

Supplemental Figure Titles and Captions

Supplemental Figure 1: Term Frequency – Inverse Document Frequency (TF-IDF) Equation

Supplemental Figure 2: Model Development and Validation

UCSF- University of California, San Francisco, BIDMC – Beth Israel Deaconess Medical Center, AUROC – Area Under the Receiving-Operator Curve

Supplemental Figure 3: Calibration Curves for Penalized Logistic Regression Models

The bottom bar graph reports the number of patients found in each decile of predicted mortality. Mortality on the y axis of the line graph represents actual computed mortality rate for patients in each decile. The TF-IDF model does not report a 10th point because the model did not predict any patients to have a probability of mortality between 0.9 and 1.

Supplemental Figure 4: Calibration Curves for Feedforward Neural Network Models

The bottom bar graph reports the number of patients found in each decile of predicted mortality. Mortality on the y axis of the line graph represents actual computed mortality rate for patients in each decile.

Supplemental Figure 5: Calibration Curves for Random Forest Models

The bottom bar graph reports the number of patients found in each decile of predicted mortality. Mortality on the y axis of the line graph represents actual computed mortality rate for patients in each decile.

Term Frequency: $tf(t,d)$ = Number of times term t is present in document d

Document Frequency (df) = Number of times term t is present in the document corpus N

Inverse Document Frequency of term t
 $idf(t) = \log(N/(df + 1))$

TF-IDF Formula: for term t in document d within a document corpus N

$$tf-idf(t,d) = tf(t,d) * \log(N/(df+1))$$







