ORIGINAL ARTICLE

Predicting thromboembolic complications in COVID-19 ICU patients using machine learning

Davy van de Sande[1], Michel E. van Genderen[1*], Babette Rosman[1], Maren Diether[1,2], Henrik Endeman[1], Johannes P.C. van den Akker[1], Martijn Ludwig[1,2], Joost Huiskens[1,3], Diederik Gommers[1], Jasper van Bommel[1]

1. Department of Adult Intensive Care, Erasmus University Medical Center, Rotterdam, the Netherlands
2. Deloitte Netherlands, Analytics & Cognitive, Amsterdam, the Netherlands
3. SAS Institute, Health Care Analytics, Huizen, the Netherlands

*Corresponding author
Michel E. van Genderen
Erasmus Medical Center , Department of Adult Intensive Care, Room Ne-411, Doctor Molewaterplein 40, 3015 GD Rotterdam, the Netherlands
Email: m.vangenderen@erasmusmc.nl

ABSTRACT

BACKGROUND

The coronavirus disease 2019 (COVID-19) pandemic is a challenge for intensive care units (ICU) in part due to the failure to identify risks for patients early and the inability to render an accurate prognosis. Previous reports suggest a strong association between hypercoagulability and poor outcome. Factors related to hemostasis may therefore serve as tools to improve the management of COVID-19 patients.

AIM

The purpose of this report is to develop a model to determine whether it is possible to early identify COVID-19 patients at risk for thromboembolic complications.

METHODS

We analyzed electronic health record data of 108 consecutive COVID-19 patients admitted to the adult ICU of the Erasmus University Medical Center between February 27 and May 20, 2020. By training a decision tree classifier on 66% of the available data, a model for the prediction of thromboembolic complications was developed.

RESULTS

The median (inter quartile range) age was 62 [53-70] years and 73% were male. Forty-three patients (40%) developed a thromboembolic complication during their ICU stay. Mortality was higher for patients in the thromboembolic complications group compared to the control group (26% vs 8%, p = 0.03). Lactate dehydrogenase, standardized bicarbonate, albumin, and leukocytes were identified by the Decision Tree classifier as most powerful predictors for thromboembolic complications two days prior to the onset of the thromboembolic complication, with a positive predictive value of 67% and a positive likelihood ratio of 2.7.

CONCLUSIONS

Clinically relevant thromboembolic complications frequently occur in critically ill COVID-19

patients. These can successfully be predicted using a decision tree model. Although this model could

be of special importance to aid clinical decision making, its generalizability and clinical impact should

be determined in a larger population.

Relevance for patients

Recently, severe thromboembolic complications were observed in COVID-19 patients with

progressive respiratory failure warranting ICU treatment. Timely identification of patients at risk of

developing thromboembolic complications (TC) is critical inasmuch as it would enable clinicians to

initiate potentially salvaging therapeutic anticoagulation.

LIST OF ABBREVIATIONS

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

Coronavirus disease (COVID-19)

Intensive care unit (ICU)

Acute respiratory distress syndrome (ARDS)

Computed tomography (CT)

Thromboembolic complications (TC)

Polymerase chain reaction (PCR)

Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis

(TRIPOD)

Body mass index (BMI)

C-reactive protein (CRP)

Interleukin-2 receptor (IL-2R)

Interleukin-6 (IL-6)

High-sensitive troponin t (hsTnT)

N-terminal prohormone of brain natriuretic peptide (NT-proBNP)

Alanine aminotransferase (ALAT)

Krebs von den Lungen-6 (KL-6)

Neutrophil gelatinase-associated lipocalin (NGAL)

Procalcitonin (PCT)

Prothrombin time international normalized ratio (PR INR)

Soluble urokinase-type plasminogen activator receptor (suPAR)

Activated partial thromboplastin time (APTT)Lactate dehydrogenase (LDH)

Interquartile range (IQR)

Area under the receiver operating characteristics (AUROC)

1. BACKGROUND

A pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection responsible for coronavirus disease (COVID-19) [1] has led to large numbers of moderately and severely ill patients. Up to 5% of COVID-19 patients need intensive care unit (ICU) treatment [2] due to severe pulmonary dysfunction, resulting in acute respiratory distress syndrome (ARDS). [3,4] Previous reports suggest that endothelial dysfunction leads to secondary activation of the complement system and coagulation cascades, culminating in hyperinflammation, the formation of microthrombi, and corollary gas exchange disturbances and even fatal obstructive shock. [5-8] Early clinical identifcation of impending pulmonary issues is therefore crucial to initiate timely treatment.

Recently, severe thrombotic complications were observed in COVID-19 patients with progressive respiratory failure warranting ICU treatment.[9] Because there is failure to identify and accurately manage this risk, the National Institute for Public Health of the Netherlands advised to lower the threshold of initiating therapeutic anticoagulation.[10,11] The clinical relevance of coagulopathy in these patients is mainly characterized by a positive correlation between elevated D-dimer levels and poor prognosis. Accordingly, therapeutic anticoagulation in high-risk individuals has yielded beneficial outcomes. [12,13] Furthermore, previous studies suggest that patients with a high probability of a pulmonary embolism could benefit from preemptive therapeutic anticoagulation, even before diagnostics based on computed tomography (CT).[14] To date, studies regarding the early recognition of these high-risk patients are lacking. Timely identification of patients at risk of developing thromboembolic complications (TC) is critical inasmuch as it would enable clinicians to initiate potentially salvaging therapeutic anticoagulation. Especially in a stressful time, such as the COVID-19 pandemic, diagnosis of thromboembolic disease is difficult because few clinical CT investigations are performed either due to the overwhelming workload or due the severity of critical illness (clinical instability). Also, CT scans are often performed without iodinated intravenous contrast and very high D-dimer levels are often observed (secondary to inflammation in the later phase), complicating the interpretation.[15,16]

Vast amounts of patient data are generated at a high rate during the COVID-19 pandemic and it has been difficult for clinicians to keep up and interpret these data for each individual patient. Data analysis in the form of machine learning is capable of processing large amounts of complex patient data and may contribute to making more efficient use of data and generate accurate, patient-specific predictions.[17]

In the current study, we describe the number of TCs in critically ill COVID-19 patients warranting ICU treatment and provide a decision tree model to predict the risk of TCs.

2. METHODS

This is a single center retrospective cohort study at the adult ICU of the Erasmus University Medical Center, Rotterdam, the Netherlands. In order to be included, patients must have a confirmed SARS-CoV-2 infection via nasopharyngeal swab and PCR and had to be older than 18 years of age. The first patient was enrolled on February 27, 2020. Patient data were censored at the time of data cutoff on May 20, 2020. All patients had at least 14 days of hospital follow-up. All data were retrieved from the patient's electronic health records. The study was approved by the institutional review board of Erasmus University Medical Center and the need for written informed consent was waived. All data were de-identified. The manuscript has been prepared in accordance with the TRIPOD guideline for multivariable prediction models.[18]

2.1. Predictors

As candidate predictors for the model, we considered demographic, diagnostic, and treatment data collected during ICU stay (including laboratory test results and radiologic assessments) (Table 1). Laboratory tests were performed routinely and radiologic assessments, including plain chest

radiography, arterial and/or venous ultrasound, and CT pulmonary angiogram were performed at the

discretion of the treating physician.[9]

Table 1. Predictors used in the prediction of thromboembolic complications.

Patient information (3 features)	Age, gender, BMI
Prior diseases (8 features)	Hypertension, diabetes mellitus, other cardiovascular diseases, ischemic stroke, tumor, chronic renal insufficiency, chronic lung disease, congestive heart failure
Laboratory results (32 features)	Lymphocytes, neutrophils, neutrophil/lymphocyte ratio, CRP, leukocytes, eosinophils, Il-2R, IL-6, thrombocytes, LDH, D-dimer, CKMB, hsTNT, NT-proBNP, albumin, ALAT (GPT), BSE, CK, cystatine-C, ferritine, fibrinogen, KL6, creatinine, NGAL, PCT, pO_2, PTINR, bicarbonate, suPAR, triglyceride, APTT, APTT ratio

2.2. Outcomes

The primary outcome was a TC. Patients were considered TC-positive (event) if they were diagnosed

with pulmonary embolism established by lung CT or cardiac ultrasound (direct visualization of

thromboembolism in the heart, right ventricular dilation and/or pulmonary hypertension) or with deep

vein thrombosis during hospital admission. Plain chest radiography, arterial and/or venous ultrasound,

and CT pulmonary angiogram were performed at the discretion of the treating physician. Patients were

considered TC-negative (reference group/non-event) if 1) radiological assessment ruled out a TC

during ICU stay and/or 2) if they were discharged from the ICU during the study period and had no

clinical indication for radiological assessment during ICU stay nor during follow-up.

2.3. Sample size

No statistical sample size calculation was performed a priori. Sample size was equal to the number of

patients treated during the study period.

2.4. Data preprocessing

Predictors that contained ≥ 30% missing data were removed. We started by considering the day of

diagnosis of a TC as mark point. We analyzed all data until 5 days prior to this mark point for the TC-

positive patients. In the TC-negative patients with radiological assessment, we selected all data from

the day with most available measurements in the period from two to zero days before radiological

assessment. For the TC-negative patients who were already discharged and had no clinical indication

for radiological assessment, data was collected from the day with most available measurements in the

period of day 2 to day 6 of ICU stay. We chose day six because this was the median day of TC

diagnosis. This time window was selected to ensure that the collected data would be comparable to

that of the other patients included in the study. To additionally test model performance for the day of

ICU admission, we also included data on the day of admission.

We chose to focus on 2 days prior to the mark point because this time point is clinically relevant and

demonstrated good values in preliminary analysis. When multiple measurements of the same item

were available per day, we selected only the first measurement, usually occurring between 6 a.m. and

9 a.m. as part of the daily routine laboratory testing. BR and MD were responsible for data analysis

and modeling.

2.5. Statistical analysis

Continuous variables are presented as median and interquartile range (IQR). Categorical variables

were reported as counts and percentages (%). To assess whether a particular predictor was

significantly different between groups, we performed a Wilcoxon rank-sum test for continuous

variables and a Fisher's exact test for categorical (binary) variables. A p-value of < 0.05 was

considered statistically significant. Furthermore, the correlation between continuous variables was evaluated using Spearman's rank correlation coefficient. Data processing and statistical analyses were performed in R 3.6.3 (R foundation) [19] and Python 3.7 (Python software foundation). [20]

2.6. Model development

To create the prediction model we trained a decision tree classifier. Our choice to select a decision tree is motivated by our attempt to create an easily clinical interpretable and an easy-to-use prediction model. We collected 53 predictors that encompassed several inflammation-related and coagulation-related markers (Table 1). Only predictors that exhibited a p-value of ≤ 0.05 between TC-positive and TC-negative patients were considered for modeling. We opted for this p-value cutoff to 1) select the patient characteristics that are more likely associated with TC and 2) to reduce the number of predictors.

To enable machine learning, patients with missing data in one of the selected predictors were removed from the dataset. The predictive model was obtained by training a decision tree classifier on a randomly selected subset of 66% out of the available data. Hyperparameters (maximum tree depth: 3, minimum samples per leaf: 3, minimum samples per split: 7, criterion: gini, class weights: 1 (TC-negative), 2.5 (TC-positive)) were obtained using a grid search with 5-fold repeated 5-fold cross-validation to prevent the model from overfitting to the data. The adjusted class weights achieve a higher probability of false positives compared to false negatives (missing less patients that will experience TCs). To assess generalizability to earlier days, we back tested the model after admission to the ICU and of the five days prior to the diagnosis. We assessed model performance by generating the area under the receiver operating characteristics (AUROC) using class probabilities. [21,22] In addition, we calculated the models' sensitivity (%), specificity (%), and positive likelihood ratio. The 'decision tree classifier' of the Python sklearn-library [23] was used for modeling.

3. RESULTS

From the date of the first confirmed case up to May 20, a total of 134 patients were admitted to our

ICU, of whom 108 patients were included in the analysis (Figure 1).
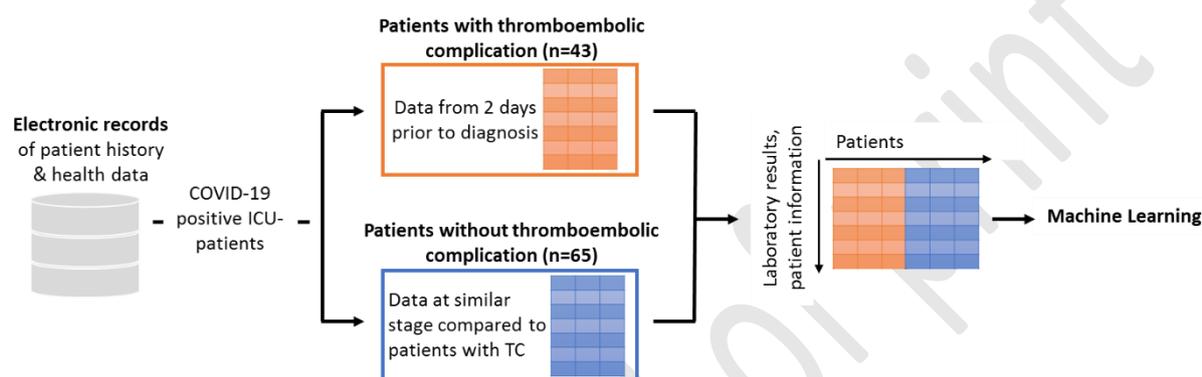
Figure 1. Schematic depiction of the extraction of patient data for machine learning. Data of COVID-19 positive ICU patients was extracted from the in-house database. For patients with thromboembolic complications, data from 2 days prior to diagnosis was selected, whereas for all other patients, the data was selected to be comparable regarding length of stay in the ICU (see section 2.4 'Data preprocessing' for more details). Finally, the predictive model was trained and tested using the resulting patient data table.

Demographic characteristics are presented in Table 2. The median age was 62 years (IQR, 53-70);

73% were male. Patients had a median BMI of 28.5 (IQR, 25.9-33.0 kg/m2). Hypertension (38

patients; 35%) and diabetes mellitus (34 patients; 31%) were the most common chronic medical

conditions. Comparing patients with TCs and those without TCs, we found no difference in the

occurrence of chronic conditions. Forty-three (40%) patients developed TCs of whom 35/43 (81%)

manifested as a pulmonary embolism. Twenty-six percent of the patients with a TC died in the ICU

compared to 8% in the non-TC group (p = 0.03).

Table 2. Patient characteristic of patients on the day of a thromboembolic complication.

		No TCs, n = 65	TCs, n = 43	p-value
Patient information	**Age [years]**	61 [46,70]	65 [56.5,70.5]	0.14
	Sex [% male]	44 (68%)	35 (81%)	0.13
	BMI [kg/m²]	30 [26,33]	27 [25,32]	0.10
Prior Diseases	**Hypertension [%]**	22 (34%)	16 (38%)	0.83
	Diabetes mellitus [%]	18 (27%)	16 (38%)	0.62
	Other cardiovascular diseases	3 (6%)	1 (2%)	0.63
	Ischemic stroke	3 (6%)	0 (0%)	0.25
	Tumor (hematological/solid)	9 (14%)	1 (2%)	0.07
	Chronic renal insufficiency	0 (0%)	3 (8%)	0.08
	Chronic lung disease	14 (22%)	9 (20%)	0.99
	Congestive heart failure	10 (16%)	8 (18%)	0.99

Continuous variables are presented as median and interquartile range [25th and 75th percentile per group]. Categorical variables were reported as counts and percentages (%). The p-value was calculated using Wilcoxon rank-sum test (continuous variables) or Fisher's exact test (categorical variables), a p-value of < 0.05 was considered significant.

3.1. Algorithm derivation and validation

To early identify patients at risk for developing TCs, we trained a decision tree model on the available patient data. Albumin, D-dimer, leukocytes, lymphocytes, standard bicarbonate (bicarbonate), and lactate dehydrogenase (LDH) were selected ($p < 0.05$) to compare TC-positive to TC-negative patients (Table 3). The neutrophil/lymphocyte ratio was excluded due to a high correlation with lymphocytes (Spearman rank correlation > 0.8).

Table 3. Patient characteristic on the day of a thromboembolic complication.

|		No TCs, n = 65	TCs, n = 43	p-value
Patient characteristics	**CT scan [%]**	25 (38%)	43 (100%)	**<0.01**
	Pulmonary embolism [%]	0 (0%)	35 (81%)	**<0.01**
	Venous thrombosis [%]	0 (0%)	8 (21%)	**<0.01**

Day of CT scan (number of days in ICU)	4 [3,11]	7 [4.5,10]	0.234
ICU days (until day when lab values were extracted)	4 [3,4]	4 [2.5,8]	0.06
Total ICU days	4 [4,5]	5 [2.5,8.5]	0.14
SOFA	5 [4,7.5]	5 [5,7]	0.47
ICU mortality [%]	4 (8%)	8 (26%)	**0.03**
pO$_2$	10.55 [9.1,12.65]	10.7 [9.9,12.3]	0.42
Activated partial-thromboplastin time [s]	31.5 [26.3,46.5]	27.5 [24.5,33]	0.19
Activated partial-thromboplastin time ratio	1.2 [1,1.6]	1.05 [0.95,1.33]	0.27
D-dimer, mg/L	1.205 [0.85,2.425]	3.275 [1.7275,4.26]	**<0.01**
Fibrinogen, g/L	6.8 [5.85,7.95]	7 [5.95,8.2]	0.49
PT INR	1.3 [1.2,1.5]	1.2 [1.1,1.25]	0.05
Platelets, × 10^9/mL	276 [197,403]	309 [261,433]	0.13
CK, U/L	116 [50,425]	133 [69,248]	0.63
CKMB, U/L	1.4 [0.8,3.5]	1.3 [0.8,2.6]	0.57
hsTnT ng/L	18 [9,40]	16 [10,53]	0.58
NT-pro-BNP, pmol/L	25 [13,62]	37 [18,70]	0.17
Eosinophils, × 10^9/mL	1.4 [0.575,2.125]	1 [0.2,1.75]	0.36
Leukocytes, × 10^9/mL	8.2 [6.1,10.95]	10.9 [9.25,15.6]	**<0.01**
Lymphocytes, %	12.6 [8.8,17.75]	9.5 [6.8,12.8]	**0.02**
Neutrophil/lymphocyte ratio	4.7 [3.5,8.65]	7.6 [5.25,10.5]	**0.01**
Neutrophils, × 10^9/mL	74.3 [68.93,79.85]	79.1 [74.4,83.1]	**0.04**
Serum ferritin, µg/L	992 [492,1819]	1255.5 [787.5,2376.25]	0.31
CRP, mg/L	190 [89,242]	189.5 [127.5,317.75]	0.09
IL-6, pg/L	88 [37,125.5]	102 [64.5,336.5]	0.07
Procalcitonin, ng/mL	0.645 [0.2,1.895]	0.815 [0.325,1.715]	0.31
Triglycerides, mmol/L	2.105 [1.573,3.17]	2.23 [1.7075,2.9925]	0.95
Creatinine, mcmol/L	76 [64,102.75]	83.5 [74.75,110.75]	0.16
NGAL, ng/mL	171[107,224.25]	233 [176,441]	**<0.01**
suPAR, ng/mL	12.5 [9,20.5]	13 [11.5,19]	0.71
ALAT (GPT), U/L	42 [26,74]	44 [28.5,67.5]	0.81
KL6, U/L	378 [249,637]	548 [399,701]	**0.02**
LDH, U/L	301 [247,415.5]	355 [315.5,450.5]	**<0.01**
Bicarbonate, mmol/L	26.3 [23.975,28.7]	28.5 [24.8,31.6]	**0.02**
Albumin [g/L]	20 [17,23]	16 [14,18.75]	**<0.01**

Continuous variables are presented as median and interquartile range [25th and 75th percentile per group]. Categorical variables were reported as counts and percentages (%). The p-value was calculated using Wilcoxon rank-sum test (continuous variables) or Fisher's exact test (categorical variables). A p-value of < 0.05 was considered significant.

Patients with missing data in one of the selected predictors were removed, yielding a final set of 76

patients for predictive modeling (33 TC-positive, 43 TC-negative). The training dataset included 50

patients (22 TC-positive, 28 TC-negative) and the testing dataset included 26 patients (11 TC-positive,

15 TC-negative).

The strongest resulting decision model assessed the levels of LDH (U/L), leukocytes ($\times 10^9$/L),

bicarbonate (mmol/L) and albumin (g/L) (Figure 2). Five-fold cross validation of the model on the

training dataset (66% of available data) yielded an accuracy of 72% and a recall of 80% on average.
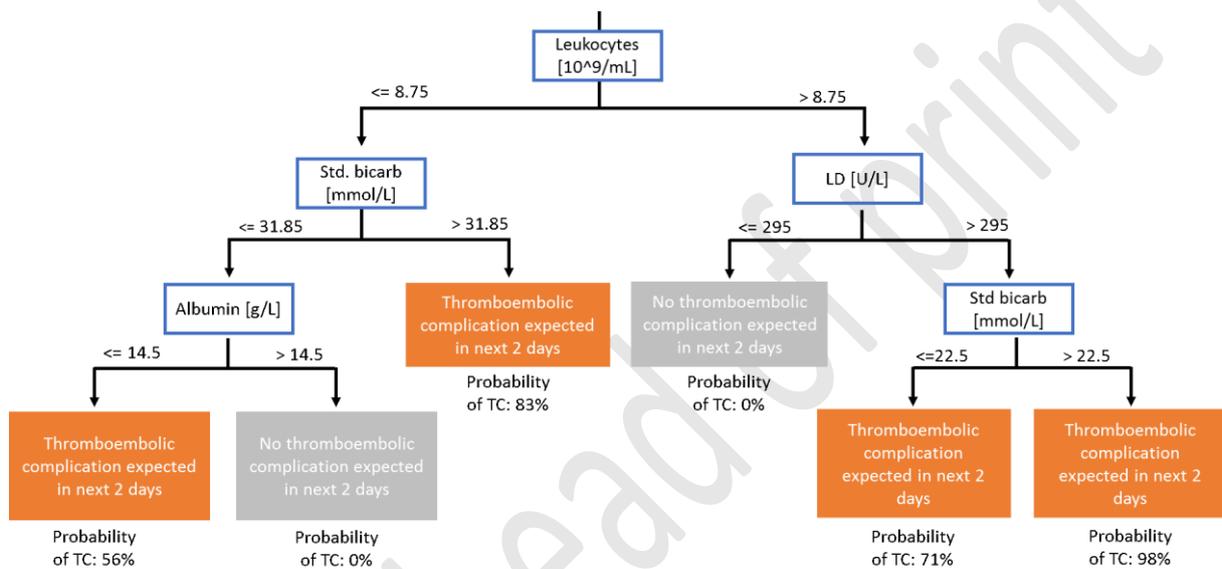
Figure 2. Schematic representation of the decision model to identify ICU patients at risk for thromboembolic complications. By following the depicted decision rules, assessing LDH, standard bicarbonate, albumin, and leukocytes, patients that are likely to develop thromboembolic complications in the next two days can be identified.

3.2. Algorithm performance

Evaluation of this model on 33% remaining data (26 patients) yielded an AUROC of 0.76, a sensitivity

of 73%, a specificity of 73%, and a positive likelihood ratio of 2.7. Based on these findings, we

subsequently assessed whether the model could identify patients up to five days before TC diagnosis

(Figure 3). Over time, the model demonstrated a varying sensitivity ranging from 50%, five days prior

to TC, to over 90% two days prior to TC. Specificity increased to 77% on two days prior to TC

diagnosis.
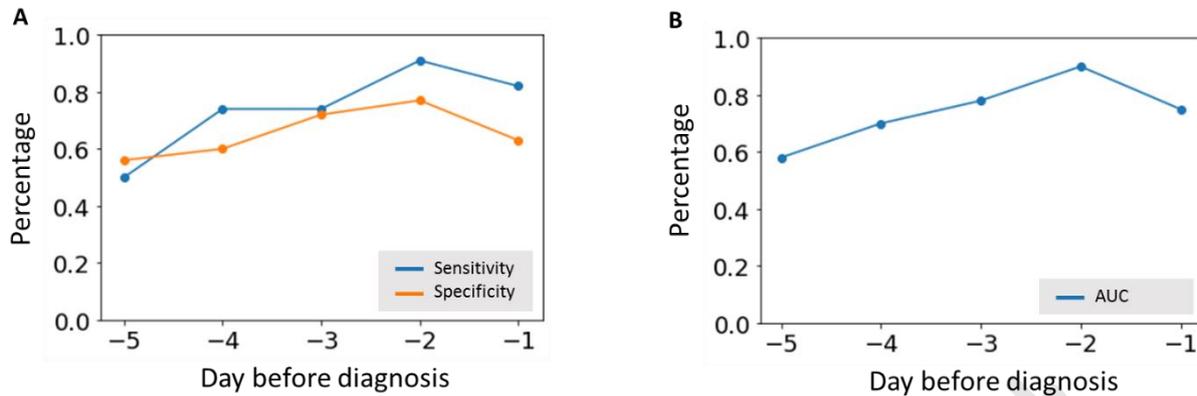
Figure 3. Model performance on the days before diagnosis. A) Sensitivity and specificity of the decision model on the five days before diagnosis expressed in percentages. B) AUC of the decision model on the five days before diagnosis.

4. DISCUSSION

In this single-center case series study, we demonstrated that a machine learning-derived decision algorithm, using commonly available laboratory values, is able to predict TCs in critically ill COVID-19 patients. Model testing demonstrated a high sensitivity and specificity with a positive predictive value of 67%, indicating that this decision model can improve risk stratification.

We found a high incidence of TCs in critically ill COVID-19 patients, mostly manifesting as pulmonary embolism, associated with increased mortality and as such expanding on the work of Tang et al. [24] and Tu et al.[7] Because the global number of COVID-19 ICU patients is still increasing drastically, early detection of complications is paramount to enable appropriate and effective interventions. This study not only affirms the association between coagulation and poor outcome but also adds clinical translation to actual prediction of such TCs.

Our model could benefit global treatment of COVID-19 ICU patients because it is generalizable and can be reproduced in other ICU departments across the globe. Moreover, our model uses standard laboratory assessments that are easily available. Additional tests (CT-pulmonary angiography and trans-thoracic ultrasound) can be targeted to true patients at risk of developing a TC, thereby shortening the diagnostic delay and enabling early initiation of therapeutic anticoagulation therapy.

However, the individual risk of bleeding must be assessed very carefully for each patient before initiating therapeutic anticoagulation.

Previous studies demonstrated that D-dimer levels were significantly elevated in COVID-19 patients admitted to the ICU with severe disease.[25-27] Increased levels of D-dimer and other fibrin degradation products have been reported in fatal cases [24] and it has even been suggested that therapy that blunts the increase in D-dimer levels might improve outcome. [28,29] In line with these observations, the National Institute for Public Health of the Netherlands advised to lower the threshold of initiation of therapeutic anticoagulation. However, this is not without possible harm because of the side effects of systemic anticoagulation and consequent bleeding. Poisey et al. therefore recently advised to be more reticent.[11] Although D-dimer was a strong predictor of TCs, and was significant different between groups, this value was not selected by the machine learning. In preliminary analysis a model including D-dimer was created but revealed no superior accuracy, sensitivity, and specificity over the current model. In addition, D-dimer contained 10% more missing data than the other variables considered for modelling and thus reduced the number of patients that could be included. Our model instead selected LDH, leukocytes, bicarbonate, and albumin. LDH is an intracellular enzyme found in lung tissue (isozyme 3) and can be released to the circulation in severe infections as part of cytokine-mediated tissue damage. Correspondingly, it makes sense from a pathophysiological point of view that LDH is released in high risk COVID-19 patients. Earlier studies demonstrated that LDH levels can be used as hallmark of disease in in patients with MERS and COVID-19. [30] Metabolic acidosis, which is influenced by bicarbonate, is not a benign condition and signifies an underlying disorder that needs to be corrected to improve outcome. Metabolic acidosis was found to be markedly lower in deceased COVID-19 patients than in recovered patients.[15]

ICUs provide a highly challenging environment for healthcare workers. Continuous routine monitoring in combination with repeated diagnostic assessments provide large amounts of data throughout the day. Rapid and appropriate decision making is therefore required to increase diagnostics throughput and to avoid a delay in treatment. Machine learning algorithms can provide decision support by uncovering hidden clinically relevant patterns. By using routinely collected data,

our method can easily be translated to other clinics. Given their routine adoption in clinical chemistry and role in COVID-19-related pulmonary pathophysiology, it makes perfect sense that our model selected LDH, leukocytes, bicarbonate, and albumin as strongest predictors of our model. Applying our model to all data up to five days prior to diagnosis revealed a sensitivity ≥ 74% already starting on day 4 before diagnosis of a TC, which is a powerful statistic despite our small sample size.

This study has several limitations. First, not all patients in the TC-negative group underwent radiological assessment (n = 25, 38%). The indication for additional radiological assessment was based on the decision of the attending physician in the wake of lacking international guidelines.[10] Second, we used a decision tree model to identify the early predictors of TCs. Decision trees are one of the many machine learning options to make use of the available data. In our preliminary analysis we also considered and shortly investigated other algorithms such as a logistic regression and neural networks. We selected a decision tree classifier over other algorithms, since the decision tree is easier to interpret and implement in clinical settings than the output of for example a logistic regression.[31] Third, the small size of the validation cohort might limit the reliability of our model. External validation in a larger or prospective dataset could further address this limitation but is not yet possible as we are currently still building a national registry in the Netherlands (https://covidpredict.org) to perform more comprehensive data analysis.

5. CONCLUSIONS

In this single-center case series study in COVID-19 patients admitted to the ICU, we demonstrated a high incidence of TCs associated with increased mortality. A simple and generalizable decision tree demonstrated a powerful model to early recognize patients at risk for TCs. This model uses conventional clinical biochemical tests and as such represents an easy stratification tool. Further research should focus on improving model and classification performance, evaluating its performance, generalizability, and clinical impact in a larger ICU patient population.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

REFERENCES

1   Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus I, Research T. A novel coronavirus from patients with pneumonia in china, 2019. N Engl J Med 2020;382:727-733.

2   Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, Yuen KY, Chen RC, Tang CL, Wang T, Chen PY, Xiang J, Li SY, Wang JL, Liang ZJ, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY, Zhong NS, China Medical Treatment Expert Group for C. Clinical characteristics of coronavirus disease 2019 in china. N Engl J Med 2020;382:1708-1720.

3   Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, Fan Y, Zheng C. Radiological findings from 81 patients with covid-19 pneumonia in wuhan, china: A descriptive study. Lancet Infect Dis 2020;20:425-434.

4   Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, Cereda D, Coluccello A, Foti G, Fumagalli R, Iotti G, Latronico N, Lorini L, Merler S, Natalini G, Piatti A, Ranieri MV, Scandroglio AM, Storti E, Cecconi M, Pesenti A, Network ftC-LI. Baseline characteristics and outcomes of 1591 patients infected with sars-cov-2 admitted to icus of the lombardy region, italy. JAMA 2020

5   Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china. JAMA 2020;323:1061-1069.

6   Campbell CM, Kahwash R. Will complement inhibition be the new target in treating covid-19 related systemic thrombosis? Circulation 2020

7   Tu WJ, Cao J, Yu L, Hu X, Liu Q. Clinicolaboratory study of 25 fatal cases of covid-19 in wuhan. Intensive Care Med 2020

8   Langer-Gould A, Smith JB, Gonzales EG, Castillo RD, Garza Figueroa J, Ramanathan A, Li BH, Gould MK. Early identification of covid-19 cytokine storm and treatment with anakinra or tocilizumab. Int J Infect Dis 2020

9   Endeman H, van der Zee P, van Genderen ME, van den Akker JPC, Gommers D. Progressive respiratory failure in covid-19: A hypothesis. Lancet Infect Dis 2020

10   Oudkerk M, Buller HR, Kuijpers D, van Es N, Oudkerk SF, McLoud TC, Gommers D, van Dissel J, Ten Cate H, van Beek EJ. Diagnosis, prevention, and treatment of thromboembolic complications in covid-19: Report of the national institute for public health of the netherlands. Radiology 2020:201629.

11   Poissy J, Goutay J, Caplan M, Parmentier E, Duburcq T, Lassalle F, Jeanpierre E, Rauch A, Labreuche J, Susen S. Pulmonary embolism in covid-19 patients: Awareness of an increased prevalence. Circulation;0

12   Rotzinger DC, Beigelman-Aubry C, von Garnier C, Qanadli SD. Pulmonary embolism in patients with covid-19: Time to change the paradigm of computed tomography. Thromb Res 2020;190:58-59.

13   Rico-Mesa JS, Rosas D, Ahmadian-Tehrani A, White A, Anderson AS, Chilton R. The role of anticoagulation in covid-19-induced hypercoagulability. Curr Cardiol Rep 2020;22:53.

14   Blondon M, Righini M, Aujesky D, Le Gal G, Perrier A. Usefulness of preemptive anticoagulation in patients with suspected pulmonary embolism: A decision analysis. Chest 2012;142:697-703.

15   Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: A retrospective cohort study. Lancet 2020;395:1054-1062.

16   Klok FA, Kruip M, van der Meer NJM, Arbous MS, Gommers D, Kant KM, Kaptein FHJ, van Paassen J, Stals MAM, Huisman MV, Endeman H. Incidence of thrombotic complications in critically ill icu patients with covid-19. Thromb Res 2020

17   Johnson AE, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine learning and decision support in critical care. Proc IEEE Inst Electr Electron Eng 2016;104:444-466.

18   Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): Explanation and elaboration. Annals of Internal Medicine 2015;162:W1-W73.

19   Team RDC. Index of /src/base/r-3, R Development Core Team, 2020,

20   Foundation PS. Python 3.7.0, Python Software Foundation, 2018,

21   Hanley JA, Mcneil BJ. The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology 1982;143:29-36.

22   Bradley AP. The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognition 1997;30:1145-1159.

23   Fabian Pedregosa GV, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Edouard Duchesnay. Scikit-learn: Machine learning in python. Journal of Machine Learning Research 2011;12:2825-2830.

24   Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost 2020;18:844-847.

25   Wang Y, Lu X, Chen H, Chen T, Su N, Huang F, Zhou J, Zhang B, Li Y, Yan F, Wang J. Clinical course and outcomes of 344 intensive care patients with covid-19. Am J Respir Crit Care Med 2020

26   Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. Lancet 2020;395:497-506.

27   Yin S, Huang M, Li D, Tang N. Difference of coagulation features between severe pneumonia induced by sars-cov2 and non-sars-cov2. Journal of Thrombosis and Thrombolysis 2020

28   Tang N, Bai H, Chen X, Gong J, Li D, Sun Z. Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. Journal of Thrombosis and Haemostasis;n/a

29   Liu X, Li Z, Liu S, Chen Z, Zhao Z, Huang Y-y, Zhang Q, Wang J, Shi Y, Xu Y, Sun J, Xian H, Fang R, Bai F, Ou C, Xiong B, Lew A, Cui J, Huang H, Zhao J, Hong X, Zhang Y, Zhou F, Luo H-B. Therapeutic effects of dipyridamole on covid-19 patients with coagulation dysfunction, medRxiv, 2020,

30   Henry BM, Aggarwal G, Wong J, Benoit S, Vikse J, Plebani M, Lippi G. Lactate dehydrogenase levels predict coronavirus disease 2019 (covid-19) severity and mortality: A pooled analysis. Am J Emerg Med 2020;38:1722-1726.

31   Doupe P, Faghmous J, Basu S. Machine learning for health services researchers. Value in Health 2019;22:808-815.