaluation**

| Variables | OR | RR | Univariate *p*-value | Multivariate *p*-value | Number of Patches | Control Group |
|---|---|---|---|---|---|---|
| Black | 0.880 | 0.922 | <0.001* | 0.248 | 2,630 | White |
| Other* | 0.749 | 0.828 | <0.001* | 0.050* | 1,160 | White |
| 50-60y/o | 0.881 | 0.918 | <0.001* | 0.315 | 1,798 | <50y/o |
| 60-70y/o | 0.823 | 0.875 | <0.001* | 0.163 | 1,434 | <50y/o |
| >70y/o | 0.89 | 0.924 | <0.001* | 0.482 | 840 | <50y/o |
| BI-RADS density B | 1.132 | 1.060 | <0.001* | 0.079 | 311 | BI-RADS density A |
| BI-RADS density C | 0.752 | 0.862 | 0.490 | 0.590 | 2,327 | BI-RADS density A |
| BI-RADS density D | 1.239 | 1.103 | 0.015* | 0.756 | 321 | BI-RADS density A |
| Benign * | 0.567 | 0.927 | <0.001* | 0.011* | 499 | Never Biopsied |
| Cancer | 0.778 | 0.971 | <0.001* | 0.533 | 118 | Never Biopsied |
| Mass * | 0.596 | 0.921 | <0.001* | 0.010* | 761 | No Mass |
| Asymmetry * | 0.751 | 0.854 | <0.001* | 0.040* | 3,127 | No Asymmetry |
| AD * | 2.575 | 1.037 | 0.575 | <0.001* | 413 | No AD |
| Calcification | 0.744 | 0.934 | <0.001* | 0.075 | 1,248 | No Calcification |

**Note:** Univariate two-sample Student's t-test was conducted to compare the difference in the false negative rate of bootstrap performance between subgroups and control groups. Demographic and clinical/imaging features were evaluated using a multivariate logistic regression model for descriptive analysis to control for confounding effects between the selected features. A total of 6,142 patches were inspected.
AD = Architectural Distortion; BI-RADS = Breast Imaging Reporting and Data System; OR = Odds Ratio; RR = Risk Ratio.
*Statistically significant, p<.05

CLEMSON PUBLIC HEALTH SCIENCES

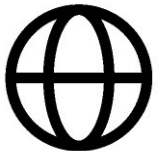
# Our Multivariate Approach with NIS

| Variables | OR | RR | p-value | Control Group |
|---|---|---|---|---|
| **Income quartile 2** | 0.988 | 0.992 | 0.308 | Income quartile 1 |
| **Income quartile 3** | 1.001 | 1.001 | 0.912 | Income quartile 1 |
| **Income quartile 4** | 0.981 | 0.987 | 0.174 | Income quartile 1 |
| **White *** | 1.279 | 1.26 | <0.001 | Asian or Pacific Islander |
| **Black *** | 1.446 | 1.412 | <0.001 | Asian or Pacific Islander |
| **Hispanic *** | 1.089 | 1.084 | <0.001 | Asian or Pacific Islander |
| **Native American *** | 1.37 | 1.344 | <0.001 | Asian or Pacific Islander |
| **Other races *** | 1.142 | 1.134 | <0.001 | Asian or Pacific Islander |
| **Medicare *** | 1.28 | 1.269 | <0.001 | Self-pay |
| **Medicaid *** | 1.186 | 1.179 | <0.001 | Self-pay |
| **Private insurance *** | 1.29 | 1.278 | <0.001 | Self-pay |
| **No charge** | 1.016 | 1.015 | 0.880 | Self-pay |
| **Other payment methods *** | 1.203 | 1.196 | <0.001 | Self-pay |
| **New England *** | 1.406 | 1.321 | <0.001 | West South Central |
| **Middle Atlantic *** | 1.352 | 1.281 | <0.001 | West South Central |
| **East North Central *** | 1.224 | 1.182 | <0.001 | West South Central |
| **West North Central *** | 1.186 | 1.152 | <0.001 | West South Central |
| **South Atlantic *** | 1.231 | 1.188 | <0.001 | West South Central |
| **East South Central *** | 1.129 | 1.106 | <0.001 | West South Central |
| **Mountain *** | 1.178 | 1.146 | <0.001 | West South Central |
| **Pacific *** | 1.338 | 1.27 | <0.001 | West South Central |
| **≥1 million population "Central" counties *** | 1.176 | 1.165 | <0.001 | Not metropolitan or micropolitan counties |
| **≥1 million population "Fringe" counties** | 1.042 | 1.04 | 0.059 | Not metropolitan or micropolitan counties |
| **250,000-999,999 population** | 1.034 | 1.032 | 0.127 | Not metropolitan or micropolitan counties |
| **50,000-249,999 population** | 1.012 | 1.012 | 0.604 | Not metropolitan or micropolitan counties |
| **Micropolitan counties *** | 0.942 | 0.945 | 0.012 | Not metropolitan or micropolitan counties |

*statistically significant, $p \leq 0.05$.

# THANK YOU

mwoo@clemson.edu



https://www.clemson.edu/cbshs