Machine Learning Classifier Models Can Identify ARDS Phenotypes Using Readily Available Clinical Data

Pratik Sinha, Matthew M Churpek, and Carolyn S Calfee

ONLINE DATA SUPPLEMENT

XGBOOST

XGBoost is an implementation of the gradient boosted machine algorithm, and we utilized the "xgbtree" setting, which is based on an ensemble of individual decision trees. A decision tree is an algorithm that partitions the data by using a series of discrete cut-points that best separate those individuals that have the event of interest vs. those who do not. In XGBoost, these individual trees are sequentially combined together in order to correct for errors of the incumbent trees in the model. Several hyperparameters can be tuned in order to alter the complexity and behavior of the algorithm, including how much each individual tree is weighted in the model (Eta), how large each tree is allowed to grow (Max depth), and how many trees are included in the final model (Nround). These hyperparameters are explained in more detail below:

Eta (learning rate): This hyperparameter represents the shrinkage at each boosting step when new trees are added and limits the rate of learning. Smaller eta results in fewer corrections at each step. Grids were created to seek an optimal eta between a range of 0.01 to 0.5.

Max_depth: This parameter dictates the size and complexity of the tree. Increasing the depth of the tree increases complexity of the model and risks overfitting. Grid search was conducted to seek an optimal maximum depth between 2 and 10.

Nround: This denotes the number of individual decision trees in the final model. Increasing the number of trees can increase the model complexity but also increases the risk of overfitting. Grid search was conducted to seek the optimal number of trees between 10 and 2000.

Ten-fold cross-validation in the training datasets using the R package *caret* was used to seek the values for each hyperparameter that maximized the area under the receiver operating characteristic curve (AUC). All other hyperparameters were set at default setting.

XGBoost offers several advantages over standard regression algorithms (e.g., logistic regression) in that it automatically includes non-linear associations and deep interactions between variables in the model. It has the advantage over other gradient boosted machine implementations due to the fact that it implements a system of parallel tree construction which allows rapid and efficient processing (i.e., the "extreme" in XGBoost ("eXtreme Gradient Boosting") refers to its computational efficiency. Another advantage of XGBoost is its ability to handle missing values automatically. On the assumption that the values are missing at random, XGBoost models have a unified method of dealing with these data. When constructing individual decision trees, at each node, the direction for missing data is selected based on learning from the data such that algorithm performance is maximized.(1) The missing data in this study are summarized in **Table E8**.

Figure E1: Differences in the plasma biomarker levels in the validation dataset (SAILS) at baseline in the hypo-inflammatory and hyper-inflammatory phenotypes as identified by the clinical-classifier model developed in the primary analysis. P-values represent the Wilcoxon rank sum test. E1A Intercellular adhesion molecule-1 (Y-axis upper limit restricted to 2000 with two hyper-inflammatory observations censored)**. E1B Plasminogen activator factor-1** (Y-axis upper limit restricted to 100 with seven hyper-inflammatory observations censored)**.**

Table E1. Comparison of variables between the training dataset, composed of three component cohorts (ARMA, ALVEOLI, FACTT) and the validation dataset (SAILS).

	Training	Validation
Number of patients (n)	2022	745
Sex Female (n)	901 (45%)	380 (51%)
Race		
White	1409 (70%)	590 (79%)
Other	613 (30%)	155 (21%)
Body mass index (kg/m^2)	27.9 $\pm$ 7.3	30.7 $\pm$ 10
Age (years)	50 $\pm$ 17	54 $\pm$ 16
Temperature (°C)	38.4 $\pm$ 1.0	38.1 $\pm$ 1.0
Systolic Blood Pressure (mmHg)	88 $\pm$ 17	85 $\pm$ 16
Heart rate (bpm)	125 $\pm$ 22	118 $\pm$ 23
PaO_2/FiO_2 ratio	131 $\pm$ 61	139 $\pm$ 64
Tidal Volume (mL)	518 $\pm$ 139	413 $\pm$ 87
Minute Ventilation (L/min)	12.5 $\pm$ 4.0	10.8 $\pm$ 3.2
PEEP (cm H_2O)	10 (5 - 12)	10 (5 - 11)
$PaCO_2$ (mmHg)	39.1 $\pm$ 9.6	40.3 $\pm$ 10.8
Respiratory rate (breath/min^{-1})	32 (26 - 40)	32 (27 - 38)
Haematocrit (%)	30 $\pm$ 6	30 $\pm$ 6
White Blood Cells (10^3/µL)	14.7 $\pm$ 12	15.7 $\pm$ 12.4
Platelets (10^3/µL)	183 $\pm$ 125	186 $\pm$ 125
Sodium (mmol/L)	137 $\pm$ 6	138 $\pm$ 5
Creatinine (mg/dL)	1.52 $\pm$ 1.4	1.53 $\pm$ 1.2
Glucose (mg/dL)	129 $\pm$ 60	125 $\pm$ 49
Albumin (g/dL)	2.2 $\pm$ 0.6	2.2 $\pm$ 0.6
Bilirubin (mg/dL)	0.8 (0.5-1.7)	0.8 (0.5-1.4)
Bicarbonate (mmol/L)	21.4 $\pm$ 5.5	21.8 $\pm$ 5.5
ARDS Risk Factors:		
Trauma	178 (9%)	6 (1%)
Sepsis	478 (24%)	145 (19%)
Aspiration	305 (15%)	49 (7%)
Pneumonia	837 (41%)	529 (71%)
Other	224 (11%)	16 (2%)
Vasopressor at Enrollment (n)	647 (32%)	407 (55%)
Hyperinflammatory phenotype	591 (29%)	277 (37%)
Ventilator Free Days	17 (0 - 23)	20 (0 - 25)
Mortality at 90 days	575 (28%)	204 (27%)

PEEP = Peak end-expiratory pressure

Table E2. 2 X 2 table comparing accuracy of phenotype classification derived by the clinical classifier model (Round 2) to latent class analysis (LCA) derived classification in the secondary analysis validation dataset (FACTT) using a probability cut-offs of ≥ 0.5 to assign phenotype.

	LCA Assigned Hyper-inflammatory class	LCA Assigned Hypo-inflammatory Class	Total
Clinical-classifier derived Hyper-inflammatory class	213 (Sensitivity 0.86)	109	322
Clinical-classifier derived Hypo-inflammatory Class	36	642 (Specificity 0.85)	678
Total	249	751	

The presented LCA-assigned phenotypes are extracted from the merged dataset combining ARMA, ALVEOLI and FACTT.

Table E3. 2 X 2 table comparing phenotype classification derived by the clinical classifier model (Round 3) to latent class analysis (LCA) derived phenotypes in the secondary analysis validation dataset (ALVEOLI) a probability cut-off of ≥ 0.5 to assign phenotype.

	LCA Assigned Hyper-inflammatory class	LCA Assigned Hypo-inflammatory Class	Total
Clinical-classifier derived Hyper-inflammatory class	136 (Sensitivity 0.82)	41	177
Clinical-classifier derived Hypo-inflammatory Class	29	343 (Specificity 0.89)	372
Total	165	384	

The presented LCA-assigned phenotypes are extracted from the merged dataset combining ARMA, ALVEOLI and FACTT.

Table E4 Area under the receiver operating characteristic curves (AUC) for classifier models comprising of the combined sparse variable groups in the validation cohort (SAILS) of the primary analysis. The groups were sequentially added starting with the group with the highest AUC followed by the next highest.

Classifier Model Composition	Area under the curve (CI)
Laboratory only	0.917 (0.90 – 0.94)
Laboratory and Vital Signs	0.944 (0.93 – 0.96)
Laboratory, Vital Signs, and Respiratory	0.946 (0.93 – 0.96)
Laboratory, Vital Signs, Respiratory and demographics	0.950 (0.94 – 0.96)

CI: Confidence Interval

Table E5 Model performance and accuracy of clinical-classifier Secondary Model 1 over a range of probability cut-offs in the validation dataset (FACTT). For each phenotype, proportions of patients, mortality at day-90 and p-values for interaction term of phenotypes with randomized intervention (with mortality as outcome) are also presented.

Probability Cut-off	Sensitivity	Specificity	Accuracy	Total patients, n (%)		Mortality at day 90, n (%)		p-value for treatment interaction
				Hypo-inflammatory	Hyper-inflammatory	Hypo-inflammatory	Hyper-inflammatory	
≥ 0.3	0.89	0.80	0.82	629 (63%)	371 (37%)	123 (20%)	161 (43%)	0.0128
≥ 0.4	0.88	0.84	0.85	659 (66%)	341 (34%)	136 (21%)	148 (43%)	0.0068*
≥ 0.6	0.83	0.88	0.87	705 (70%)	295 (30%)	155 (22%)	129 (44%)	0.0150*
≥ 0.7	0.77	0.92	0.88	748 (75%)	252 (25%)	168 (22%)	116 (46%)	0.0090*

* Denotes $p < 0.05$

Table E6 Model performance and accuracy of clinical-classifier Secondary Model 2 over a range of probability cut-offs in the validation dataset (ALVEOLI). For each phenotype, proportions of patients, mortality at day-90 and p-values for interaction term of phenotypes with randomized intervention (with mortality as outcome) are also presented.

Probability Cut-off	Sensitivity	Specificity	Accuracy	Total patients, n (%)		Mortality at day 90, n (%)		p-value for treatment interaction
				Hypo-inflammatory	Hyper-inflammatory	Hypo-inflammatory	Hyper-inflammatory	
≥ **0.3**	0.90	0.80	0.83	324 (59%)	225 (41%)	62 (19%)	86 (38%)	0.0490*
≥ **0.4**	0.86	0.85	0.85	349 (64%)	200 (36%)	67 (19%)	81 (41%)	0.0567
≥ **0.6**	0.76	0.93	0.88	397 (72%)	152 (28%)	79 (20%)	69 (45%)	0.0061*
≥ **0.7**	0.66	0.95	0.86	419 (76%)	130 (24%)	87 (21%)	61 (47%)	0.0533

* Denotes $p < 0.05$

Table E7 Model performance and accuracy of sparse model (composed of vital signs and laboratory variables) over a range of probability cut-offs. For each phenotype, proportions of patients, mortality at day-90 and p-values for interaction term of phenotypes with randomized intervention (with mortality as outcome) are also presented. A: Round 1 with SAILS as the validation dataset. B: Round 2 with FACTT as the validation dataset. C: Round 3 with ALVEOLI as the validation dataset.

Probability Cut-off	Sensitivity	Specificity	Accuracy	Total patients, n (%)		Mortality at day 90, n (%)		p-value for treatment interaction
				Hypo-inflammatory	Hyper-inflammatory	Hypo-inflammatory	Hyper-inflammatory	
A: Primary Analysis: SAILS as validation								
≥ 0.3	0.85	0.88	0.87	455 (61%)	290 (39%)	95 (21%)	109 (38%)	0.9989
≥ 0.4	0.75	0.92	0.86	497 (67%)	248 (33%)	112 (23%)	92 (37%)	0.5107
≥ 0.6	0.60	0.98	0.84	571 (77%)	174 (23%)	133 (23%)	71 (41%)	0.1447
≥ 0.7	0.49	0.99	0.80	602 (81%)	143 (19%)	147 (24%)	57 (40%)	0.3368
B: Secondary Analysis Model 1: FACTT as validation								
≥ 0.3	0.87	0.79	0.81	625 (62%)	375 (38%)	135 (22%)	149 (40%)	0.0688
≥ 0.4	0.84	0.83	0.84	665 (66%)	335 (34%)	140 (21%)	144 (43%)	0.0449*
≥ 0.6	0.80	0.89	0.87	719 (72%)	281 (28%)	163 (23%)	121 (43%)	0.0159*
≥ 0.7	0.76	0.91	0.87	742 (74%)	258 (26%)	171 (23%)	113 (44%)	0.0351*
C: Secondary Analysis Model 2: ALVEOLI as validation								
≥ 0.3	0.86	0.86	0.86	355 (65%)	194 (35%)	75 (21%)	73 (38%)	0.0392*
≥ 0.4	0.79	0.90	0.87	379 (69%)	170 (31%)	81 (21%)	67 (39%)	0.0162*
≥ 0.6	0.67	0.95	0.87	420 (77%)	129 (23%)	86 (20%)	62 (48%)	0.0375*
≥ 0.7	0.56	0.98	0.85	450 (82%)	99 (18%)	95 (21%)	53 (54%)	0.0136*

Table E8. Summary of the number of missing observations in the variables used as predictors

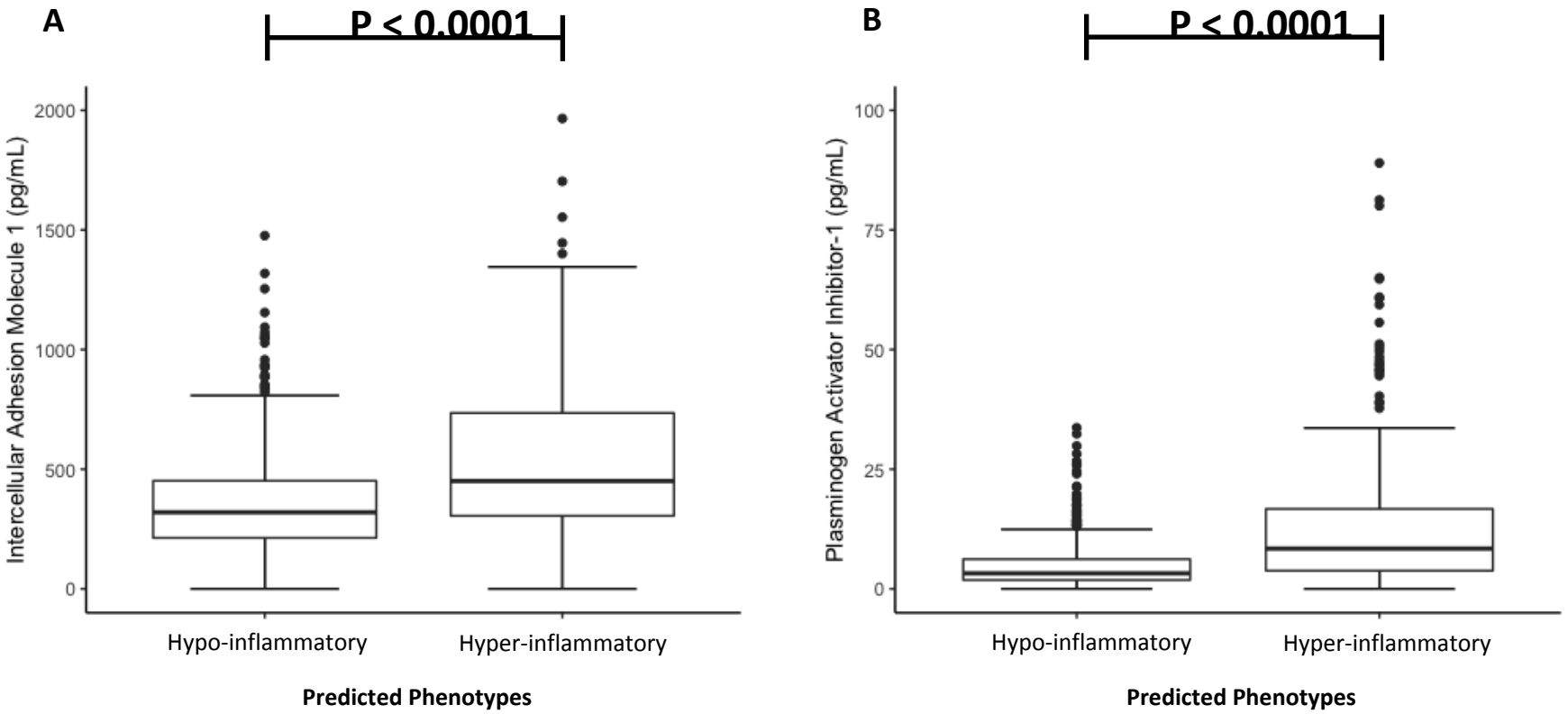
	ARMA	ALVEOLI	FACCT	Merge	SAILS
Number of patients	473	549	1000	2022	745
Body mass index	32	44	84	160	2
Temperature	0	1	2	3	0
Systolic Blood Pressure	1	1	2	4	0
Heart rate	0	1	2	3	0
PaO_2/FiO_2 ratio	1	0	0	1	0
Tidal Volume	146	42	121	309	213
Minute Ventilation	4	5	27	36	43
PEEP	0	3	4	7	22
$PaCO_2$	34	25	39	98	12
Respiratory rate	0	1	2	3	4
Haematocrit	3	1	5	9	1
White Blood Cells	26	4	13	43	1
Platelets	5	5	8	18	0
Sodium	9	1	3	13	0
Creatinine	27	2	5	34	1
Glucose	10	4	20	34	1
Albumin	62	35	220	317	96
Bilirubin	45	33	268	346	92
Bicarbonate	4	1	15	20	5
Vasopressor at Enrollment	0	0	0	0	1

for the classifier models in the cohorts used in the analysis. For the primary analysis, Merge served as the training dataset and SAILS as the validation dataset.

Reference:

1. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA: ACM; 2016. p. 785-794.

Figure E1


A
P < 0.0001
Intercellular Adhesion Molecule 1 (pg/mL)
2000
1500
1000
500
0
Hypo-inflammatory
Hyper-inflammatory
Predicted Phenotypes
B
P < 0.0001
Plasminogen Activator Inhibitor-1 (pg/mL)
100
75
50
25
0
Hypo-inflammatory
Hyper-inflammatory
Predicted Phenotypes