

Data Processing

EMR data included vital signs, lab results, ventilator status and settings, demographics, and existing conditions. A full list of the EMR variables is represented in Supplemental Table 1. For data such as lab results, where there could be a lag time between the collection time and when the results are reported, collection time was used. Preprocessing was performed to remove outliers (most likely due to input errors). Acceptable ranges (with consideration of abnormal condition) for most vitals and lab variables were derived from clinical input. Outing values from these acceptable ranges were removed from the dataset. All data was initially aggregated on an hourly basis. If multiple measurements were taken within an hour for a particular variable, the median value was taken. Statistics were reported for each feature from the 8 hourly values in the window of the case or control. For continuous features, the statistics were mean, median, minimum, maximum, skewness, variance, and kurtosis. This process is depicted in Supplemental Figure 1. For procedures an indicator variable was generated, e.g. admission of a vasopressor, etc.

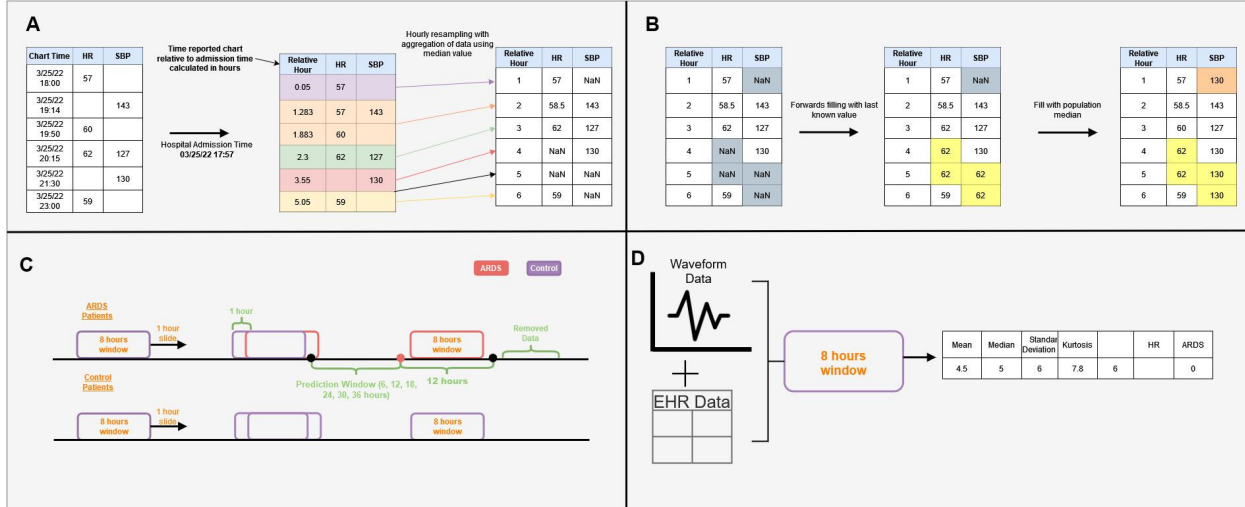
All waveform derived variables were continuous and were treated in the same fashion as the EMR continuous variables to create features. A total list of variables is available in Supplemental Figure 2. Many of the variables can be separated into categories such as time domain, frequency domain, or entropy based.

Supplemental Table1: EMR Variables

Category	Sub-category	Variable
6 statistics	Vitals	BMI Daily Weight Heart Rate Mean Arterial Pressure Respiratory Rate Systolic Blood Pressure Temperature
	Lab Values	Albumin Bicarbonate Bilirubin Calcium Chloride Creatinine FIO2 Glucose Hematocrit Lactate PCO2 PaO2 Platelet Potassium Sodium White Blood Count
	Ventilator Settings	FIO2 GCS PF Ratio Vent Rate Vent Tidal Rate Tidal Volume
	Metric	GCS
Maximum value	Concurrent Conditions	Cardiac Arrhythmia

		Chronic Pulmonary Congestive Heart Failure Diabetes complicated Diabetes Uncomplicated Hypertension Liver Disease Metastatic Cancer Obesity Renal Failure Rheumatoid Arthritis Valvular Disease
	Treatment	Dobutamine Dopamine Epinephrine Midodrine Milrinone Norepinephrine Number of Concurrent Inotropes Number of Concurrent Vasopressors Phenylephrine Vasopressin Ventilation Duration (Min.)
Demographics	Demographics	Age Gender Race

Supplemental Figure 1: Depiction of Hourly Resampling (A), Accounting for Missing Data (B), Windowing Data for Classification (C), and Statistics Used for Continuous Variables That Make Up Features (D)



Supplemental Table 2: Waveform Derived Variables

Category	Variable
Time Domain	NNmean
	NNmedian
	NNmode
	NNvariance
	NNskew
	NNkurt
	NNiqr
	SDNN
	pnn50
	RMSSD
	avgsqi
Frequency Domain	ulf
	vlf
	lf
	hf
	lffh
	ttlpr
Other	ac
	dc
	SD1
	SD2
	SD1SD2
Entropy	SampEn
	ApEn
	SEscale1
	SEscale2

	SEscale3
	SEscale4
	SEscale5
	SEscale6
	SEscale7
	SEscale8
	SEscale9
	SEscale10
	SEscale11
	SEscale12
	SEscale13
	SEscale14
	SEscale15
	SEscale16
	SEscale17
	SEscale18
	SEscale19
	SEscale20