Supplemental Table 1: Stop Words in the Python Natural Language Tool Kit (NLTK)

Default Stop words in Python NLTK package
i, me, my, myself, we, our, ours, ourselves, you, you're, you've, you'll, you'd, your, yours, yourself, yourselves, he, him, his, himself, she, she's, her, hers, herself, it, it's, its, itself, they, them, their, theirs, themselves, what, which, who, whom, this, that, that'll, these, those, am, is, are, was, were, be, been, being, have, has, had, having, do, does, did, doing, a, an, the, and, but, if, or, because, as, until, while, of, at, by, for, with, about, against, between, into, through, during, before, after, above, below, to, from, up, down, in, out, on, off, over, under, again, further, then, once, here, there, when, where, why, how, all, any, both, each, few, more, most, other, some, such, no, nor, not, only, own, same, so, than, too, very, s, t, can, will, just, don, don't, should, should've, now, d, ll, m, o, re, ve, y, ain, aren, aren't, couldn, couldn't, didn, didn't, doesn, doesn't, hadn, hadn't, hasn, hasn't, haven, haven't, isn, isn't, ma, mightn, mightn't, mustn, mustn't, needn, needn't, shan, shan't, shouldn, shouldn't, wasn, wasn't, weren, weren't, won, won't, wouldn, wouldn't

Supplemental Table 2: Terms Removed From the Default Stop Word List in the Python Natural Language Tool Kit (NLTK)

Stop words removed from the default NLTK package that represent negation
No, nor, not, don, don't, wasn, wasn't, weren, weren't

Supplemental Table 3: Impact of pre-processing strategies on machine learning model performance

	Feed Forward Neural Network		Random Forest	
	AUCROC (95th% CI) for UCSF Model Validated on UCSF Data	AUCROC for UCSF Model Validated on BIDMC	AUCROC (95th% CI) for UCSF Model Validated on UCSF Data	AUCROC for UCSF Model Validated on BIDMC
Raw Text	0.85 (0.80-0.90)	0.76	0.87 (0.82-0.93)	0.67
Cleaned Text	0.87 (0.82-0.92)	0.78	0.88 (0.82-0.93)	0.68
Cleaned and Stemmed Text	0.86 (0.81-0.90)	0.80	0.88 (0.82-0.94)	0.71
TF-IDF	0.88 (0.84-0.92)	0.81	0.89 (0.83-0.94)	0.79
1-3 N-grams	0.88 (0.84 -0.92)	0.78	0.88 (0.83-0.93)	0.77

AUCROC = Area Under the Receiver-Operating Characteristic Curve, CI = Confidence Interval, TF-IDF = Term Frequency – Inverse Document Frequency

Supplemental Table 4: Comparison of statistical model selection and pre-processing strategy on model performance.

	Penalized Logistic Regression		Feed Forward Neural Network		Random Forest	
Model Pre-Processing	Raw Text	TF-IDF	Raw Text	TF-IDF	Raw Text	TF-IDF
AUCROC	0.72	0.83	0.76	0.81	0.67	0.77

AUCROC = Area Under the Receiver-Operating Characteristic Curve. TF-IDF = Term Frequency – Inverse Document Frequency

Supplemental Table 5: Top 30 most predictive terms, ranked by absolute value of contribution to probability of mortality, for the penalized logistic regression model depending on the stage of text pre-processing

	Raw Text			Clean Text			Stem			TF-IDF			N-Grams	
	Term	Impact on Risk of Mortality		Term	Impact on Risk of Mortality		Term	Impact on Risk of Mortality		Term	Impact on Risk of Mortality		Term	Impact on Risk of Mortality
1	risks	Survival	1	silt	Survival	1	silt	Survival	1	pupil	Mortality	1	pupil	Mortality
2	pupils	Mortality	2	risks	Survival	2	psychiatr	Survival	2	normal	Survival	2	extub	Survival
3	equal	Survival	3	pupils	Mortality	3	cardiovascular	Mortality	3	worsen	Mortality	3	famili	Mortality
4	cardiovascular	Mortality	4	poor	Mortality	4	gastrointestin	Survival	4	famili	Mortality	4	poor	Mortality
5	monitoring	Survival	5	cardiovascular	Mortality	5	clean	Survival	5	intub	Mortality	5	worsen	Mortality
6	atraumatic	Mortality	6	psychiatric	Survival	6	pupil	Mortality	6	metastat	Mortality	6	normal	Survival
7	psychiatric	Survival	7	phone	Mortality	7	phone	Mortality	7	call	Mortality	7	well	Survival
8	role	Survival	8	gastrointestinal	Survival	8	poor	Mortality	8	diffus	Mortality	8	goal	Mortality
9	maker	Mortality	9	equal	Survival	9	extub	Survival	9	poor	Mortality	9	headach	Survival
10	planned	Survival	10	epinephrine	Mortality	10	epinephrin	Mortality	10	ascit	Mortality	10	clear	Survival
11	face	Survival	11	motion	Mortality	11	glu	Survival	11	comfort	Mortality	11	support	Mortality
12	death	Mortality	12	signs	Survival	12	face	Survival	12	hemorrhag	Mortality	12	call	Mortality
13	bedside	Mortality	13	clean	Survival	13	reconstruct	Survival	13	support	Mortality	13	anesthesia	Survival
14	gastrointestinal	Survival	14	commands	Survival	14	structur	Mortality	14	extub	Survival	14	cough	Mortality
15	lumbar	Mortality	15	clear	Survival	15	adult	Survival	15	failur	Mortality	15	bedsid	Mortality
16	these	Survival	16	monitoring	Survival	16	motion	Mortality	16	deni	Survival	16	sodium	Mortality
17	glu	Survival	17	atraumatic	Mortality	17	encount	Mortality	17	care	Mortality	17	diffus	Mortality
18	wheezes	Survival	18	removal	Survival	18	needl	Survival	18	bedsid	Mortality	18	care	Mortality
19	signs	Survival	19	headaches	Survival	19	anesthesia	Survival	19	anesthesia	Survival	19	progress	Mortality
20	commands	Survival	20	face	Survival	20	equal	Survival	20	well	Survival	20	stop	Mortality
21	q15	Survival	21	needle	Survival	21	distend	Mortality	21	discuss	Mortality	21	report	Survival
22	po4	Survival	22	iliac	Survival	22	pack	Mortality	22	effus	Mortality	22	intraven continu	Survival
23	symptoms	Survival	23	support	Mortality	23	symptom	Survival	23	goal	Mortality	23	deni	Survival
24	motion	Mortality	24	reported	Survival	24	death	Mortality	24	resect	Survival	24	albumin	Mortality
25	91	Mortality	25	glu	Survival	25	headach	Survival	25	epinephrin	Mortality	25	diseas	Mortality
26	zofran	Mortality	26	set	Survival	26	sign	Survival	26	peripher	Mortality	26	intraven prn	Mortality
27	alert	Survival	27	occupational	Mortality	27	uop	Survival	27	oper	Survival	27	alert	Survival
28	resection	Survival	28	encounter	Mortality	28	transaminas	Mortality	28	malign	Mortality	28	anterior	Mortality
29	very	Mortality	29	Night	Survival	29	clear	Survival	29	patient	Mortality	29	symptom	Survival
30	sch	Survival	30	reason	Mortality	30	anterior	Mortality	30	headach	Survival	30	discuss	Mortality

TF-IDF = Term Frequency – Inverse Document Frequency