Appendix

A. ML classification models and feature selection.

The variables used to build up the model:

- *physical examination on admission*: temperature, HR BPM, SBP, DBP, RR /min. SpO_2, SpO_2 on RA vs. O_2 Therapy, GCS, SOFA score
- *symptoms on admission*: cough, sputum, sore throat, chest pain, SOB, fever, headache, confusion, having any gastrointestinal symptoms (e.g., nausea, vomiting, diarrhea), myalgia, malaise, loss of smell or taste.
- *laboratory findings on admission*: the count of WBC, platelet, and lymphocyte; the concentration of hemoglobin, total bilirubin, D-Dimer, creatinine, sodium, C-reactive protein, troponin, ferritin, fibrinogen; the activity of ALT, AST, CK, LDH; APTT.

Feature selection:

To check if there are unique patterns within the data that can unambiguously identify if the patient is going to be transferred to the intensive care unit, we utilized ML algorithms.

To assess the importance of the features fed to the ML models as predictors of admitted to ICU patients, we employed four ensemble tree-based estimators such as AdaBoost, Gradient Boosting, Random Forest, and Extra Trees. These models were trained on the whole dataset and used to rank the features in ascending order concerning their predictive potential. Figure 1 and Table 1 display the averaged values of impurity-based attribute ranks, where the average for each feature is calculated as the mean of rank values for the four ML methods mentioned above.

Table 1: Ranking scores of the variables selected for predicting the disease severity

Score	Feature	Score	Feature	Score	Feature	Score	Feature
0.19429	SOFA score	0.02520	Temperature	0.01164	Total bilirubin	0.00466	Sore Throat
0.10168	Clinical severity	0.01748	SOB	0.01135	SBP	0.00445	Troponin
0.08745	O2 therapy	0.01712	ALT	0.00983	Fever	0.00367	Confusion
0.08061	RR/min	0.01623	APTT	0.00969	GCS	0.00309	GI symptoms
0.04127	LDH	0.01595	Hemoglobin	0.00896	Ethnicity	0.00287	Cough
0.03829	Lymphocytes	0.01545	SpO2 on RA vs O2 Therapy	0.00732	HR BPM	0.00188	Malaise
0.03223	SpO2	0.01505	Na	0.00637	Myalgia	0.00186	Chest pain
0.03212	D-Dimer	0.01383	AST	0.00633	Sputum	0.00141	Smell/taste loss
0.03125	CRP	0.01382	CK	0.00524	DBP	0.00000	Creatinine
0.03067	Platelet	0.01360	WBC	0.00513	Headache	0.00000	Ferritin
						0.00000	Fibrinogen

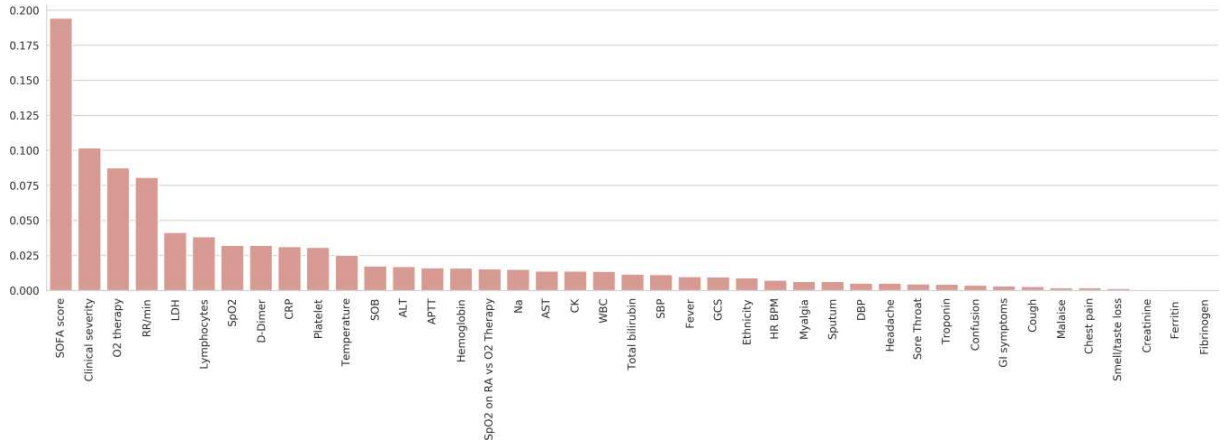

Figure 1: Feature selection for predicting whether a patient is going to be transferred to ICU

Table 2: Specificity and sensitivity of the ML model applied to the all features as predictors of the severity of the disease

ML model	Specificity	Sensitivity
Gradient Boosting	0.5972	0.9734
AdaBoost	0.6667	0.9775
Extra Trees	0.7361	0.9693
Random Forest	0.75	0.9795
NN	0.9938	1.0
SVM linear	0.6806	0.9508
Logistic regression	0.6667	0.952
Logistic regression (L1 penalty)	0.7083	0.959

Table 3: Confusion matrix to assess the accuracy of classification with a three-layer dense NN model to predict the severity of the disease

		Predicted	
		Not admitted to ICU	Admitted to ICU
Actual	Not admitted to ICU	485	3
	Admitted to ICU	0	72

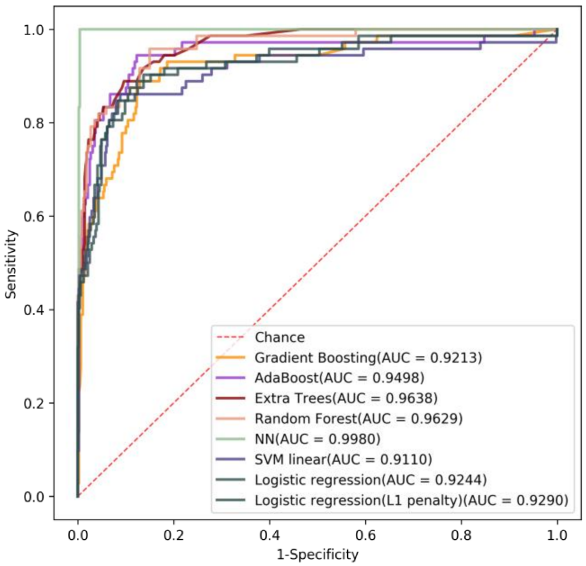

Figure 2: The performance of the employed NN classification method.

Prediction of transferring to ICU. We utilized three-layer fully connected NN with the following configuration of hidden layers (35, 30, 10) and with the stochastic gradient descent optimizer. The learning rate hyperparameter of the model was assigned to 0.1. The model was also regularized using L2 penalty with 0.0001 alpha value. NN was trained for maximum 100 epochs or before converged. Convergence implies that the loss function is not improving by at least 0.0001 for 10 consecutive iterations.

To evaluate the classifier output quality, we trained several ML classification models using a stratified 10-fold cross-validation technique to generalize the models to the true rate error. For each fold, we used 90% of the data to train the model and then tested it on the rest 10%.

The decision matrices built on the test dataset for all folds were combined and used to calculate the performance metrics. The best performance measures were obtained with a three-layer fully connected NN.

Table 4: Classification metrics of the NN model to predict the event of being transferred to ICU

	Recall	Precision	F1 score	Support
Not admitted to ICU	1.00	0.99	1.00	488
Admitted to ICU	0.96	1.00	0.98	72
accuracy			0.99	560
macro average	0.98	1.00	0.99	560
weighted average	0.99	0.99	0.99	560

B. ROC curves for laboratory tests used as input to NN.

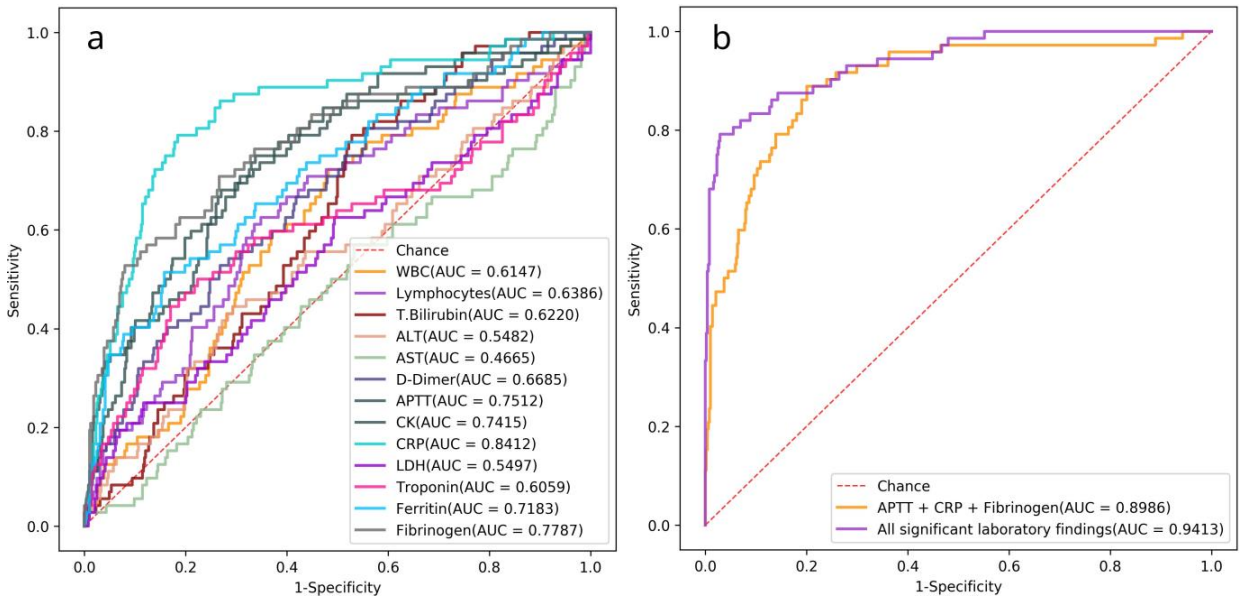

Figure 3: ROC curves for the laboratory tests used as input to NN separately (a) and in the combination (b). The models are trained with 10 folds cross-validation.

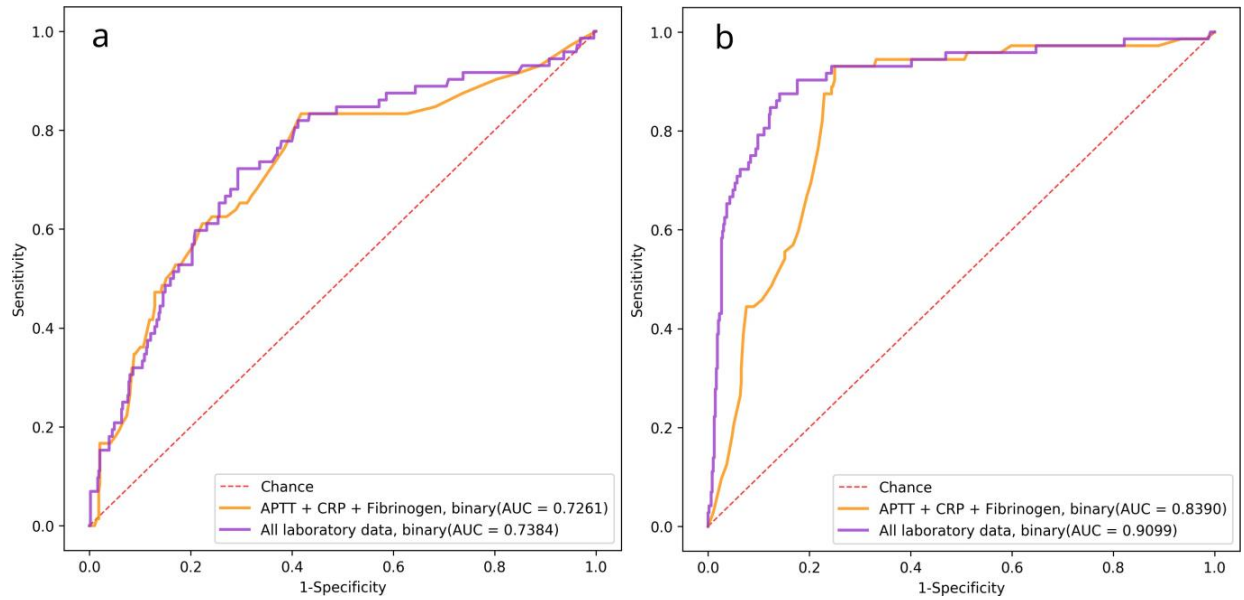

Figure 4: The performance of the 10 folds cross-validation logistic regression model trained on binary data with the threshold moving technique returned by the ML estimator (a), and with the cut-off level set to the 25^{th} percentile for lymphocyte count and 75^{th} for the other features (b).