A. MIMIC3 Preprocessing for Survival Forecasting

We use the publicly available dataset MIMIC3 (Johnson et al., 2016) and then follow the preprocessing of Harutyunyan et al. (2017) for the in-hospital mortality prediction task. It excludes the neonatal and pediatric patients and patients with multiple ICU stays. The training set consists of $35,725$ patients with 10.81% mortality rate, and test set has $6,294$ patients with 9.94% mortality rate. We then split 15% of our training set as our validation set.

For classifier training, we uniformly take 6 timepoints within the last 24 hours of each patient trajectory. For each prediction point, we set the label as 1 if the the patient die in the encounter and 0 otherwise. For each patient, we evaluate on a uniformly spaced grid points with separation of 3 hours starting backward from the patient's last time until the maximum 24 hours, resulting in maximum 7 points per patient. We only include the prediction points with at least 3 hours of history and 5 measurement values. The ultimate goal is to predict whether patient dies within 24 hours given the past 24 hours of observations.

For RL, we take the last 24 hours of each dying patient and discretize it into 30 minutes interval. We only include the patients with at least 12 hours to remove unstable trajectories. Note that we set the RL trajectory and classifier prediction horizon as the last 24 hours of the patient. This could avoid the label confounding problem (Paxton et al., 2013) as there might be unrecorded intervention that increase patient's health condition and change the label.

We select 38 static demographic features and clinical. Age, Gender, Ethnicity, congestive heart failure, cardiac arrhythmias, valvular disease, pulmonary circulation, peripheral vascular, hypertension, paralysis, other neurological, chronic pulmonary, diabetes uncomplicated, diabetes complicated, hypothyroidism, renal failure, liver disease, peptic ulcer, aids, lymphoma, metastatic cancer, solid tumor, rheumatoid arthritis, coagulopathy, obesity, weight loss, fluid electrolyte, blood loss anemia, deficiency anemias, alcohol abuse, drug abuse, psychoses, depression. The features are curated from the official MIMIC repository [1] with the comorbidity concept.

We show the feature choices and their count in Appendix Table 2. We select 39 time-series measurements with counts at least 1% of the count of heart rate, which is the largest count of the measurement in our data. We further log-transform, remove outliers outside of 2 inter-quantile regions (IRQ), and standardize time series measurement values for zero-mean unit-variance for each feature.

[1] https://github.com/MIT-LCP/mimic-code/tree/master/concepts

B. Classifier Training Details and Performances

We train the RNN as follows. We use LSTM with 1 hidden layer of 32 nodes. We regularize the neural network with $\lambda = 1\text{e}{-5}$ as ℓ_2 regularization and dropout rate of 0.3 for input and output layer, and 0.5 for the hidden layer. We use mean imputation (set the missing value as the feature mean) for the time-series features, and add missingness indicators for each feature (Lipton et al.). We discretize the time series into 1-hour interval and take average value if there are multiple measurements per interval. The RNN takes these 1-hour discretized grid point for up to 24 hours time point to classify. We train two other baselines: Logistic Regression (LR) with ℓ_2 regularization as $\lambda = 1\text{e}{-5}$ (selected by cross validation), and Random Forest (RF) with 500 trees. We concatenate all the features, as long as missingness indicators across all the time points, resulting in $24 * (39 * 2) + 38 = 1910$ features. We also use mean imputation for these features.

Table 2. Time variant features and their counts after preprocessing

Feature	Count	Relative Count %
Anion gap	213442	0.051
Bicarbonate	219802	0.052
Blood urea nitrogen	220854	0.053
Calcium (total)	185718	0.044
Chloride (blood)	225476	0.054
Creatine kinase	44459	0.011
Creatinine (blood)	221715	0.053
Diastolic blood pressure	3929745	0.935
Glascow coma scale total	627577	0.149
Glucose (blood)	313798	0.075
Heart Rate	4204926	1.0
Hematocrit	253045	0.06
Hemoglobin	196859	0.047
Magnesium	218030	0.052
Mean blood pressure	3904218	0.928
Mean corpuscular hemoglobin	194995	0.046
Phosphate	189261	0.045
Platelets	205492	0.049
Potassium	241110	0.057
Prothrombin time	139231	0.033
Red blood cell count (blood)	194997	0.046
Sodium	229893	0.055
Systolic blood pressure	3930865	0.935
Temperature (C)	797435	0.19
White blood cell count (blood)	196268	0.047
CO2 (ETCO2, PCO2, etc.)	263161	0.063
Oxygen saturation	101518	0.024
Partial pressure of carbon dioxide	263153	0.063
Partial thromboplastin time	149675	0.036
pH (blood)	285076	0.068
Bilirubin (total)	47707	0.011
Lactate	84510	0.02
Lactic acid	89347	0.021
Positive end-expiratory pressure	53689	0.013
Fraction inspired oxygen	375335	0.089
Calcium ionized	140283	0.033
Alanine aminotransferase	46850	0.011
Alkaline phosphate	45809	0.011
Asparate aminotransferase	46808	0.011

Table 3. Relative action frequency of physician policy in 24h mortality forecasting task

Feature	Relative action frequency
Anion gap	0.0021
Bicarbonate	0.0118
Blood urea nitrogen	0.0022
Calcium (total)	0.0120
Chloride (blood)	0.0022
Creatine kinase	0.0122
Creatinine (blood)	0.0059
Diastolic blood pressure	0.0101
Glascow coma scale total	0.0046
Glucose (blood)	0.0125
Heart Rate	0.0022
Hematocrit	0.0123
Hemoglobin	0.0642
Magnesium	0.0029
Mean blood pressure	0.0085
Mean corpuscular hemoglobin	0.0148
Phosphate	0.0662
Platelets	0.0143
Potassium	0.0112
Prothrombin time	0.0023
Red blood cell count (blood)	0.0024
Sodium	0.0122
Systolic blood pressure	0.0635
Temperature (C)	0.0111
White blood cell count (blood)	0.0029
CO2 (ETCO2, PCO2, etc.)	0.0059
Oxygen saturation	0.0073
Partial pressure of carbon dioxide	0.0103
Partial thromboplastin time	0.0116
pH (blood)	0.0009
Bilirubin (total)	0.0134
Lactate	0.0068
Lactic acid	0.0112
Positive end-expiratory pressure	0.0126
Fraction inspired oxygen	0.0643
Calcium ionized	0.0133
Alanine aminotransferase	0.0193
Alkaline phosphate	0.0112
Asparate aminotransferase	0.0075

Table 4. Hyperameters and ranges for Dueling DQN

Parameter	Range
Num. of representation layers	$\{1, 2, 3, 4\}$
Num. of dueling layers	$\{1, 2, 3, 4\}$
Dim. of NN layers	$\{16, 32, 64, 128\}$
Learning rate	$\{5e-2, 1e-3, 5e-3, 1e-4, 5e-4, 1e-5, 5e-5, 1e-6\}$
L2 reg. constant	$\{5e-1, 1e-1, 5e-2, 1e-2, 5e-3, 1e-3, 5e-4, 1e-4\}$
Dropout keep prob.	$\{1.0, 0.9, 0.8, 0.7, 0.6, 0.5\}$
Training batch size	$\{32, 64, 128, 256, 512\}$
Action cost coefficient λ	$\{1e-4, 5e-4, 1e-3, 5e-3, 1e-2\}$

Table 5. Hyperameters and ranges for information gain estimator in OPPE

Parameter	Range	The best model
Num. of representation layers	$\{1, 2, 3, 4\}$	1
Dim. of NN layers	$\{16, 32, 64, 128, 256, 512\}$	64
Learning rate	$\{1e-2, 1e-3, 1e-4, 1e-5, 1e-6, 1e-7\}$	1e-3
L2 reg. constant	$\{1e-2, 1e-3, 1e-4, 1e-5, 1e-6, 1e-7\}$	1e-4
Dropout keep prob.	$\{1.0, 0.9, 0.8, 0.7, 0.6, 0.5\}$	0.7
Training batch size.	$\{64, 128, 256, 512, 1024\}$	64