

# Analysis of medical data using machine learning methods (MLHC)

Michal Alayev, Lian Galanti

## supplementary material

**Supplementary Material 1:** Percentage of cohort patients without any documentation of the data/variable during the hospitalization prior blood culture sampling.

Data/variable	MIMIC Model A	MIMIC Model B	eICU Model A
Antibiotics treatment	76.42	79.05	75.4
Pressor sedatives treatment	70.96	80	72.22
Presence of lines	71.8	80	80.95
Liquid input	65.27	76.19	73.81
Liquid output	65.85	76.19	7.94
Mean Airway Pressure	24.52	24.76	84.92
Minute Volume (Obser)	50.28	40	* -
Peak Insp. Pressure	26.05	26.67	83.33
PEEP	50.15	40	34.13
Respiratory Rate	0.85	0	2.38
Tidal Volume (observed)	25.92	26.67	-
Tidal Volume (spontaneous)	40.33	35.24	-
Albumin	32.36	29.52	22.22
Anion Gap	0.07	0	14.29
Arterial Blood Pressure diastolic	77.11	84.76	61.9
Arterial Blood Pressure systolic	77.07	84.76	61.9
Base Excess	10.34	8.57	21.43
Basophils	25.14	25.71	10.32
Bicarbonate	0.07	0	0
Bilirubin, Total	31.22	28.57	22.22
Calcium, Total	1.72	0.95	6.35
Calculated Total CO2	10.34	8.57	80.16
Chloride	0.07	0	0
Creatinine	0.1	0	0
Glucose	0.13	0	0
Heart Rate	0.68	0	2.38
Hematocrit	0.07	0	0
Hemoglobin	44.59	45.71	0
Lactate	19.77	14.29	38.1
Lymphocytes	25.14	25.71	8.73
MCV	0.1	0	0

Neutrophils	25.14	25.71	15.87
pCO2	10.34	8.57	12.7
pH	9.04	6.67	11.9
Platelet Count	0.1	0	0
pO2	10.31	8.57	12.7
Red Blood Cells	0.1	0	0
SpO2	0.75	0	2.38
Temperature Celsius	75.93	65.71	3.17

\* - indicates that there was no available data in eICU database for patients in the cohort.

**Supplementary Material 2:** The data included in the datasets for both models:

- **Demographic:** gender, age, duration (in days) of hospitalization and ICU stay prior to the BC sampling time, number of distinct ICU stays during the hospitalization, mortality in hospital, duration (in hours) from hospital admission to ICU admit and to BC sampling time, from ICU admit time to BC sampling time, and from BC sampling time to discharge.
- Height (cm), weight (kg) and BMI on admission, and daily weight measurements.
- **Vital signs:** heart rate (bpm), arterial blood pressure diastolic/systolic/mean (mmHg), O2 saturation pulseoximetry (SpO2) (%), temperature (Celsius), non-invasive blood pressure diastolic/systolic/mean (mmHg).
- **Respiratory parameters:** FiO2 (Fraction of inspired oxygen), respiratory rate (insp/min), PEEP (positive end-expiratory pressure), minute ventilation, inspiratory time, mean airway pressure, peak insp. pressure, plateau pressure, and tidal volume.
- **Laboratory Measurements:**
  - Complete blood count: C-Reactive Protein, Red Blood Cells, Hemoglobin, Hematocrit, MCV, MCH, MCHC, RDW, Platelet Count.
  - Leukocytes differential: Neutrophils, Lymphocytes, Monocytes, Eosinophils, Basophils.
  - Basic metabolic panel: sodium (mEq/L), chloride (mEq/L), potassium (mEq/L), magnesium (mEq/L), calcium (mg/dL), glucose (mg/dL), bicarbonate (mEq/L), creatinine (mg/dL), urea nitrogen (mg/dL), phosphate (mg/dL), anion gap (mEq/L), pH.
  - Blood gases: base excess (mEq/L), calculated total CO2 (mEq/L), PCO2 (mmHg), PO2 (mmHg), lactate (mmol/L).
  - Coagulation panel: PTT (sec), INR (PT), PT (sec)
  - Liver function: total bilirubin, total protein, albumin, Alkaline Phosphatase (ALP), Alanine Aminotransferase (ALT) (GPT), Aspartate Aminotransferase (AST) (GOT).
- **Daily fluid balance:** Daily urine output (ml), and daily liquid input (fluids intake, blood products/colloids) (ml).
- **Medications:** Timing and dosage of antibiotic treatment and pressor-sedatives medications.

- **Interventions:** Presence and insertion timing of central venous lines, arterial lines, and dialysis catheters.
- **GCS score:** describes the level of consciousness in a person following a traumatic brain injury. The scores are divided to motor, eye, and verbal response, including total score.
- **Microbiology:** Blood, urine, stool, sputum, and IV-catheter tip cultures, that were taken up to 3 or 2 days (in model A or B respectively) prior to the cohort's BC sampling time.

**Supplementary Material 3:** Data cleaning- the plausible ranges that we defined for vital signs, laboratory tests, and respiratory parameters:

Variable	Minimal value	Maximal value
Heart Rate	30	220
Temperature Celsius	25	45
Mean Arterial Blood Pressure Mean Non Invasive Blood Pressure	15	250
Diastolic Arterial Blood Pressure Diastolic Non Invasive Blood Pressure	15	200
Systolic Arterial Blood Pressure Systolic Non Invasive Blood Pressure	30	250
Alanine Aminotransferase (ALT) Aspartate Aminotransferase (AST)	0	20000
Urea Nitrogen	2	200
Calcium, Total	0	21
Glucose	0	2000
Albumin	0.5	6.5
Bilirubin, Total	0.1	70
Creatinine	0.1	28
Potassium	0.05	13
Sodium	90	215
Lactate	0.1	30
pH	6.6	7.8
SpO2	0	100
Bicarbonate Calculated Total CO2	0	9999
Alkaline Phosphatase Chloride MCH MCHC MCV	0.1	9999
C-Reactive Protein Platelet Count pO2	0	1000
pCO2	0	200
Hemoglobin	2	25

Red Blood Cells	1	8
RDW	5	40
Hematocrit	5	75
Lymphocytes		
Neutrophils	0.2	100
INR (PT)	0.5	15
PT	5	100
PTT	5	200
Minute Volume	0	50
Tidal Volume	0	1200
PEEP	0	30
Inspiratory Time	0	10
Respiratory Rate	0	65
FiO2	0	100
Weight	0	300
Height	100	240

**Supplementary Material 4:** Features selected as an input to the models.

**Model A** - the 40 selected features, ordered by F-value and p-value, as was derived from feature selection algorithm using ANOVA test (KBest function).

Model A Features	F-value	P-value
Got any Antibiotics during hospital stay	590.2325	2.84E-124
Got any Antibiotics in the last day before target time	394.5226	5.29E-85
Number of different Antibiotics the patient got	376.5579	2.47E-81
Got any Pressor-Sedatives during hospital stay	372.0001	2.11E-80
Patient had Central Venous Line during hospital stay	371.1773	3.11E-80
Number of Blood Cultures taken in the ICU prior to target	332.9902	2.16E-72
Number of invasive lines inserted during hospital stay	332.5341	2.68E-72
Hours from ICU intime to target time	328.8644	1.53E-71
Number of different Pressor-Sedatives the patient got	302.0902	5.22E-66
Patient had Arterial Line during hospital stay	301.7064	6.27E-66
Central Venous Line, amount during hospital stay	300.4978	1.12E-65
Doses of any Pressor-Sedatives the patient got	288.0862	4.18E-63
Heart Rate max	284.0275	2.91E-62
Doses of any Pressor-Sedatives in the last day before target time	282.4547	6.19E-62
Hours from first Antibiotics dose to target time	278.497	4.11E-61
Got any Pressor-Sedatives in the last day before target time	273.8462	3.82E-60
Number of different Antibiotics in the last day before target time	267.0907	9.75E-59
Heart Rate, delta between last and min	257.8917	8.08E-57
pH, hours to target from min	254.5403	4.05E-56
Heart Rate last	253.3127	7.31E-56

Arterial Line, amount during hospital stay	248.1093	8.94E-55
Bicarbonate, delta between last and min	242.6419	1.24E-53
Calculated Total CO2, hours to target from min	237.3592	1.59E-52
Hematocrit amount	236.4172	2.51E-52
Doses of any Antibiotics in the last day before target time	234.6219	5.96E-52
pCO2, hours to target from min	233.5288	1.01E-51
Platelet Count, hours to target from max	233.1534	1.21E-51
Central Venous Line, hours from first insertion to target	231.3353	2.91E-51
Hours from first Pressor-Sedatives dose to target time	230.7847	3.80E-51
GCS Total, hours to target from first	229.6775	6.49E-51
Number of different Pressor-Sedatives in the last day before target time	222.0058	2.65E-49
Sodium amount *	219.3154	9.73E-49
Magnesium amount *	213.5737	1.57E-47
Potassium amount *	213.5522	1.59E-47
pO2 min	212.3097	2.90E-47
Glucose amount *	212.0007	3.36E-47
Heart Rate, delta between last and 25th percentile	211.6543	3.98E-47
Base Excess, delta between last and min	206.188	5.64E-46
Anion Gap amount *	203.5418	2.04E-45
Respiratory Rate, hours to target from min	203.1949	2.41E-45

\* amount is the number of times the test was taken from the patient

**Model B** - the 10 selected features, ordered by the logistic regression coefficients, as was derived from the logistic regression feature selection algorithm.

Model B Features	coefficients
Minute Volume (Observed), hours to target from max	0.156396
Tidal Volume (spontaneous), delta between last and 25th percentile	0.132404
Heart Rate, hours to target from last	0.131909
Respiratory Rate, hours to target from last	0.131885
GCS Total, hours to target from last	0.121135
SpO2, hours to target from last	0.120791
Tidal Volume (spontaneous), delta between last and median	0.120728
Hematocrit, delta between last and 25th percentile	0.117973
Tidal Volume (spontaneous), delta between last and average	0.116917
Respiratory Rate max	-0.116838