Table 1: Descriptive statistics for alive and dead patients

Metric	Deceased patients	Alive patients	Function to complete
Event Count 1. Average Event Count 2. Max Event Count 3. Min Event Count	1 8635 982.01	1 12627 498.12	event_count_metrics
Encounter Count 1. Average Encounter Count 2. Max Encounter Count 3. Min Encounter Count	1 203 23.04	1 391 15.45	encounter_count_metrics
Record Length 1. Average Record Length 2. Max Record Length 3. Min Record Length	0 1972 127.53	0 2914 159.2	record_length_metrics

Table 2: Model performance

Model	Accuracy	AUC	Precision	Recall	F-Score
Logistic Regression	0.7381	0.7375	0.6804	0.7333	0.7059
SVM	0.7381	0.7389	0.6768	0.7444	0.7090
Decision Tree	0.6714	0.6569	0.6329	0.5556	0.5917

Table 4: Cross Validation

CV Strategy	Accuracy	AUC
K-Fold	0.7250	0.7100
Randomized	0.7381	0.7188

My Model

I aggregated events by counting the diagnostics, medication, and lab events. Based on tests performed against features_svmlight.validate data, the Ada Boost ensemble algorithm achieved the best AUC. It fits a sequence of weak learners on repeatedly modified versions of the data. The results from these learners are assigning weights (higher for learners that more accurately predicted data and lower for those that were less accurate) and combined. I used 40 estimators in my model, achieving an AUC of approximately 77% (a 4% improvement from the SVM classifier)